Com SIS

Computer Science and Information Systems

Volume 22, Number 3, June 2025

Contents

ISSN: 2406-1018 (Online)

Computer Science and Information Systems

Vol 22, No 3, June 2025

Com SIS

Computer Science and Information Systems

Published by ComSIS Consortium

Volume 22, Number 3
June 2025

ComSIS is an international journal published by the ComSIS Consortium

ComSIS Editorial Office:
University of Novi Sad, Faculty of Sciences,
Department of Mathematics and Informatics
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
Phone: +381 21 458 888; **Fax:** +381 21 6350 458
www.comsis.org; Email: comsis@uns.ac.rs

Volume 22, Number 3, 2025
Novi Sad

Computer Science and Information Systems

ISSN: 2406-1018 (Online)

The ComSIS journal is sponsored by:

Ministry of Education, Science and Technological Development of the Republic of Serbia
http://www.mpn.gov.rs/

Computer Science and Information Systems

AIMS AND SCOPE

Computer Science and Information Systems (ComSIS) is an international refereed journal, published in Serbia. The objective of ComSIS is to communicate important research and development results in the areas of computer science, software engineering, and information systems.

We publish original papers of lasting value covering both theoretical foundations of computer science and commercial, industrial, or educational aspects that provide new insights into design and implementation of software and information systems. In addition to wide-scope regular issues, ComSIS also includes special issues covering specific topics in all areas of computer science and information systems.

ComSIS publishes invited and regular papers in English. Papers that pass a strict reviewing procedure are accepted for publishing. ComSIS is published semiannually.

Indexing Information

ComSIS is covered or selected for coverage in the following:

- Science Citation Index (also known as SciSearch) and Journal Citation Reports / Science Edition by Thomson Reuters, with 2024 two-year impact factor 1.8,
- Computer Science Bibliography, University of Trier (DBLP),
- EMBASE (Elsevier),
- Scopus (Elsevier),
- Summon (Serials Solutions),
- EBSCO bibliographic databases,
- IET bibliographic database Inspec,
- FIZ Karlsruhe bibliographic database io-port,
- Index of Information Systems Journals (Deakin University, Australia),
- Directory of Open Access Journals (DOAJ),
- Google Scholar,
- Journal Bibliometric Report of the Center for Evaluation in Education and Science (CEON/CEES) in cooperation with the National Library of Serbia, for the Serbian Ministry of Education and Science,
- Serbian Citation Index (SCIndeks),
- doiSerbia.

Information for Contributors

The Editors will be pleased to receive contributions from all parts of the world. An electronic version (LaTeX), or three hard-copies of the manuscript written in English, intended for publication and prepared as described in "Manuscript Requirements" (which may be downloaded from http://www.comsis.org), along with a cover letter containing the corresponding author's details should be sent to official journal e-mail.

Criteria for Acceptance

Criteria for acceptance will be appropriateness to the field of Journal, as described in the Aims and Scope, taking into account the merit of the content and presentation. The number of pages of submitted articles is limited to 20 (using the appropriate LaTeX template).

Manuscripts will be refereed in the manner customary with scientific journals before being accepted for publication.

Copyright and Use Agreement

Computer Science and Information Systems

Volume 22, Number 3, June 2025

CONTENTS

Computer Science and Information Systems 22(3): i–iii https://doi.org/10.2298/CSIS250300iI

Editorial

Mirjana Ivanović, Miloš Radovanović, and Vladimir Kurbalija

University of Novi Sad, Faculty of Sciences
Novi Sad, Serbia
{mira,radacha,kurba}@dmi.uns.ac.rs

Volume 22, Issue 3 of Computer Science and Information Systems consists of 15 regular articles and one special section: "Deep Meta-Learning and Explainable Artificial Intelligence (XAI): Methodologies, Interactivity and Applications" (15 articles). As always, we are thankful for the hard work and enthusiasm of our authors, reviewers, and guest editors, without whom the current issue and the publication of the journal itself would not be possible.

We are happy to announce the updated impact factors of our journal for 2024: the new two-year IF 1.8, and the five-year IF 1.5, which is a considerable increase compared to the previous year(s).

In the first regular article, "A GAN-Based Hybrid Approach for Addressing Class Imbalance in Machine Learning," Dae-Kyoo Kim and Yeasun K. Chung present an approach based on generative adversarial networks (GAN) which uses hybrid models that combine oversampling, undersampling and ensemble techniques to reduce model overfitting, i.e., bias towards the majority class on highly imbalanced data. Experimental evaluation demonstrates improved performance compared to two variants of the synthetic minority oversampling technique (SMOTE).

The second regular article, "Empirical Analysis of Python's Energy Impact: Evidence from Real Measurements" by Elisa Jimenez et al., presents a study on whether the different ways of programming in Python have an impact on the energy consumption of the resulting programs. Conclusions include: (1) compiling Python code is a good option if it is done using the py compile module, but not Nuitka, and (2) use of dynamically typed variables seems to decrease considerably the graphics and processor energy consumption.

Sergio Ferrer-Gilabert et al., in "Stages and Critical Success Factors in ERP Implementation: Insights from Five Case Studies," propose a conceptualization of the implementation stages of an enterprise resource planning (ERP) system and identify the critical factors that ensure success. This was achieved in two stages: (1) identifying the critical success factors (CSFs) through a questionnaire distributed to information systems experts, and (2) confirming the relevance of the identified CSFs in the different stages of the proposed ERP life cycle model through semistructured interviews.

"Image Clustering Using Zernike Moments and Self-Organizing Maps for Gastrointestinal Tract," by Parminder Kaur et al. proposes a novel algorithm based on unsupervised neural classifier systems for in-vivo image clustering to address the gap between image feature representations and image semantics. Visual features are represented using the wavelet transform and Zernike moments, and a self-organizing map is utilized for the clustering of images. The system is then trained for categorizing gastral images in the respective clusters using feature-based similarity.

Laura C. Rodriguez-Martinez et al., in their article "An MDA-based Requirements Analysis Process for Service-Oriented Computing Applications," identify elements from

previously proposed requirements processes in terms of phases, activities, products, and roles/viewpoints. Using these elements, a requirements analysis process based on model-driven architecture (MDA), specifically aimed at service-oriented computing (SOC) applications. The general development process is structured in two dimensions: (1) Four general activities – requirements, design, construction and operation; (2) Three MDA models – the computational independent model, the platform independent model, and the platform specific model.

"PI2M-ITGov – Panel of Indicators for Monitoring and Maintaining the Information Technology Governance: Method and Artefacts," by Altino J. Mentzingen Moraes and Álvaro Rocha presents a method which covers 12 identified information technology (IT) areas and consists of 12 key monitoring indicators (KMIs) and their 36 sub-KMIs (3 sub-KMIs for each of the 12 identified IT areas). Simulation through a case study demonstrates a high level of acceptance of the tools as a practical IT governance alternative.

The article "Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy in Spark," authored by Yanhao Zhang et al., tackles the problem of low CPU and network resource utilization in the task scheduler process of the Spark and Flink computing frameworks by proposing a delay-aware resource-efficient interleaved task scheduling strategy (RPTS). The algorithm can schedule parallel tasks in a pipelined fashion, effectively improving the system resource utilization and shortening job completion times.

In "Identification and Detection of Illegal Gambling Websites and Analysis of User Behavior," Zhimin Zhang et al. propose a machine learning method to identify illegal gambling websites sites and analyze user behavior which combines extracting key data from post messages, generating word vectors via Word2Vec with TF-IDF, feature extraction using a stacked denoising auto encoder (SDAE), primary agglomerative clustering of websites, and secondary clustering of users' operational behaviors within website clusters.

"Classification and Forecasting in Students' Progress Using Multiple-Criteria Decision Making, K-Nearest Neighbors, and Multilayer Perceptron Methods," by Slađana Spasić and Violeta Timašević, addresses the assessment of students' performance in higher education, proposing the use of the MCDM method – Promethee II to assess students' knowledge and the K-nearest neighbors (KNN) and multilayer perceptron (MLP) methods for grade classification, with the goal of tracking and diagnosing students' knowledge levels, predicting their outcomes, and providing tailored recommendations.

In their article entitled "Image Semantic Segmentation Based on Multi-Layer Feature Information Fusion and Dual Convolutional Attention Mechanism," Lin Teng et al. propose a novel image semantic segmentation model that uses the SegFormer network as the backbone, fusing multi-scale features of encoder output with overlapping features. A dual convolutional attention module is used to fuse high-level semantic information, avoiding the loss of feature information caused by up-sampling and the influence of introducing noise.

Hui Liu et al., in "Efficient Algorithms for Collecting the Statistics of Large-Scale IP Address Data," present two algorithms for collecting the statistics of large-scale IP addresses that balance time efficiency and memory consumption. The proposed solutions take into account the sparse nature of the statistics of IP addresses while maintaining a dynamic balance among layered memory blocks. Experimental results on several synthetic

datasets show that the proposed method substantially outperforms the baselines with respect to time and memory space efficiency.

In their article "PFLIC: A Novel Personalized Federated Learning-Based Iterative Clustering," Shiwen Zhang et al. propose an iterative clustering algorithm PFLIC with the goal of mitigating data heterogeneity and improving the efficiency of federated learning (FL). The approach is combined with sparse sharing to facilitate knowledge sharing within the system for personalized FL. To ensure fairness, a client selection strategy is proposed to choose relatively "good" clients to achieve fairer federated learning without sacrificing system efficiency.

"A Spatio-Temporal Graph Neural Network for EEG Emotion Recognition Based on Regional and Global Brain," by Xiaoliang Wang et al. proposes a novel multi-scale spatio-temporal graph neural networ (MSL-TGNN), which integrates local and global brain information for the task of emotion recognition from electroencephalography (EEG) data. A multi-scale temporal learner is designed to extract EEG temporal dependencies. Also, a brain region learning block and an extended global graph attention network are introduced to explore the spatial features.

The article "Boundary-Aware Semantic Segmentation of Remote Sensing Images via Segformer and Snake Convolution," by Yanting Xia et al. introduces a novel hybrid image segmentation model that combines Segformer for global context extraction with dynamic snake convolution to better capture fine-grained, boundary-aware features. An auxiliary semantic branch is introduced to improve feature alignment across scales.

Finally, "Extending Hybrid SQL/NoSQL Database by Introducing Statement Rewriting Component," by Srđa Bjeladinović, presents a process model for applying automatic hybrid statements' rewriting, extending the architecture for the hybrid database with the newly developed statement rewriting component (SRC). Test use cases were examined on the example of Oracle/MongoDB/Cassandra hybrid before and after introducing SRC, demonstrating decrease in the average execution times of the system when SRC is used.

Computer Science and Information Systems 22(3): v–vii https://doi.org/10.2298/CSIS250300vC

Guest Editorial
Deep Meta-Learning and Explainable Artificial Intelligence (XAI): Methodologies, Interactivity and Applications

Mu-Yen Chen[1], Miltiadis D. Lytras[2], Mary Gladence[3], and Ivan Luković[4]

[1] National Cheng Kung University, Taiwan
mychen119@gs.ncku.edu.tw
[2] American College of Greece, Greece
miltiadis.lytras@gmail.com
[3] Sathyabama Institute of Science and Technology, India
marygladence.it@sathyabama.ac.in
[4] University of Belgrade, Faculty of Organizational Sciences, Serbia
ivan.lukovic@fon.bg.ac.rs

When it comes to technological development, many countries nowadays consider Artificial Intelligence (AI) to be a critical area of interest. Deep learning (DL) and machine learning (ML) are major branches of AI. Current AI systems can run many high-performance algorithms and provide advanced recognition capabilities. However, there is concern over transparency in AI development because we can only know the input information and the output results without having access to the entire computation process and other data. The 'black box' nature of DL and ML makes their inner workings difficult to understand and interpret. The deployment of explainable artificial intelligence (XAI) can help explain why and how the outputs of DL and ML models are generated. As a result, an understanding of the functioning, behavior, and outputs of models can be garnered, reducing bias and error and improving confidence in decision-making.

Explainable artificial intelligence (XAI) is one of the interesting issues that has emerged recently. Many researchers are trying to address the subject from different dimensions, and interesting results have emerged. Currently, machine learning and deep learning have been applied to many complex fields, such as medicine, finance, self-driving cars and other fields related to daily life. Since these models have been applied to these fields with good results, future applications will also expand to cognitive assistance, explainable science, and the development of reliable models. Although good results can be obtained using AI models, these models lack the disclosure of key information and the explanation of model operation. Therefore, many researchers suggest the AI model should not be just a black box, as nobody knows the reason or the detailed relationship between features and results. Hence, the concept of explainable AI was advocated. The key concept is to make the entire process of AI algorithms – from input and the decision-making process to output results – accessible and traceable. As a result, users and operators can utilize XAI to produce transparent explanations and reasons for the decisions made, reinforcing trust and confidence in an AI system's reliability.

For this special issue, it aims to explore XAI applications and researches in more areas of study and see how XAI models can take a vast amount of available data and discover undiscovered phenomena, retrieve useful knowledge, and draw conclusions and

reasoning. There are 15 papers accepted for publication. A quick overview of the papers in this issue is provided below, and we expect the content may draw attention from the public readers, and furthermore, promote societal development.

The article entitled "Analyzing the Operational Efficiency of Online Shopping Platforms Integrated with AI-Powered Intelligent Warehouses", by Wang et al. adopts the Data Envelopment Analysis (DEA) model to evaluate the operating efficiency of online shopping platforms integrated with AI, IoT, and big data analysis. The proposed system can provide convenient, secure, and novel shopping experiences for consumers.

The article entitled "Deep Learning-Driven Decision Tree Ensembles for Table Tennis: Analyzing Serve Strategies and First-Three-Stroke Outcomes", by Chang et al., develops a hybrid AI system by proposing the deep learning-driven decision tree ensemble algorithms (DLDDTEA) for table tennis match analysis. The results show the proposed system can provide a comprehensive framework for table tennis match analysis, understanding of players' strengths and weaknesses, and facilitate suitable training and competitive strategies.

The article entitled "Exploring Factors Affecting User Intention to Accept Explainable Artificial Intelligence", by Wang and Chiou, proposes a research model grounded in the characteristics of XAI and prior technology acceptance studies. The results point out that perceived value and perceived need would be the key determinants of users' intention to adopt XAI technologies and applications.

The article entitled "Development of an Explainable AI-Based Disaster Casualty Triage System", by Hsiao et al., designs and implement an XAI-based disaster casualty triage scenario system. The proposed system can develop different scenarios using generative AI (GAI) and utilize XAI to improve data transparency. Through the simulation in the training games, users can improve their judgment and responsiveness, further strengthen their rapid reaction and deal with the disaster scenarios.

The article entitled "The Integration of Artificial Intelligence and Ethnic Music Cultural Inheritance under Deep Learning", by Chang, applies the AI technology to improve the ethnic music cultural inheritance, and analyzes its music background and content in advance. This research can improve the accuracy of music emotion recognition, thus protecting the inheritance of ethnic music more effectively.

The article entitled "The Analysis of Deep Learning-based Football Training under Intelligent Optimization Technology", by Luan et al., integrates the Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNN) to optimize the action recognition, and related football data analysis in the college football training courses.

The article entitled "Three-Dimensional Visualization Design Strategies for Urban Smart Venues under the Internet of Things", by Liu, proposes the various 3D visualization methods to examine the proposed application in urban smart venues more effectively. The empirical results indicate the combination of databases and browsers would affect the 3D visualization rendering performance in IoT environments significantly.

The article entitled "Smart Home Management Based on Deep Learning: Optimizing Device Prediction and User Interface Interaction", by Liang et al., integrates the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to improve the accuracy of smart device status predictions. The proposed system can also decrease the utilization rates of the CPU, memory, GPU, and network bandwidth by 15%, 18%, 25%, and 20%, respectively.

The article entitled "Application of Deep Learning-Based Personalized Learning Path Prediction and Resource Recommendation in Graduate Education", by Li and Ding, adopts the LSTM network to predict the personal learning paths of learners. Besides, they also develop a hybrid recommendation mechanism by combining collaborative filtering and content-based filtering methods to provide the resource recommendation for the learners.

The article entitled "Effectiveness of Game Technology Applied to Preclinical Training for Nurse Aides in Implementing Contact Isolation Precautions", by Lin and Subeq, adopts the game technology and scaffolding theory to design and develop a mobile app for isolation protection capabilities of nurse aides interactive teaching. The results indicate that the proposed mobile app can significantly improve cognition, skills, and self-efficacy for the nurse aides, and demonstrates the usefulness of game technology in e-learning.

In "The Analysis of Intelligent Urban Form Generation Design based on Deep Learning", by Lian et al., presents a DL-based framework for generating intelligent urban morphology, including the network structures of deep learning models, and fine-tuning for the hyperparameter optimization.

In "Impact of Inspirational Film Appreciation Courses on College Students by Voice Interaction System and Artificial Intelligence", by Fan and Song, raises the mental health education level of college students and recommend the adoption of AI in college education for freshmen in the universities.

In "Leveraging AI and Diffusion Models for Anime Art Creation: A Study on Style Transfer and Image Quality Evaluation", by Shen et al., concentrates on the capabilities of an open-source AI image generation model. The proposed method can solve the widespread challenges of style consistency and image quality issues.

In "Usage Intention of the Reservation System of Taipei Sports Center from the Perspective of Technology Readiness Index", by Lin et al., investigates the usage intention of the reservation system of Taipei Sports Center from the perspective of Technology Readiness Index. The proposed framework can focus on the application of technology readiness theory in the sports research domain and offer a good reference for the acceptance of technology services.

The last but not the least paper is "Applying MSEM to Analyze People's Cognitive Behavior towards Virtual Reality Sport Experience", by Li et al. The authors investigate the virtual reality experience to examine the motivations and cognitive behaviors of the public opinion towards sports and fitness.

Acknowledgments. The guest editors are thankful to 38 reviewers whom were included in the review process for their effort in reviewing the manuscripts. We also thank the Editor-in-Chief, Professor Mirjana Ivanovic for the supportive guidance during the entire process.

Computer Science and Information Systems 22(3):673–691 https://doi.org/10.2298/CSIS231113014K

A GAN-Based Hybrid Approach for Addressing Class Imbalance in Machine Learning

Dae-Kyoo Kim[1] and Yeasun K. Chung[2]

[1] Computer Science and Engineering, Oakland University
Rochester, Michigan 48309, USA
kim2@oakland.edu
[2] Spears School of Business, Oklahoma State University
Stillwater, Oklahoma 74078, USA
y.chung@okstate.edu

Abstract. Class imbalance is a common problem in machine learning where the majority class has a significantly higher number of instances than the minority class, which leads to bias towards the majority class. The problem can be effectively addressed by using Generative Adversarial Network (GAN) to generate realistic synthetic samples. In this work, we present a GAN-based approach that makes use of hybrid models that combine oversampling techniques with undersampling and ensemble techniques to reduce overfitting. The proposed approach was evaluated on two datasets with different level of class imbalance using six widely used classifiers and compared with two popular class balancing techniques – SMOTEENN and SMOTETomek. The results show that the proposed approach outperforms them in highly imbalanced datasets.

Keywords: class imbalance, classification, GAN, hybrid model, machine learning.

1. Introduction

Class imbalance is a prevalent issue that can arise in many machine learning problems where one class has a significantly higher number of instances than the other class. For instance, in fraud detection (e.g., [11,35]), the majority of observations may be negative cases (e.g., no fraud), while only a few cases are positive, but critical (e.g., fraudulent transactions). Class imbalance may lead machine learning to be biased towards the majority class at the expense of the minority class [14,15,20], resulting in lower performance metrics (e.g., precision, recall, F1 score) [5].

To address class imbalance, oversampling techniques such as Synthetic Minority Oversampling Technique (SMOTE) [7,36] have been used [4]. More recently, there has been much work (e.g., [1,3,9,10,18,19,22,24,31,32]) using Generative Adversarial Network (GAN) [13] for oversampling which can generate more diverse and realistic synthetic samples, leading to competitive and, in some cases, superior performance to SMOTE. Oversampling can address class imbalance by creating artificial samples for the minority class, but it may lead to an overfitting problem. To mitigate the problem, hybrid models such as SMOTEENN and SMOTETomek have been proposed, combining an oversampling technique with an undersampling techniques like Edited Nearest Neighbours (ENN) [33] and Tomek-Links [29], which help remove noisy samples in the majority

class. However, the use of GAN as an oversampling technique with undersampling techniques has not been studied.

In this paper, we present a GAN-based hybrid approach to address class imbalance. The approach introduces a set of GAN-based hybrid models – GANBoost, GANENN, GANRUS, GANRUSBoost, and GANTomek, which combine a GAN with undersampling or ensemble techniques such as AdaBoost [12], Edited Nearest Neighbors (ENN) [7], Random Under Sampling (RUS) [27], RUSBoost (RUS+AdaBoost), and TomekLinks [29]. These models enable generating realistic samples by GANs, while removing ambiguous examples in the majority class using undersampling and ensemble techniques

We evaluated the proposed approach using two datasets – Hotel Booking Cancellation (HBC) and Financial Fraud Detection (FFD) on six different classifiers – Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), XGBoost (XGB), k-Nearest Neighbors (KNN), and Light Gradient Boosting Machine (LGBM). The datasets were purposely chosen from different domains with notably varying sizes and degree of class imbalance for the sake of diversity. The classifiers were chosen for their popular use for binary classification as concerned in the datasets. The performance of the GAN-based hybrid models are compared with SMOTEENN and SMOTETomek which are widely used hybrid models for class imbalance [4]. The results showed that GAN-based hybrid models consistently outperform both SMOTEENN and SMOTETomek in the highly imbalanced FFD dataset, which suggests the effectiveness of the proposed approach for significant class imbalance. The contributions of the work are as follows.

- Introducing a GAN-based hybrid approach to address class imbalance.
- Evaluating the approach on datasets that vary in domain, size, and degree of class imbalance.
- Comparing the performance of the approach with widely used hybrid models, and demonstrating the effectiveness of the approach on highly imbalanced datasets.

The paper is organized as follows. Section 2 gives an overview of related work using GAN to address class imbalance. Section 3 describes commonly used hybrid techniques for dealing with class imbalance. Section 4 presents the proposed GAN-based hybrid approach. Section 5 discusses the benchmark datasets used in the evaluation of the approach. Section 6 describes the results of the evaluation and compares them with existing hybrid techniques. Finally, Section 7 concludes the paper with a discussion of future research.

2. Related Work

Odena et al. [24] introduced the Auxiliary Classifier Generative Adversarial Network (AC-GAN) model to improve the training and quality of GANs for image synthesis. This model incorporates label conditioning to enable the generation of high-resolution images with greater global coherence across all classes of the ImageNet dataset. The AC-GAN model includes an auxiliary classifier within the discriminator to output class labels for training data, making the model class-conditional but also capable of reconstructing class labels. The model was evaluated using discriminability and diversity metrics, showing that 128x128 samples were more than twice as discriminable as 32x32 samples and 84.7% of classes matched or exceeded the diversity of real ImageNet data.

Mariani et al. [22] introduced Balancing Generative Adversarial Network (BAGAN) to address class imbalance in image classification by generating image samples via initializing the autoencoder of the discriminator and generator of a GAN. Class conditioning was used in the latent space to steer the generation process of image samples in a certain direction. They reported that the initialization of the autoencoder helped learn class conditioning and reduce convergence issues that arise in conventional GANs.

Antoniou et al. [3] presented Data Augmentation Generative Adversarial Network (DAGAN) to learn a model of a larger invariance space using a conditional GAN. The learned DAGAN was used to improve few-shot target domains by augmenting the data in Matching networks [30] with relevant comparator points generated from the DAGAN. The approach was evaluated on three datasets, Omniglot, EMNIST, and VGG-Face for classification tasks via vanilla classifiers (e.g., DenseNet) and Matching networks.

Wang et al. [31] introduced a traffic data augmentation method called PacketCGAN, which extends their earlier work [32] using Conditional GAN. The method was evaluated using three deep learning (DL) models on four types of encrypted traffic datasets and compared with ROS and SMOTE, and vanilla GAN. They reported that the PacketCGAN outperformed the compared methods.

Wang et al. [32] proposed FlowGAN, a GAN-based method to generate synthesized samples for encrypted traffic classification. The synthesized data was combined with real data to create a new training dataset. They evaluated the method against a Multi-Layer Perceptron (MLP) on three datasets and reported that FlowGAN outperformed the MLP on both the imbalanced dataset and oversampled dataset.

Fiorea et al. [10] used a GAN to address class imbalance in credit card fraud detection. The GAN was used to generate synthetic examples from the original minority class which were merged with original data to obtain an augmented training set. They reported that the classifier trained on the augmented set outperformed the same classifier trained on the original data.

Ali-Gombe and Elyan [1] proposed Multiple Fake Class Generative Adversarial Network (MFC-GAN) to address class imbalance in multi-classification tasks. MFC-GAN uses a multi-fake class GAN to preserve the structure of the minority class and generate samples for each class by learning the correct data distribution. The approach conditions sample generation on real class labels only and modifies the classification objective to reduce noise appearing in generated samples. They reported that MFC-GAN results in improved performance.

Jiang et al. [16] presented an anomaly detection method using GANs, specifically designed for imbalanced time series data. The method involves an encoder-decoder-encoder architecture within the generator and is trained on normal samples to understand the distribution of normal data. This approach addresses class imbalance by focusing on learning the distribution of the normal class and identifying deviations as anomalies during testing. The method was tested on benchmark rolling bearing datasets, which are widely used in the field of mechanical engineering to evaluate the performance of diagnostic algorithms and models. They reported that their method achieved 100% accuracy in distinguishing between normal and abnormal samples.

Lei et al. [19] proposed Imbalanced Generative Adversarial Fusion Network (IGAFN) to address class imbalance in credit scoring task. The network consists of a fusion module and a balance module, The fusion model is used for feature exploration, leveraging a feed-

forward neural network and bidirectional long short-term memory (Bi-LSTM) network to learn user profile and behavior data. The balance module is used for data generation, applying an imbalanced generative adversarial network (IGAN) to approximate the real data distribution and generate new samples for the minority class. The network uses a minimax algorithm to optimize the generative and discriminative networks.

Similar to Mariani et al.'s work [22], Kim et al. [17] proposed a GAN-based model to address class imbalance in defect detection within industrial inspections. The model incorporates an autoencoder as the generator, alongside two distinct discriminators for normal and anomalous inputs. They introduced Patch Loss and Anomaly Adversarial Loss functions to optimize the training process. The model was evaluated on MNIST, Fashion MNIST, CIFAR-10/100, and a real-world industrial dataset concerning smartphone case defects, achieving an average accuracy of 99.03% on the latter.

Qasim et al. [25] presented Red-GAN, a GAN-based approach equipped with class conditioning and a segmentor to mitigate class imbalance in medical imaging. This GAN is conditioned both at the pixel-level and global-level, allowing for controlled synthesis of image-label pairs tailored to specific classes. A segmentor is incorporated to ensure that the synthesized images are relevant for segmentation tasks. The method was experimented on two medical datasets – BraTS and ISIC, showing increases in DICE scores of up to 5% and 2%, respectively.

Lee and Park [18] used a vanilla GAN to address imbalanced data in intrusion detection and compared the performance of the GAN with SMOTE and single RF with no handling of data imbalance. They reported that their model outperformed both SMOTE and single RF.

Similar to Lei et al.'s work [19], Engelmann and Lessmann [9] proposed a GAN-based approach for oversampling data in credit scoring. The method was evaluated on seven credit scoring datasets and compared against benchmark oversampling methods (e.g., SMOTE, Random Oversampling) as well as without any oversampling methods, using five different classification algorithms (e.g., RF, DT, KNN). Their method outperformed four variants of SMOTE and Random Oversampling on most datasets. However, predictions made without any oversampling method generally performed better than those using SMOTE variants, and maintaining the original class distribution also delivered competitive results.

Yang and Zhou [34] introduced Imbalanced Data Augmentation Generative Adversarial Network (IDA-GAN) to tackle class imbalance in datasets. IDA-GAN incorporates a variational autoencoder to learn class distributions in the latent space, allowing for the generation of diverse samples of minority classes, thus mitigating the mode collapse issue [26] in GANs. The approach was experimented on five benchmark datasets (MNIST, Fashion-MNIST, SVHN, CIFAR-10, and GTSRB). They compared their work with AC-GAN [24] and BAGAN [22], reporting outperformance in precision, recall, and F1-score.

Bhagwani et al. [6] used GANs to address class imbalance in datasets by generating synthetic samples of the minority class. The method was evaluated on a credit card fraud detection dataset using a Support Vector Machine (SVM) for the classification of the generated samples. They reported that their approach demonstrates a classification accuracy of 99.89%, compared to SMOTE's 58.29%.

Deng et al. [8] presented Imputation Balanced Generative Adversarial Network (IB-GAN) to address the classification of multivariate time series data with strong class im-

balance. IB-GAN combines data augmentation and classification in a unified process using an imputation-balancing approach. This method employs imputation and resampling techniques to generate synthetic samples from randomly masked vectors, improving the classification of minority classes. The approach was tested on the UCR data and a 90K product dataset. They reported that their work outperforms similar existing work in F1-score.

Sharma et al. [28] proposed SMOTified-GAN which combines SMOTE and GANs to address class imbalance in datasets. It uses SMOTE for the initial oversampling of the minority class, which is then refined through GANs to produce more samples. The method was experimented on various datasets with performance improvements of up to 9% on F1-score measurements compared to other algorithms.

In summary, most of the discussed works reported the positive effect of GAN on addressing class imbalance. However, no work studied the effect of GAN with hybrid techniques of oversampling and undersampling.

3. Hybrid Class Balancing Techniques

In this section, we discuss two commonly used hybrid techniques for addressing class imbalance, namely SMOTEENN and SMOTETomek.

SMOTEENN is a data sampling method that combines SMOTE [7] and the ENN algorithm [33] to address class imbalance in datasets. SMOTEENN applies the SMOTE method to create synthetic samples for the minority class, which are generated by interpolating between the nearest neighbors of each minority class observation. This increases the number of minority class samples and helps balance the class distribution. However, SMOTE can also generate noisy samples that can negatively impact the model's performance. To address this issue, the ENN algorithm is applied, which removes samples that are misclassified by their k-nearest neighbors. For each sample, the k-NN algorithm is used to locate its nearest neighbors. Subsequently, the class of each neighbor is compared to the class of the observation. If there is a difference in class, indicating potential anomalies, both the observation and its corresponding k-nearest neighbors are eliminated from the dataset. Let D denote the dataset, and m be an individual observation within D. The class of observation m is then denoted as C_m, while the majority class of its k-nearest neighbors, denoted as k, is represented by C_k. If C_m is not equal to C_k, it signifies a discrepancy in class labels. In such cases, the observation m and its k-nearest neighbors are removed from the dataset using the set operation $D \setminus m \cup k$. This process ensures that only relevant samples are included in the dataset, which improves the overall quality of the data and reduces the impact of noise. Figure 1 illustrates the under-sampling by ENN when $k = 4$ and an example of before and after application of SMOTEENN. The example uses a synthetic classification dataset with 300 samples, 4 features, and 2 classes in a weight of 0.8 for the majority class and a weight of 0.2 for the minority class. The combination of SMOTE and ENN leads to a more balanced dataset with reduced noise, making it a useful method for improving the performance of machine learning models on imbalanced datasets [4].

SMOTETomek is another hybrid technique that combines SMOTE with Tomek Links [29]. After generating synthetic samples by SMOTE, Tomek Links identifies pairs of samples belonging to opposite classes where one sample is from the majority class and the

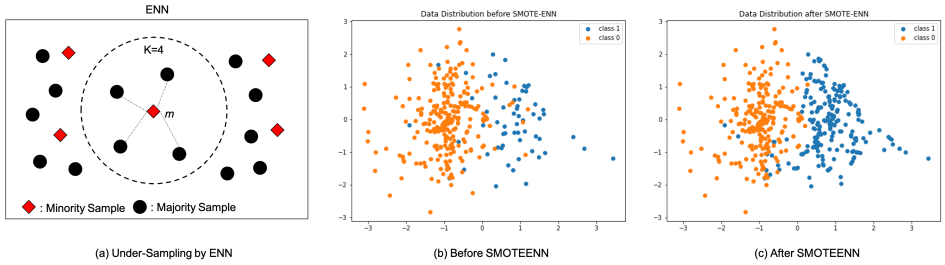


Fig. 1. Data Balancing by SMOTEENN

other is from the minority class. These paired samples are considered the closest neighbors to each other. To determine if a pair of samples, denoted as (x_i, x_j), forms a Tomek link, the Euclidean distance between them, represented as $d(x_i, x_j)$, is calculated. For a given minority class sample x_i and a majority class sample x_j, a Tomek link exists if there is no other sample x_k in the k-nearest neighbors such that $d(x_i, x_k) < d(x_i, x_j)$ or $d(x_j, x_k) < d(x_j, x_i)$. That is, a Tomek link is present when there are no neighboring samples of x_i and x_j that are closer to each other than x_i and x_j. The decision boundary between classes is improved by removing the majority class instance from a Tomek link. The process of applying SMOTE and Tomek links is repeated until a balanced dataset is achieved. Figure 1 illustrates the undersampling by Tomek when $k = 3$ and an example of before and after application of SMOTETomek.

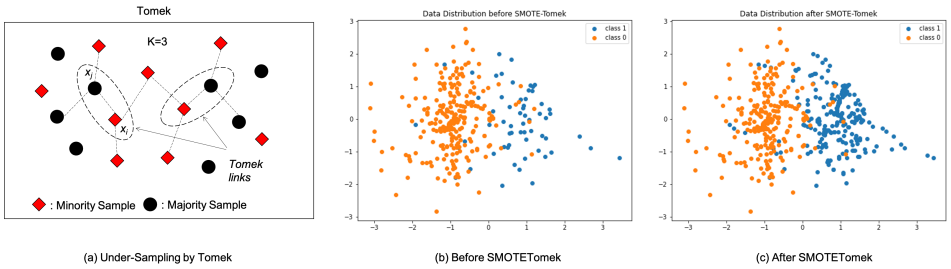


Fig. 2. Data Balancing by SMOTETomek

4. GAN-Based Hybrid Models

In this work, we present a GAN-based hybrid approach that makes use of GANs as an oversampling technique together with undersampling techniques (e.g., ENN, Tomek Links, RUS) or ensemble techniques (e.g., AdaBoost, RUSBoost) to address class imbalance problems. The approach enables creating more diverse and realistic synthetic samples to better capture the underlying data distribution using a GAN, while removing noisy,

irrelevant, or ambiguous examples in the majority class using an undersampling/ensemble technique, which helps reduce overfitting.

A GAN consists of the generator G, the discriminator D, and the classifier C [23]. G takes random noise z as input and generates synthetic samples $G(z)$, learning to map the noise distribution to the distribution of the minority class. The function $G : \mathcal{Z} \rightarrow \mathcal{T} \rightarrow \mathbb{R}^{n_i}$ can be defined as follows:

$$G(z) = g_i(t(z|i))$$

where $t : \mathcal{Z} \rightarrow \mathcal{T}$ maps the latent space z to the intermediate space $\mathcal{T}$, and $g_i : \mathcal{T} \rightarrow \mathbb{R}^{n_i}$ maps the intermediate space $\mathcal{T}$ to a vector of weights for the existing points n_i in the minority class X_i using the softmax activation function. The objective of the generator is to generate synthetic samples that resemble the real samples from the minority class by fooling the discriminator, aiming to maximize the probability of the discriminator misclassifying the synthetic samples as real:

$$\max_{\forall s \in G(z)} D(s)$$

The discriminator D distinguishes between real and synthetic samples, learning to classify whether a given sample is from the minority class or generated by the generator. The function $D : \mathbb{R}^{n_i} \rightarrow [0, 1]$ takes a real sample x from the minority class and outputs a probability $D(x)$. The objective of the discriminator is to correctly identify the real samples and distinguish them from the synthetic ones. That is, the discriminator aims at maximizing the probability of correctly classifying real samples, while minimizing the probability of misclassifying synthetic samples:

$$\max_{\forall x \in X_i} D(x) \wedge \min_{\forall s \in G(z)} D(s)$$

The classifier C evaluates the performance of the generated synthetic samples. It is initially trained on the original imbalanced dataset and then is used to assess the quality of the synthetic samples produced by the generator through a two-player game between the generator and the discriminator. The generator tries to produce synthetic samples that can fool the discriminator, while the discriminator aims to correctly classify between real and synthetic samples. This adversarial training process continues iteratively through backpropagation and gradient descent until the generator is capable of generating realistic synthetic samples that are indistinguishable from the real minority class samples according to the discriminator.

Undersampling techniques reduce the number of instances in the majority class to balance the distribution of the target variable. In addition to ENN and Tomek Links discussed in Section 3, we also use Random Under Sampling (RUS) [27] for undersampling. RUS reduces the size of the majority class by randomly selecting a subset of examples from the majority class, resulting in a more balanced dataset. The removed examples are discarded or used for validation purposes. This technique is simple and computationally efficient, making it a popular choice for addressing class imbalance. When combined with GANs, RUS can help reduce overfitting in training.

Ensemble techniques combine multiple classifiers to improve classification performance. For ensemble techniques, we utilize Adaptive Boosting (AdaBoost) [12] and Random Under-Sampling Boosting (RUSBoost). AdaBoost is a boosting algorithm that combines multiple weak classifiers into a strong ensemble. In each iteration, AdaBoost assigns

more weight to the misclassified examples, which allows the subsequent weak classifiers to focus on the misclassified examples and improve their performance. By combining the predictions of all the weak classifiers, AdaBoost generates a strong classifier that can accurately classify both the majority and minority classes. RUSBoost is a hybrid method that combines AdaBoost with random undersampling. RUSBoost first randomly undersamples the majority class to balance the class distribution and then applies the AdaBoost algorithm to the balanced dataset. By randomly removing examples from the majority class, RUSBoost reduces the computational cost of the AdaBoost algorithm while still providing a balanced dataset for training.

Figure 3 presents the proposed GAN-based hybrid approach. It first computes the number of minority samples by summing up all the samples labelled "1" (representing the minority class) and the number of majority samples by subtracting the minority count from the total number of samples. Then, it determines the significance of class imbalance by comparing the ratio of the minority count to the majority count with the threshold value. If the class imbalance is significant, it computes the number of synthetic minority samples to generate based on the difference between the majority and minority classes and the user-defined percentage of synthetic samples to generate. The GAN is then trained on the minority class samples, and synthetic minority samples are generated using the trained GAN. These synthetic minority samples are combined with the original dataset. The combined dataset undergoes balancing using Tomek, ENN, RUS, RUSBoost, or AdaBoost techniques to remove noisy and ambiguous samples introduced by the GAN. Finally, the algorithm returns the resampled dataset with a reduced class imbalance.

An instance of the GAN model is created using the $build\_gan()$ function based on the dimensionality of the input data and the dimensionality of the noise vector input. The function first builds a generator and a discriminator for the GAN instance using the $build\_generator()$ and $build\_discriminator()$ function, respectively and then combine them sequentially to create a GAN instance which is compiled with binary cross-entropy loss and $Adam$ optimizer.

The $build\_generator()$ function takes in a noise vector of length as input and outputs a synthetic sample with the same shape as the input data of dimension. It creates a sequential model with three layers. The first layer is a dense layer with 128 neurons and the input dimension (i.e., the noise vector). The second layer is the $LeakyReLU$ activation function with a slope of 0.2, which helps prevent the generator from collapsing and improves the quality of the generated samples. The third layer is another dense layer with the hyperbolic tangent ($tanh$) activation function, which scales the output values to be between -1 and 1, similar to the range of the real data.

The $build\_discriminator()$ function takes in the input dimension of the discriminator network and creates a sequential model with four layers. The first layer is a dense layer with 128 nodes and the second layer is an activation layer with $LeakyReLU$ introducing nonlinearity to the output of the first layer. A dropout layer is added after the activation function to prevent overfitting. It randomly drops out 50% of the nodes in the layer during training. Lastly, a dense output layer with a single node and the $sigmoid$ activation function is added to produce a scalar output indicating the probability that the input data is from the minority class. The model is then compiled with binary cross-entropy loss and $Adam$ optimizer.

Algorithm 1: GAN-based hybrid model

Input: X: array-like or sparse matrix, shape (n_samples, n_features); y: array-like, shape (n_samples,); $synth\_ratio$: float, the percentage of synthetic samples to generate; $threshold$: float, the threshold value for significant class imbalance; $epochs$: int, the number of epochs to train the GAN; $batch\_size$: int, the number of samples per batch to use when training the GAN; $latent\_dim$: int, the size of the noise vector input to the generator;

Output: X_bal, y_bal: balanced dataset with synthetic samples

```
G = (majority_count - minority_count) * synth_ratio          ▷ Computes # of synthetic minority samples to generate
gan = build_gan(X_min, latent_dim)                           ▷ Build a GAN on the minority class in X
for epoch in 1 to epochs do
    train_gan(X_min, generator, discriminator, gan, latent_dim, batch_size)   ▷ Train the GAN
end
X_syn = gan.generate_samples(G)                              ▷ Generate synthetic minority samples
y_syn = [1] * G                                              ▷ Assign a label of 1 to the synthetic minority samples
X_bal = combine_datasets(X, X_syn)
y_bal = combine_datasets(y, y_syn)
perform Tomek|ENN|RUS|RUSBoost|AdaBoost on combined samples
return X_bal, y_bal
def build_gan(input_dim, latent_dim):
    generator = build_generator(input_dim, latent_dim)
    discriminator = build_discriminator(input_dim)
    gan = create_model()
    gan.add(generator)
    gan.add(discriminator)
    gan.compile(loss='binary_crossentropy', optimizer='adam')
    return GAN(generator, discriminator, gan)
def build_generator(input_dim, latent_dim):
    model = create_model()
    model.add(Dense(128, input_dim=latent_dim))
    model.add(LeakyReLU(alpha=0.2))
    model.add(Dense(input_dim, activation='tanh'))
    return model
def build_discriminator(input_dim):
    model = create_model()
    model.add(Dense(128, input_dim=input_dim))
    model.add(LeakyReLU(alpha=0.2))
    model.add(Dropout(0.5))
    model.add(Dense(1, activation='sigmoid'))
    model.compile(loss='binary_crossentropy', optimizer='adam')
    return model
def train_gan(X, generator, discriminator, gan, latent_dim, batch_size):
    noise = generate_noise(batch_size, latent_dim)
    gen_samples = generator(noise)
    idx = get_random_indices(X.shape[0], batch_size)
    real_samples = X[idx]
    X_combined = combine_array(real_samples, gen_samples)
    y_combined = combine_label(ones(batch_size), zeros(batch_size))
    d_loss = discriminator.train_on_batch(X_combined, y_combined)
    noise = generate_noise(batch_size, latent_dim)
    y_mislabeled = ones(batch_size)
    g_loss = gan.train_on_batch(noise, y_mislabeled)
def generate_samples(num_samples):
    noise = get_random_indices(num_samples, latent_dim)
    synthetic_samples = generator(noise)
    return synthetic_samples
minority_count = count_minority(y)                           ▷ Compute # of minority samples
majority_count = count_majority(y)                           ▷ Compute # of majority samples
if majority_count / minority_count < threshold then
    X_bal = X, y
    return X_bal
end
```

Fig. 3. GAN-based hybrid model

The $train\_gan()$ function trains the GAN model, taking in the input data, the number of epochs, and batch size. It generates fake samples using the generator and noise and selects real samples from the input data. Then, it combines the generated and real samples with their corresponding labels (0 for fake and 1 for real) and trains the discriminator on

the combined dataset. After that, it generates noise and mislabels the generated samples as real to train the generator.

The $generate\_samples()$ function generates $num\_samples$ number of new samples using the generator and passes it along with noise to the generator to generate new samples.

Figure 4 illustrates data balancing by the proposed GAN-based hybrid models on a synthetic classification dataset with 300 samples, 4 features, and 2 classes of a weight (0.8, 0.2). Figure 4(a) shows the data distribution before applying the approach and Figure 4(b)-(f)) show the application of the approach in order of GANRUS, GANENN, GANRUSBoost, GANBoost, and GANTomek.

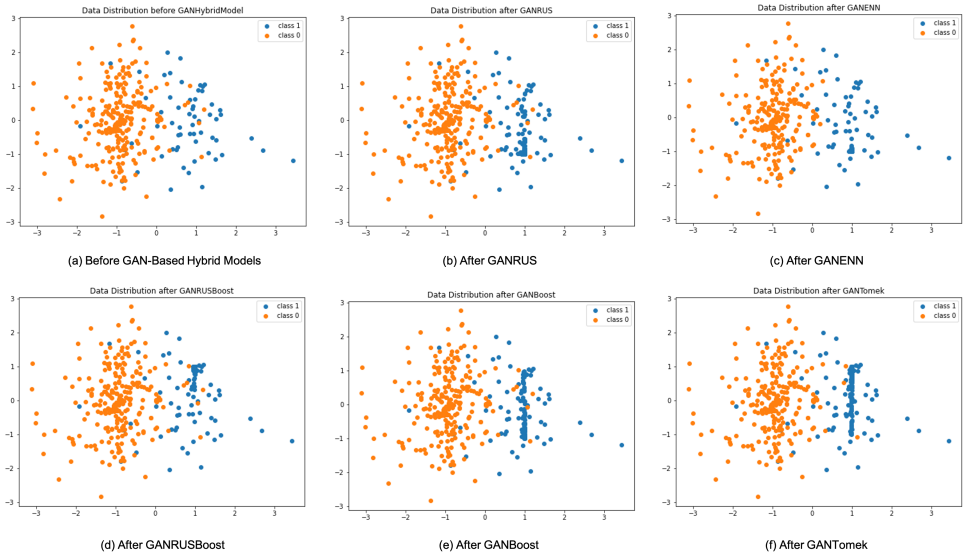


Fig. 4. Data Balancing by GAN-Based Hybrid Models

5. Datasets

We evaluate the proposed approach using two datasets with different levels of class imbalance. One is hotel booking cancellation [2] obtained from two hotels situated in Portugal, a resort hotel (RH) and a city hotel (CH) during the period between July, 2015 and August, 2017. The dataset has a size of 119,390, consisting of 40,060 RH bookings and 79,330 CH bookings. Out of these, 11,120 bookings (28%) of RH and 33,079 (42%) bookings of CH were canceled, resulting in a total of 44,224 (37%) cancellations. The dataset features are presented in Table 5. In order to ensure accurate learning, we preprocessed the dataset for null values and non-numerical values. The $country$, $agent$, and $company$ attribute contained 488, 16,340, and 112,593 null values, respectively where the null values for

$country$ were replaced with *Unknown* and those for $agent$ and $company$ were replaced with 0. The *Undefined* value for $meal$ was replaced with *SC* (self-catering). The samples with zero guests were excluded, and categorical feature values were replaced with numerical values to facilitate efficient learning.

Variable	Description
hotel	Hotel - Resort Hotel, City Hotel
is_canceled	Value indicating if the booking was canceled (1) or not (0)
lead_time	Number of days that elapsed between the entering date of the booking and the arrival date
arrival_date_year	Year of arrival date
arrival_date_month	Month of arrival date with 12 categories: "January" to "December"
arrival_date_week_number	Week number of the arrival date
arrival_date_day_of_month	Day of the month of the arrival date
stays_in_weekend_nights	Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
stays_in_week_nights	Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
adults	Number of adults
children	Number of children
babies	Number of babies
meal	Type of meal booked. Categories are presented in standard hospitality meal packages - Undefined/SC, BB, HB, FB
country	Country of origin
market_segment	Market segment designation - TA (Travel Agents), TO (Tour Operators)
distribution_channel	Booking distribution channel - TA (Travel Agents), TO (Tour Operators)
is_repeated_guest	Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_cancellations	Number of previous bookings that were cancelled by the customer prior to the current booking
previous_bookings_not_canceled	Number of previous bookings not cancelled by the customer prior to the current booking
reserved_room_type	Code of room type reserved.
assigned_room_type	Code for the type of room assigned to the booking.
booking_changes	Number of changes/amendments made to the booking from the moment the booking was entered until the moment of check-in or cancellation.
deposit_type	Indication on if the customer made a deposit to guarantee the booking.
agent	ID of the travel agency that made the booking.
company	ID of the company/entity that made the booking or responsible for paying the booking.
days_in_waiting_list	Number of days the booking was in the waiting list before it was confirmed to the customer.
customer_type	Type of booking, assuming one of four categories
adr	Average Daily Rate
required_car_parking_spaces	Number of car parking spaces required by the customer
total_of_special_requests	Number of special requests made by the customer (e.g. twin bed or high floor).
reservation_status	Reservation last status, assuming one of three categories - Canceled, Check-Out, No-Show
reservation_status_date	Date at which the last status was set.

Fig. 5. Data Features of Hotel Booking Cancellation

Another dataset is financial fraud detection obtained by PaySim [21], a data simulator that emulates real transactions with malicious transactions incorporated. This dataset comprises 6.9 million transactions with 0.13% identified as fraudulent. The features of the dataset are presented in Figure 6. In the dataset, the type of the $nameOrig$ and $nameDest$ features is of object data type, while the type of the remaining features is numerical. To ensure accurate learning, the dataset underwent validation and normalization processes where null values were eliminated, and numerical variables were transformed to values ranging between -1 and 1. The $isFlaggedFraud$ feature was omitted, as it is a subset of $isFraud$.

Feature	Description
step	It maps a unit of time in the real world. 1 step is 1 hour of time. Total steps 744 (30 days simulation).
type	Type of the transaction. There are four types -- CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.
amount	Amount of the transaction in local currency.
nameOrig	Customer who started the transaction.
oldbalanceOrg	Initial balance before the transaction.
newbalanceOrig	New balance after the transaction.
nameDest	Customer who is the recipient of the transaction.
oldbalanceDest	Initial balance recipient before the transaction. Note that there is not information for customers that start with M (Merchants).
newbalanceDest	New balance recipient after the transaction. Note that there is not information for customers that start with M (Merchants).
isFraud	Transaction made by a fraudulent agent. In this dataset, the fraudulent behavior aims to take control of a customer account and transfer the funds to another account and then cashing out of the system.
isFlaggedFraud	Transaction flagged as illegal attempt (more than 200,000 in a single transaction) by the business model monitoring massive transfers from one account to another.

Fig. 6. Data Features of Financial Fraud Detection

The financial fraud detection dataset has a minority class (fraudulent) proportion of 0.13%, which is approximately 769 times more imbalanced than the hotel booking cancellation dataset with a minority class (defective) proportion of 37%. Using the two different datasets in the level of class imbalance allows for evaluating the effectiveness of GAN-based hybrid models depending on the level of class imbalance.

6. Evaluation

We evaluated the proposed GAN-based hybrid models on six widely used machine learning classifiers – Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), XGBoost (XGB), K-Nearest Neighbors (KNN), and Light Gradient Boosting Machine (LGBM) and compared them with SMOTEENN and SMOTETomek which are commonly used hybrid models for handling class imbalance. The performance of the models was measured in terms of accuracy, precision, recall, F1 score, ROC AUC, and PR AUC.

Figure 7 shows the performance of GAN-based hybrid models on the hotel booking cancellation dataset which has 10,885 samples with 63% non-cancelled and 37% cancelled, which is lllized in Figure 8. The results show that SMOTEENN outperformed all other models across multiple classifiers in terms of accuracy. Specifically, SMOTEENN achieved the highest accuracy scores for RF (0.9643), XGB (0.9368), and KNN (0.9458). On the other hand, GANENN consistently delivered the best accuracy among the hybrid GAN models. When it comes to precision, both GANENN and SMOTEENN demonstrate superior performance across most classifiers. They take the lead in precision scores, highlighting their effectiveness in minimizing false positives. In recall, SMOTEENN stands

out with the highest scores for all classifiers, except for the LR model where GANENN outperforms it. This indicates that SMOTEENN is particularly adept at capturing instances of the minority class. Similar to recall, SMOTEENN dominates the F1 metric for all classifiers, except for LR, where GANENN achieves the highest score. This demonstrates the effectiveness of SMOTEENN in achieving a balance between precision and recall. In terms of ROC AUC, SMOTEENN maintains strong performance across classifiers, followed by GANENN and GANRUSBoost. This suggests that SMOTEENN provides an effective balance between true positive rate and false positive rate. SMOTEENN also outperforms other models in PR AUC, with GANENN and GANRUSBoost closely following. This indicates that SMOTEENN is effective in capturing positive instances, while minimizing false positives. Overall, the results demonstrate that SMOTEENN consistently outperforms GAN-based hybrid models across most classifiers and metrics, while GAN-based hybrid models prove their competitive performance and GANENN outperforms other GAN-based models.

Figure 9 presents the results on the financial fraud detection dataset which contains 6.9 million samples with only 0.13% identified as fraudulent. The results are visualized in Figure 10. The results reveal that overall, GAN-based models outperformed SMOTEENN and SMOTETomek. Specifically, GANBoost and GANENN performed better on DT, RF, XGB, and LGBM in terms of accuracy, and other GAN-based models performed competitively. For precision, GANBoost and GANENN also outperformed SMOTEENN and SMOTETomek on all classifiers. In recall, SMOTEENN and SMOTETomek continue to obtain higher scores across classifiers with a few instances where GAN-based models achieve similar or marginally better performance, such as GANENN in the LR classifier. F1 scores are highly competitive among all models. GAN-based models occasionally outperform SMOTEENN and SMOTETomek across the classifiers with GANENN and GANRUSBoost showing slightly better results in the LR classifier. When considering ROC AUC and PR AUC scores, GAN-based models are either on par with or surpass SMOTEENN and SMOTETomek in most cases across the classifiers. In particular, GANBoost, GANENN, GANRUS, and GANRUSBoost demonstrate prominent performance in the RF and XGB classifiers for both ROC AUC and PR AUC scores. Overall, the proposed GAN-based hybrid models demonstrate strong performance compared to SMOTEENN and SMOTETomek across the considered classifiers.

The results from both the datasets show that GAN-based hybrid models demonstrate competitive or superior performance when compared to well-established techniques such as SMOTEENN and SMOTETomek. In the hotel booking cancellation dataset where the minority class constitutes 37%, GAN-based hybrid models exhibit competitive performance across various metrics and classifiers, although they were slightly outperformed by SMOTEENN. In particular, GANENN shows promise among the proposed GAN-based models. On the other hand, in the financial fraud detection dataset with a highly imbalanced minority class of only 0.13%, GAN-based models consistently outperformed SMOTEENN and SMOTETomek in most metrics and classifiers, demonstrating their effectiveness in addressing a high degree of class imbalance in large-scale datasets. These results witness the potential of the proposed models, especially GANENN, as an efficient and competitive technique for handling a more significant class imbalance in different domains and across various classifiers.

Model	Metric	GANBoost	GANENN	GANRUS	GANRUSBoost	GANTomek	SMOTEENN	SMOTETomek
DT	accuracy	0.8435	0.8926	0.8436	0.8436	0.8523	**0.9415**	0.8585
	precision	0.8203	0.9074	0.8200	0.8200	0.8382	**0.9440**	0.8548
	recall	0.8279	0.9032	0.8287	0.8287	0.8418	**0.9501**	0.8636
	f1	0.8241	0.9053	0.8243	0.8243	0.8400	**0.9470**	0.8592
	roc_auc	0.8447	0.8906	0.8449	0.8449	0.8549	**0.9406**	0.8622
	pr_auc	0.7620	0.8756	0.7621	0.7621	0.7857	**0.9243**	0.8131
RF	accuracy	0.8825	0.9264	0.8825	0.8824	0.8896	**0.9643**	0.8970
	precision	0.9012	0.9472	0.9014	0.9013	0.9094	**0.9687**	0.9100
	recall	0.8251	0.9220	0.8248	0.8247	0.8444	**0.9663**	0.8811
	f1	0.8614	0.9344	0.8614	0.8613	0.8757	**0.9675**	0.8953
	roc_auc	0.9499	0.9776	0.9500	0.9500	0.9558	**0.9944**	0.9628
	pr_auc	0.9494	0.9844	0.9494	0.9494	0.9575	**0.9953**	0.9662
LR	accuracy	0.7865	0.7758	0.7909	0.7775	0.7719	**0.7979**	0.7490
	precision	0.8249	0.8073	**0.8449**	0.8112	0.8053	0.8232	0.7663
	recall	0.6575	0.7957	0.6466	0.6483	0.6658	**0.8055**	0.7166
	f1	0.7317	0.8015	0.7325	0.7207	0.7289	**0.8142**	0.7406
	roc_auc	0.8402	0.8589	0.8446	0.8288	0.8345	**0.8756**	0.8206
	pr_auc	0.8584	0.9114	0.8621	0.8481	0.8592	**0.9170**	0.8556
XGB	accuracy	0.8638	0.8989	0.8636	0.8630	0.8672	**0.9368**	0.8670
	precision	0.8943	0.9212	0.8957	0.8939	0.8985	**0.9433**	0.8906
	recall	0.7852	0.8990	0.7831	0.7836	0.8024	**0.9416**	0.8368
	f1	0.8362	0.9100	0.8356	0.8351	0.8477	**0.9425**	0.8628
	roc_auc	0.9355	0.9628	0.9353	0.9354	0.9399	**0.9842**	0.9420
	pr_auc	0.9358	0.9749	0.9358	0.9358	0.9432	**0.9878**	0.9493
KNN	accuracy	0.8016	0.8900	0.8016	0.8014	0.8096	**0.9458**	0.8114
	precision	0.7955	0.9193	0.7960	0.7955	0.8084	**0.9340**	0.7867
	recall	0.7429	0.8842	0.7423	0.7425	0.7689	**0.9700**	0.8543
	f1	0.7683	0.9014	0.7682	0.7680	0.7882	**0.9517**	0.8191
	roc_auc	0.8752	0.9506	0.8751	0.8752	0.8868	**0.9860**	0.8967
	pr_auc	0.8520	0.9545	0.8520	0.8520	0.8671	**0.9822**	0.8711
LGBM	accuracy	0.8585	0.8899	0.8587	0.8586	0.8611	**0.9218**	0.8545
	precision	0.9036	0.9182	0.9035	0.9037	0.9067	**0.9346**	0.8905
	recall	0.7616	0.8852	0.7623	0.7618	0.7786	**0.9224**	0.8083
	f1	0.8265	0.9014	0.8269	0.8267	0.8378	**0.9284**	0.8474
	roc_auc	0.9311	0.9570	0.9312	0.9313	0.9349	**0.9772**	0.9325
	pr_auc	0.9319	0.9710	0.9320	0.9320	0.9389	**0.9825**	0.9410

Fig. 7. Experiment Results on Hotel Booking Cancellation

7. Conclusion

We have presented a GAN-based hybrid approach to address class imbalance by combining GANs with undersampling/ensemble techniques such as Boost, ENN, RUS, RUSBoost, and Tomek Links. The use of a GAN enables the creation of diverse and realistic synthetic samples, while undersampling/ensemble techniques help remove noisy or ambiguous examples in the majority class. The evaluation on two datasets – hotel booking cancellation and financial fraud detection shows that the proposed effective when the level of class imbalance is high.

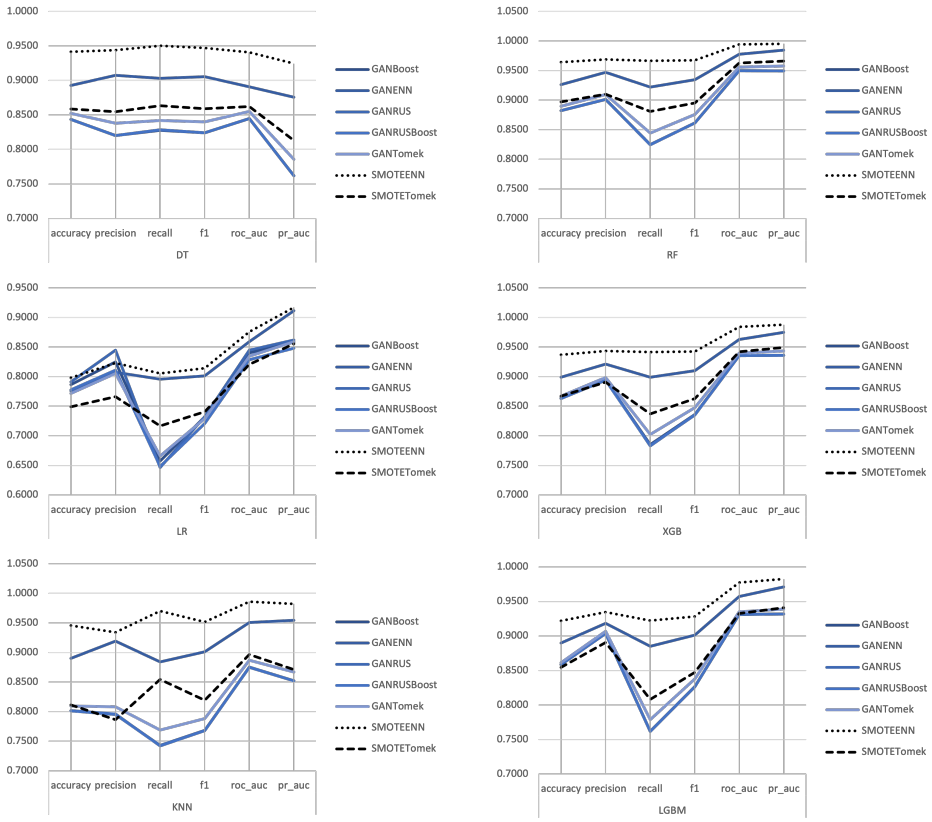


Fig. 8. Visualization of Experiment Results on Hotel Booking Cancellation

References

1. Ali-Gombe, A., Elyan, E.: MFC-GAN: Class-imbalanced dataset classification using Multiple Fake Class Generative Adversarial Network. Neurocomputing 361, 212–221 (2019)
2. Antonio, N., Almeida, A., Nunes, L.: Hotel booking demand datasets. Data in Brief 22, 41–49 (2019)
3. Antoniou, A., Storkey, A., Edwards, H.: Data Augmentation Generative Adversarial Networks. https://arxiv.org/abs/1711.04340 (2018)
4. Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter 6(1), 20–29 (2004)
5. Bekkar, M., Djemaa, H.K., , Alitouche, T.A.: Evaluation measures for models assessment over imbalanced data sets. Journal of Information Engineering and Applications 3 (2013)
6. Bhagwani, H., Agarwal, S., Kodipalli, A., Martis, R.J.: Targeting Class Imbalance Problem Using GAN. In: Proceeding of the 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT). pp. 318–322 (2021)

Model	Metric	GANBoost	GANENN	GANRUS	GANRUSBoost	GANTomek	SMOTEENN	SMOTETomek
DT	accuracy	**0.9997**	**0.9997**	0.9996	**0.9997**	0.9996	0.9994	0.9994
	precision	**0.9996**	0.9995	0.9995	0.9995	0.9995	0.9992	0.9990
	recall	0.9994	0.9995	0.9994	0.9994	0.9994	**0.9997**	**0.9997**
	f1	**0.9995**	**0.9995**	**0.9995**	**0.9995**	0.9994	0.9994	0.9994
	roc_auc	**0.9996**	**0.9996**	**0.9996**	**0.9996**	**0.9996**	0.9994	0.9994
	pr_auc	**0.9992**	**0.9992**	0.9991	**0.9992**	0.9990	0.9990	0.9989
RF	accuracy	**0.9998**	**0.9998**	**0.9998**	**0.9998**	0.9997	0.9997	0.9995
	precision	**1.0000**	**1.0000**	0.9999	**1.0000**	**1.0000**	0.9995	0.9991
	recall	0.9993	0.9994	0.9994	0.9993	0.9993	**0.9999**	**1.0000**
	f1	**0.9997**	**0.9997**	**0.9997**	**0.9997**	0.9996	**0.9997**	0.9995
	roc_auc	0.9999	**1.0000**	**1.0000**	**1.0000**	0.9999	**1.0000**	**1.0000**
	pr_auc	0.9999	**1.0000**	0.9999	0.9999	0.9999	**1.0000**	**1.0000**
LR	accuracy	0.9977	0.9976	**0.9985**	0.9984	0.8310	0.9407	0.9193
	precision	0.9949	0.9941	0.9971	**0.9974**	0.7015	0.9435	0.9176
	recall	0.9984	**0.9986**	0.9984	0.9979	0.5000	0.9379	0.9214
	f1	0.9966	0.9964	**0.9978**	0.9976	0.4986	0.9407	0.9195
	roc_auc	0.9975	0.9977	0.9985	0.9984	**0.9990**	0.9849	0.9732
	pr_auc	0.9799	0.9798	0.9878	0.9855	**0.9903**	0.9864	0.9765
XGB	accuracy	**0.9998**	**0.9998**	**0.9998**	**0.9998**	**0.9998**	0.9994	0.9993
	precision	**1.0000**	0.9999	0.9999	0.9999	0.9999	0.9990	0.9987
	recall	0.9995	0.9995	0.9995	0.9995	0.9994	0.9998	**1.0000**
	f1	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	0.9994	0.9993
	roc_auc	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
	pr_auc	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
KNN	accuracy	**0.9989**	0.9988	0.9988	0.9988	**0.9989**	0.9984	0.9982
	precision	**0.9997**	0.9995	0.9995	0.9995	0.9996	0.9991	0.9989
	recall	0.9982	0.9981	0.9981	0.9981	0.9982	0.9987	**0.9992**
	f1	0.9989	0.9988	0.9988	0.9988	0.9989	0.9989	**0.9991**
	roc_auc	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**	0.9998	0.9998
	pr_auc	**0.9999**	**0.9999**	**0.9999**	**0.9999**	**0.9999**	0.9998	0.9998
LGBM	accuracy	**0.9993**	0.9992	0.9992	0.9992	**0.9993**	0.9988	0.9986
	precision	**0.9999**	0.9998	0.9998	0.9998	**0.9999**	0.9995	0.9993
	recall	0.9987	0.9986	0.9986	0.9986	0.9987	0.9992	**0.9997**
	f1	0.9993	0.9992	0.9992	0.9992	0.9993	0.9993	**0.9995**
	roc_auc	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9999	0.9999
	pr_auc	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.9999	0.9999

Fig. 9. Experiment Results on Financial Fraud Detection

7. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Oversampling Technique. Journal Of Artificial Intelligence Research 16, 321–357 (2002)
8. Deng, G., Han, C., Dreossi, T., Lee, C., Matteson, D.S.: IB-GAN: A Unified Approach for Multivariate Time Series Classification under Class Imbalance. In: Proceedings of the SIAM International Conference on Data Mining (SDM). pp. 217–225 (2022)
9. Engelmann, J., Lessmann, S.: Conditional Wasserstein GAN-based oversampling of tabular data for Imbalanced Learning. Expert Systems With Applications 174 (2021)
10. Fiorea, U., Santisb, A.D., Perlaa, F., Zanettia, P., Palmieribi, F.: Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information sciences 479, 448–455 (2019)
11. Fotouhi, S., Asadi, S., Kattan, M.W.: A comprehensive data level analysis for cancer diagnosis on imbalanced data. Journal of Biomedical Informatics 90, 103089 (2019)
12. Freund, Y., Shapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Processings of the 2nd European Conference on Computational Learning Theory. pp. 23–37 (1995)

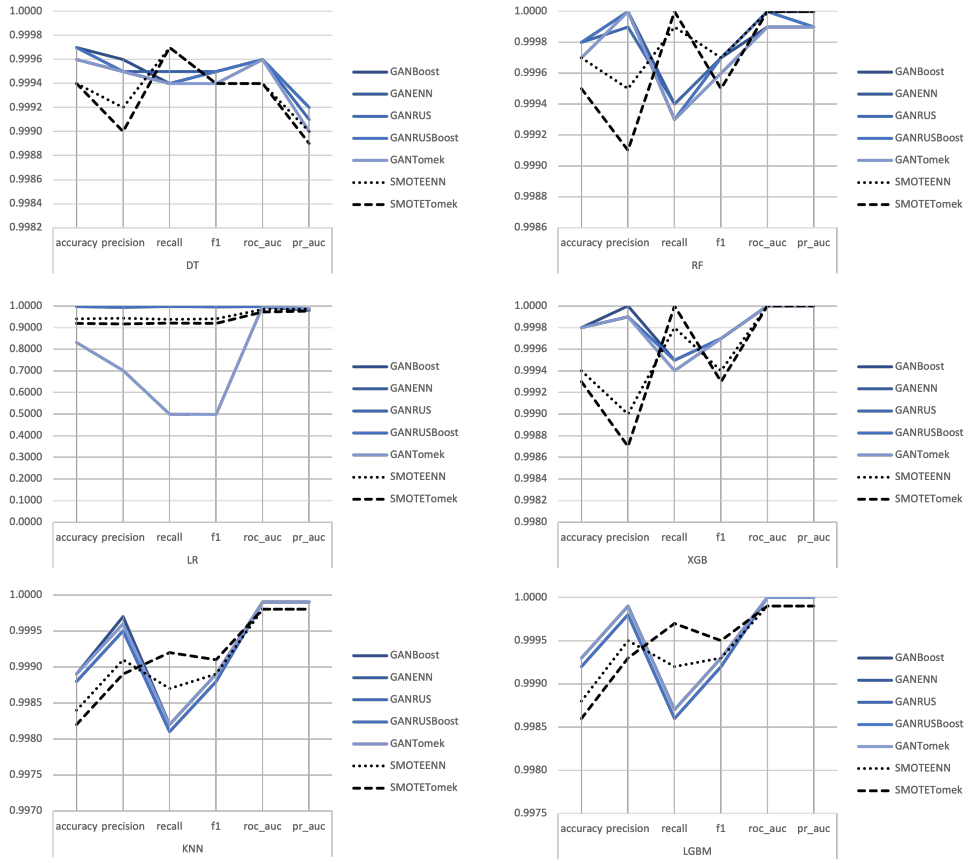


Fig. 10. Visualization of Experiment Results on Financial Fraud Detection

13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc. (2014)
14. Guo, X., Yin, Y., Dong, C., Yang, G., Guangtong, Z.: On the Class Imbalance Problem. In: Processings of the 4th International Conference on Natural Computation. pp. 192–201 (2008)
15. Japkowicz, N., Stephen, S.: The class imbalance problem: a systematic study. Intelligent Data Analysis 6, 429–449 (2002)
16. Jiang, W., Hong, Y., Zhou, B., He, X., Cheng, C.: A GAN-Based Anomaly Detection Approach for Imbalanced Industrial Time Series. IEEE Access 7, 143608–143619 (2019)
17. Kim, J., Jeong, K., Choi, H., Seo, K.: GAN-Based Anomaly Detection in Imbalance Problems. In: Proceedings of Computer Vision–ECCV Workshops: Part VI 16. pp. 128–145 (2020)
18. Lee, J., Park, K.: GAN-based imbalanced data intrusion detection system. Personal and Ubiquitous Computing 25, 121–128 (2021)
19. Lei, K., Xie1, Y., Zhong, S., Dai1, J., Yang, M., Shen, Y.: Generative adversarial fusion network for class imbalance credit scoring. Neural Computing and Applications 32, 8451–8462 (2020)

20. Liu, X., Zhou, Z., Wu, J.: Exploratory Undersampling for Class-Imbalance Learning. IEEE Transactions on Systems, Man, and Cybernetics 39, 539–550 (2009)
21. Lopez-Rojas, E.A., Elmir, A., Axelsson, S.: PaySim: A financial mobile money simulator for fraud detection. In: Proceedings of The 28th European Modeling and Simulation Symposium (2016)
22. Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., Malossi, C.: BAGAN: Data Augmentation with Balancing GAN. https://arxiv.org/abs/1803.09655 (2018)
23. Mullick, S.S., Datta, S., Das, S.: Generative Adversarial Minority Oversampling (2020)
24. Odena, A., Olah, C., Shlens, J.: Conditional Image Synthesis with Auxiliary Classifier GANs. In: International conference on machine learning. pp. 2642–2651 (2017)
25. Qasim, A.B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J.C., Sekuboyina, A., Kofler, F., Lipkova, J., Li, H., Menze, B.: Red-GAN: Attacking Class Imbalance via Conditioned Generation. Yet Another Medical Imaging Perspective. In: Proceedings of Medical imaging with deep learning. pp. 655–668 (2020)
26. Roth, K., Lucchi, A., Nowozin, S., Hofmann, T.: Stabilizing Training of Generative Adversarial Networks through Regularization. Advances in neural information processing systems 30 (2017)
27. Seiffert, C., Khoshgoftaar, T.M., Hulse, J.V., Napolitano, A.: Rusboost: A hybrid approach to alleviating class imbalance. Systems Man and Cybernetics Part A: Systems and Humans IEEE Transactions 40(1), 185–197 (2010)
28. Sharma, A., Singh, P.K., Chandra, R.: SMOTified-GAN for Class Imbalanced Pattern Classification Problems. IEEE Access 10, 30655–30665 (2022)
29. Tomek, I.: Two Modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics SMC-6, 769–772 (1976)
30. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 29, pp. 3630–3638. Curran Associates, Inc. (2016)
31. Wang, P., Li, S., Ye, F., Wang, Z., Zhang, M.: PacketCGAN: Exploratory Study of Class Imbalance for Encrypted Traffic Classification Using CGAN. In: Processings of IEEE International Conference on Communications. pp. 1–7 (2020)
32. Wang, Z., Wang, P., Zhou, X., Li, S., Zhang, M.: FLOWGAN:Unbalanced network encrypted traffic identification method based on GAN. In: Processings of IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking (2019)
33. Wilson, D.L.: Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. IEEE Transactions on Systems, Man, and Cybernetics 2(3), 408–421 (1972)
34. Yang, H., Zhou, Y.: IDA-GAN: A Novel Imbalanced Data Augmentation GAN. In: Proceedings of the 25th International Conference on Pattern Recognition (ICPR). pp. 8299–8305 (2021)
35. Zhenchuan, L., Mian, H., Guanjun, L., Changjun, J.: A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. Expert Systems with Applications 175, 114750 (2021)
36. Zheng, Z., Cai, Y., Li, Y.: Oversampling method for imbalanced classification. Computing and Informatics 34, 1017–1037 (2015)

Dae-Kyoo Kim is a professor in the Department of Computer Science and Engineering at Oakland University. He received a Ph.D. in computer science from Colorado State University in 2004. He worked as a technical specialist at the NASA Ames Research Center in 2002. His research interests include software engineering, software security,

data modeling in smart grids, and business process modeling. The author can be contacted at kim2@oakland.edu.

Yeasun K. Chung is Associate professor at Spears School of Business in Oklahoma State University. Prior to her academic career she worked in financial engineering and derivative department for a financial investment organization. Her research interests include risk management, strategic planning, business process management, and business intelligence. The author can be contacted at y.chung@okstate.edu

Received: November 13, 2023; Accepted: September 15 2024.

Computer Science and Information Systems 22(3):693–726 https://doi.org/10.2298/CSIS240228015J

Empirical Analysis of Python's Energy Impact: Evidence from Real Measurements

Elisa Jimenez, Alberto Gordillo, Coral Calero, Mª Ángeles Moraga, and Félix García

Instituto de Tecnologías y Sistemas de Información Camino de Moledores, s/n, 13005, Ciudad Real (Spain)
{elisa.jimenez,alberto.gordillo,coral.calero,mariaangeles.moraga,felix.garcia}@uclm.es

Abstract. Programming languages provide the notation for writing computer programs capable of granting our devices the desired functionalities. Even though they may seem intangible, the resulting programs involve an amount of energy consumption, which has an impact on the environment. Some studies on the consumption of programming languages indicate that while being one of the most widely used languages, Python is also one of the most demanding in terms of energy consumption. To provide developers using Python with a set of best practices on how to use it in the most energy-efficient way, this paper presents a study on whether the different ways of programming in Python have an impact on the energy consumption of the resulting programs. We have studied the relationship between Python's energy consumption and the fact that Python is a very versatile language that allows programs to be compiled and executed in many different ways. From the results obtained in our study, there seems to be a clear relationship between software energy consumption at runtime and: (1) the use of interpreted or compiled Python; (2) the use of dynamically or statically typed variables. Compiling Python code is a good option if it is done using the py_compile module. The use of interpreted code seems to decrease energy consumption over compiling using Nuitka. The use of dynamically typed variables seems to decrease considerably the graphics and processor energy consumption. In addition, we have observed that energy consumption is not always directly related to execution time. Sometimes, more power in less time increases consumption, due to the power required.

Keywords: python, efficiency, programming languages, green software

1. Introduction

In addition to serve as a means of communication and entertainment, technological devices have also become a great working tool, being almost impossible to imagine life without them. The functionalities of technological devices are enabled by software applications, which are often developed using multiple programming languages, although one language typically predominates.

In general terms, a programming language provides a structured way to express algorithms and instructions to create a software program capable of executing one or more functionalities. Obviously, it is possible to implement the same functionality using different programming languages.

In fact, the big amount of available programming languages means that developers should choose one of them to implement the software programs. As [46] points out, choosing the appropriate programming language can make or break a project.

There are several criteria that could be used to do this selection. For example [46] mentions programmer productivity, maintainability, efficiency, portability, tool support, and software and hardware interfaces as key factors, but also indicates that, depending on the type of code to develop, there is little room for choice. In [5] the authors propose a model to select the best programming language to be learned by novice programmers for Data Analytics Applications based on eight criteria: popularity, data analytics support, volume of data it can handle, speed of compiling, expressiveness, dreadfulness, programmers' recommendations and average reasonable financial cost. [32] present the results of a quantitative study about the language adoption process, identifying the availability of open-source libraries, existing code, and experience as the most influential factors. In contrast, intrinsic factors, such as a language's simplicity or safety, rank low.

These are just some examples but, as can be observed, there is no a standard set of aspects to be considered as important when choosing a programming language. Because, as remarked by [34], the nature of languages as a special software tool makes it difficult to find measures to draw objective conclusions about them.
Besides this lack of consensus on which are the best criteria to choose a programming language, it can be noted that energy efficiency does not appear as a key aspect to be considered.

However, nowadays, the energy consumption of IT (Information Technologies), including software, is becoming a concern. According to the BEREC report [12], some 2 to 4% of greenhouse gases currently come from the digital industry. Moreover, new programming trends, such as big data or Artificial Intelligence, could further increase these figures.

As remarked by [13] the existing studies that analyze the impact of choice of programming language suffer from several deficiencies with respect to methodology and the applications they consider. In [34] is indicated that most of the claims about programming languages are based on personal affinity, being the empirical comparison a good approach to provide objective information about languages. Looking into programming languages from an empirical perspective would provide supportive evidence and valuable conclusions about them.

The objective of this paper is to conduct an empirical study, which can be considered a benchmark study according to the Empirical Standard of ACM [42] (hereafter referred to as a study), to analyse the energy consumption of a programming language. In particular, the study focuses on Python's energy consumption. Python is a highly versatile programming language and is one of the most widely used languages in the current era. According to the PYPL (Popularity of Programming Language Index), [15], created by analysing how often language tutorials are searched on Google, the most popular programming language in 2023 compared to a year ago is Python, with a 27.27% increase, followed by Java, with a 16.35% increase (results of May 2023). Moreover, it is indicated that Python grew the most in the last 5 years (3.5%) [15]. Also the last updates of other rankings such as the IEEE Spectrum's Ranking of the Top Programming Languages 2022 [16] (a ranking based on nine metrics to know what languages the public is programming in) or the Tiobe Index [4] (which takes data from hundreds of different sources, compiles it, and dumps it into a list), rank Python at the top of the list.

However, according to [36] Python is one of the most energy-demanding programming languages. An analysis of the possible causes of the high consumption of Python

leads to the conclusion that it could be due to the fact that: (1) as indicated in [45], Python is a dynamically typed language or (2) the fact that Python code is typically executed without prior compilation. Therefore, we decided to study the impact of these two aspects on the energy consumption of Python to offer developers the most efficient way to use this language.

The remainder of this document is organised as follows: Section 2 presents the related work and the issues motivating this paper. Section 3 describes the background of this study, including some relevant aspects about the execution of programs written in Python language and the applied methodology to carry out the study. Section 4 contains the analysis and discussion of the results obtained, in which we analyse the difference in consumption between the different test cases to answer the research questions in Section 5. In Section 6, we present the threats to the validity of our study. Finally, Section 7 presents the conclusions of the study and gives some recommendations on the use of Python.

2. Related work

Programming languages have been studied and analysed from several points of view. For example, there are many studies that relate programming languages to the improvement of the developers' productivity from different points of view.

In [38] an experiment was conducted to compare programmer productivity and defect rates for Java and C++. They concluded that a typical C++ program had two to three times more bugs per line of code than a typical Java program. C++ also generated between 15 per cent and 50 per cent more defects per line, and perhaps took six times longer to debug. When comparing defects against development time, Java and C++ showed no difference, but C++ had two to three times more bugs per hour.

In [18] authors tried to test the Brooks assumption that annual lines-of-code programmer productivity is constant, independent of programming language used. They analysed 10 of the most popular programming languages in use in the open-source community, concluding that the programming language is a significant factor in determining the rate at which source code is written.

In an effort to study the effects of programming language fragmentation on productivity—and ultimately on a developer's problem-solving abilities—in [25] the authors presented a metric, namely language entropy, for characterising the distribution of a developer's programming efforts across multiple programming languages. They concluded that changes in language fragmentation affect a programmer working within a single paradigm less than a programmer working with multiple paradigms.

The objective of [26] is to identify how different programming languages may affect software development productivity. Each programming language has its own productivity level. The productivity of new development projects seems to be influenced by the programming language used, while the productivity of enhancement projects seems to be much less dependent on their specific programming language.

In [13] the authors propose a novel methodology which controls the development process and developer competence and quantifies how the choice of programming language impacts software quality and developer productivity. After conducting a study and statistical analysis on a set of long-running open-source projects written mainly in C and

C++ (Firefox, Blender, VLC, and MySQL), they found that the use of C++ instead of C improves software quality and reduces maintenance effort.

Finally, [34] presents a study to investigate the impact of high-level, general-purpose, programming languages on software development productivity and quality. The authors analyse 11 primary languages: JavaScript, Java, Python, Go, Objective-C, Swift, PHP, Ruby, C#, C++, and C. The conclusion of the study is that the choice of programming language can affect the development process.

Focusing on the energy consumption of programming languages, [6], studied quantitatively the impact of languages (C/C++/Java/Python), compiler optimization (GNU C/C++ compiler with O1, O2, and O3 flags) and implementation choices (e.g. using malloc instead of new to create dynamic arrays and using vector vs. array for Quicksort) on the energy-efficiency of three well-known programs: Fast Fourier Transform, Linked List Insertion/Deletion and Quicksort. Experiments showed that by carefully selecting an appropriate language, optimisation flag and data structure, a significant amount of energy can be conserved to solve the same problem with identical input size.

In [7] three metrics are proposed to categorize software implementation and optimization efficiency: Greenup, Powerup, and Speedup metrics (GPS-UP). GPS-UP metrics transform the performance, power, and energy of a program into a point on the GPS-UP software energy efficiency quadrant graph. In addition, eight categories of possible software optimisation scenarios (four energy-saving and four energy-wasting) are presented with examples on how to obtain them and the new metrics are compared with existing metrics such as the Energy Delay Product (EDP).

Connolly Bree and Ó Cinnéide [17] conducted an assessment on the impact of two popular design-level refactoring on energy consumption in the Java programming language. Specifically, they focused on the refactoring techniques of replacing Inheritance with Delegation and vice versa. The researchers assessed the energy consumption by running code snippets for both refactoring and measuring average power consumption and energy consumption. The study revealed that Inheritance proved to be more efficient than Delegation. It exhibited a 77% reduction in runtime and a 4% decrease in average power consumption when compared to Delegation. However, a significant limitation of the study was the experiments were conducted in an Interpreted mode, which does not accurately reflect real-life scenarios where Just-in-Time (JIT) enabled compilers are commonly utilized.

Pinto et al. [39] explore the energy efficiency of several Java Collection implementations, beyond their well-established characteristics in terms of performance, scalability, and thread-safety. The study involves 16 collection implementations (13 thread-safe, 3 non-thread-safe) categorized into lists, sets, and mappings. The research reveals that design decisions significantly influence energy consumption. Notably, adopting a newer hashtable version can result in a 2.19x energy savings in micro-benchmarks and up to 17% in real-world benchmarks compared to older associative implementations.

Also Pereira et al. [37] study the energy efficiency of Java Collections. They propose an approach to energy-aware development that combines application-independent energy profiling of Java Collections and static analysis to estimate the system's utilization of these collections and its intensity. The results indicate that some widely used collections, e.g. ArrayList, HashMap and Hashtable, are not energy efficient and should sometimes be avoided when energy consumption is a major concern.

Calero et al. [14] focus their efforts on investigating the suitability in the development of software applications in terms of energy consumption of Spring (a framework for the development of Java applications), and its conclusions point out that code developed using Spring require much more energy than those developed without Spring.

Finally, Lima et al. [27] investigated the energy behavior of programs written in Haskell. They conducted two in-depth and complementary studies to analyze the energy efficiency of programs from two different perspectives: strictness and concurrency. They found that making small changes can make a big difference. In one benchmark, under a specific configuration, choosing the use of the MVar (Mutable Variable) data sharing primitive instead of the TMVar (Transactional Mutable Variable), can result in up to 60% energy savings. In another benchmark, using TMVar instead of MVar can yield up to 30% energy savings.

To provide information about the differences in energy consumption of several programming languages Pereira et al. [36] conducted an investigation in which they analysed the energy behavior of twenty-seven programming languages, estimating and comparing the consumption required for the execution of ten different programs written in all of the selected programming languages. The 27 programming languages included compiled, interpreted, and virtual machine languages. As a result of the study, it was found that compiled languages tend to be, as expected, the fastest and most energy efficient. On average, virtual machine languages required almost five times more energy to execute the solutions than compiled languages. On the other hand, interpreted languages required almost 20 times more energy than compiled languages.

Some works specifically address the energy efficiency of the Python programming language. However, the existing literature mostly focuses on aspects such as performance, complexity, and optimisations, largely neglecting the crucial aspect of energy consumption. For example, in the work conducted by Redondo and Ortin [43] a meticulous evaluation of seven implementations of Python versions 2 and 3 is presented. Their aim is to assist in the selection of a suitable implementation by running 523 programs to each version. The evaluation encompasses runtime performance, memory consumption, and an exploration of significant qualitative characteristics inherent in each implementation. One of their main conclusions is that interpreter-based implementations (such as CPython) are the most energy-efficient, followed by statically compiled implementations of Python. In contrast, JIT-compiled approaches are found to be the least energy-efficient.

So far, work in the Python domain has focused predominantly on aspects such as performance, complexity, and optimisation, while energy efficiency has been neglected. The study developed by Reya [44] explores the energy efficiency of some coding patterns and techniques in Python, with the goal of guiding programmers towards more informed and energy-conscious coding practices. The research analyzes the energy consumption of a wide range of topics, such as data initialization, access patterns, structures, string formatting, sorting algorithms, dynamic programming, and performance comparisons between NumPy and Pandas, and personal computers versus cloud computing. The comparisons they present are very interesting and can offer programmers good practice in the use of Python.

Table 1 summarises the related work together with the method used to perform the energy consumption measurements.

Table 1. Summary of related works

Reference	Research goal and scope	Measurement method
[6]	Energy impact of the languages C, C++, Java, and Python based on the different implementations and compiler optimizations.	Software estimation. Intel Power Governor library (based on RAPL, Intel's Runtime Average Power Limit) estimates the energy consumption of CPU and DRAM power when implementing a few algorithms.
[14]	Execution time and energy consumption required by three applications, developed with and without Spring.	Hardware measurement. FEETINGS (Framework for Energy Efficiency Testing to Improvement eNviromental Goals of the Software), a specific framework for measuring software energy consumption, is used together with the EET (Energy Efficiency Tester) hardware measuring instrument.
[17]	How redundancy in an object-oriented design can contribute to unnecessary energy consumption and determine how software refactoring can eliminate this redundancy.	Hardware measurement. A Watts Up Pro power meter was used to record power consumption every second.
[27]	Energy behavior of programs written in Haskell. To do so, they make changes to benchmarks such as MVar and TMVar.	Software estimation. RAPL is used to collect processor energy information and extend two existing Haskell performance analysis tools (Criterion and GHC Profiler).
[36]	Power consumption of several programming languages of different types (interpreted, VM...) by using a set of benchmarks of different functionalities belonging to CBLG.	Software estimation. Intel's RAPL is used, which collects and analyzes and analyzes the resulting data on execution time execution time and memory usage.
[37]	Energy consumption of different Java Collection Framework (JFC) implementations. With the data obtained, they present an energy optimization approach for Java programs based on the calls to JFC methods in the source code of a program.	Software Estimation. To record CPU power consumption measurements, jRAPL

[39]	Energy efficiency of 16 implementations of Java collections grouped into 3 types (lists, sets, and allocations) to demonstrate that design decisions can greatly affect energy consumption.	Hardware measurement. The first type of architecture is measured using current meters across power supply lines to the CPU module. Software estimation. The second type of architecture energy values were estimated with jRAPL (framework for profiling Java programs using RAPL).
[43]	Runtime performance and memory consumption of seven language implementations of Python versions 2 and 3 to facilitate the selection of one of them.	Not specified. The methodology "Statistically Rigorous Java Performance Evaluation" is used to statistically analyze start-up and steady-state performance data. The work does not specify how the times are obtained.
[44]	Energy efficiency of various coding patterns and techniques in Python, with the objective of guiding programmers to a more informed and energy-conscious coding practices.	Software estimation. Intel's Power Gadget is used, which is a software-based tool for tracking power usage that is compatible with Intel Core i5 processors.

Two thirds of the studies use software estimation methods and therefore relatively few studies obtain more realistic measures of consumption using hardware devices. In this study we also try to contribute in this direction by providing real consumption measurements.

3. Background

This section provides an overview of the methods for executing Python programs and the methodology used in this study.

3.1. Python Execution Methods

Python generally uses dynamically typed variables, meaning that a variable can change its type during the lifetime of a program [33]. Previous studies have highlighted that this dynamic typing can impact performance, as the lack of compile-time type information reduces opportunities for compiler optimizations, and additional type checking at runtime can incur performance costs [43].

In addition to this inherent characteristic of Python, there are several methods for executing Python programs, each with its own characteristics and advantages in terms of performance and energy efficiency. These methods are described below.

3.1.1 Interpreted code Programs written in Python are generally interpreted by a specific implementation of the language, such as CPython, Jython, or IronPython [3]. In the case of CPython, the Python source code is first compiled to an intermediate format called

bytecode, which is closer to machine language but still independent of the processor architecture. This bytecode is then interpreted by the Python virtual machine.

The py_compile module [2] in CPython is used to precompile Python source files (.py) into bytecode files (.pyc). This precompilation process converts the source code into an intermediate bytecode form before execution, which helps reduce the time required to start the program. However, even though the bytecode is precompiled, its execution is still carried out by the CPython virtual machine, which interprets the bytecode at runtime. This distinction highlights that the precompilation occurs before the program's execution (at compile time), unlike Just-In-Time (JIT) compilation, which compiles code during program execution.

3.1.2 Just-In-Time (JIT) Compilation Just-In-Time (JIT) compilation [11] is a technique used to improve performance during the execution of interpreted programs. Instead of interpreting the bytecode every time it runs, JIT compiles parts of the bytecode into native machine code at runtime, which can result in faster execution.

GraalPy [9] is an implementation of Python on the GraalVM platform, which provides high-performance execution through JIT compilation. GraalPy utilizes the advanced JIT capabilities of GraalVM to dynamically compile Python code to native machine code during execution, resulting in significant performance improvements. GraalVM applies aggressive optimizations during JIT compilation, making GraalPy a powerful tool for executing Python code efficiently. Therefore, GraalPy combines features of an interpreter and a JIT compiler. It interprets and executes Python code in a manner similar to a traditional interpreter, but also performs JIT compilation to improve performance during program execution.

3.1.3 Ahead-Of-Time (AOT) Compilation Ahead-Of-Time (AOT) [10] compilation is a method where the source code is compiled directly into native machine code before execution. This process is completed during the build time, resulting in an executable that does not require further compilation or interpretation at runtime.

Nuitka [1] is a tool that compiles Python programs to C and then uses a C compiler to generate native machine code. This process is a form of Ahead-Of-Time (AOT) compilation, as it converts Python source code into an executable that can be run directly by the operating system without requiring Python to be installed. Nuitka eliminates the need to interpret bytecode at runtime and can offer significant advantages in terms of execution speed and energy efficiency. While the generated executables may still depend on certain Python libraries at runtime, Nuitka's approach aligns closely with the principles of AOT compilation, producing efficient and standalone executables.

GraalPy [9], is primarily considered a Just-In-Time (JIT) compiler but also supports AOT compilation. This feature allows GraalPy to compile Python scripts into native binaries, leveraging the performance optimizations provided by GraalVM. GraalPy's ability to perform AOT compilation can be utilized through the creation of native images [8], which are standalone executables generated by the native-image tool.

Fig. 1 also represents the cycle of a program written in Python, from source to machine code.

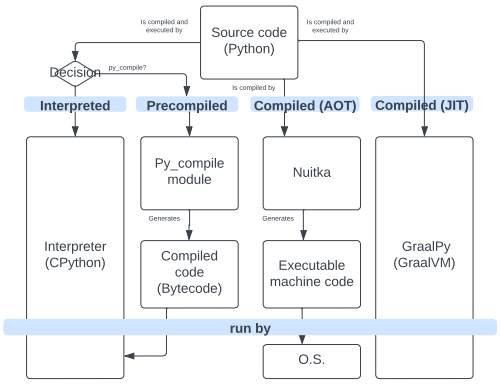


Fig. 1. Python code cycle

Although GraalPy presents potential benefits, during our study, we encountered multiple errors while executing several codes using the GraalPy compiler, likely due to the fact that GraalPy is still under development. As a result, it was not feasible to include it in our evaluation.

3.2. Green Software Measurement Process (GSMP)

As indicated by [50] Software Engineering requires a specific process for conducting experiments, as other sciences and engineering disciplines. For this reason, we have applied FEETINGS in this study, which has a methodological component, named GSMP (Green Software Measurement Process) [30] including all the activities and roles necessary to perform the measurement and analysis of the energy consumption of software, ensuring the reliability and consistency of the measurements. GSMP [29] is composed of seven phases (see Fig. 2). In a nutshell, the initial phase focuses primarily on the definition of the requirements and the software system to be evaluated. The next two phases focus on the configuration and preparation of the measurement environment. In phase four, energy consumption measurement activities are carried out. Finally, the last phases are the analysis and reporting of the data obtained. GSMP is intended to be performed in an iterative way, so the phases are interrelated to each other.

In addition to the methodological component, FEETINGS [30] has two other components: a conceptual component (GSMO-Green Software Measurement Ontology, with the terminology related to the measurement of software energy consumption) and a technological component composed by EET (Energy Efficiency Tester) [28] and ELLIOT [19]. EET is a hardware device built to capture the energy consumption of the hard disk, graphic card, and processor, as well as the overall energy consumption of the computer (namely DUT-Device Under Test) when running software. The data captured by EET are analysed by the ELLIOT tool.

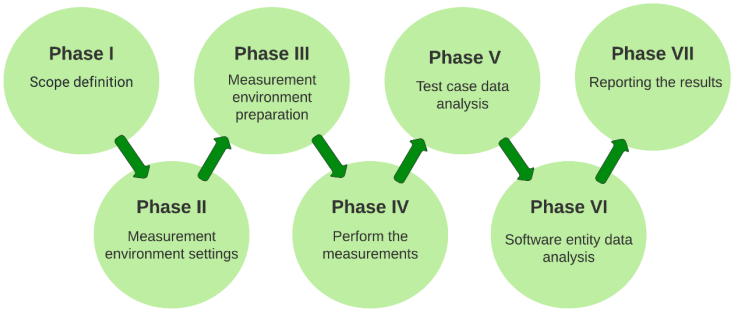


Fig. 2. GSMP phases for evaluating the energy efficiency of a software

4. Python energy consumption study

With the above considerations in mind, our research aims to quantify the energy savings achieved by statically declaring variable types. In addition, we extend our research to different Python execution methods to study the impact on consumption as well. With the results obtained, we intend to offer a set of recommendations to Python programmers on how to make better use of this programming language from an energy efficiency point of view.
Table 2 shows the research questions together with their motivation.

Table 2. Research Questions and Their Motivation

Research question	**Motivation**
RQ1. Is there any relationship between the energy consumption (by a software application at runtime) and the way it is executed?	With this research question, we want to check the difference in energy consumption between the different ways of executing programs written in Python (interpreted or compiled), in order to offer programmers some recommendations on the most efficient way to use this language.
RQ2. Is there any relationship between the energy consumption and time required (by a software application at runtime) and the way the variables are typed in the language used?	The most common use of the Python language involves the use of dynamic variables. Therefore, with this question, we want to determine whether statically declaring variables could improve the energy consumption of the language.

4.1. Application of GSMP to this study

Table 3 summarises the application of the GSMP phases and activities to our study.

Table 3. GSMP phases summary

Phase	**Application**
I. Scope Definition	– **Software Entity Class:** Python programming language – **Software Entities (SE):** Interpreted & Dinamically Typed Variables (DTV), Interpreted & Statically Typed Variables (STV), Py_compile & DTV, Py_compile & STV, Nuitka & DTV and Nuitka & STV. – **Test cases:** Ten algorithms of the Computer Language Benchmarks Game (CLBG): Binary trees, Fannkuch-redux, Fasta, Mandelbrot, K-nucleotide, N-body, Pi-digits, Reverse-complement, Regex-redux and Spectral-norm. – **Run test cases:** Each algorithm in the different ways of executing Python language.
II. Measurement Environment Setting	– **Hardware measuring instrument:** EET (Energy Efficiency Tester) – **Device Under Test (DUT):** Monitor: Philips 170s6fs LCD Motherboard: ASUS Prime B460-Plus Processor: Intel i7 10700 2900MHz RAM: 2 modules of 16GB Kingston Hiperx Fury DDR4 Graphics card: Sapphire ATI Radeon X1950 GT, 256mb RAM DDR3 Hard disk: Western Digital Blue 500GB SSD Power supply: 360 PS5805 – 580W O.S.: Gnu/Linux Ubuntu 20.4 LTS. – **Measures:** Execution time. DUT Energy Consumption. Processor Energy Consumption. Graphics Card Energy Consumption.
III. Measurement Environment Preparation	– **Before starting the measurements:** Install the Nuitka compiler for Python 3.11. – **For each Python execution way under study:** Clean the DUT, Check that there is not any software running in the background.

IV. Performing the measurement	– **For Python interpreted:** Execute algorithms using CPython interpreter. – **For py_compile:** Compile with py_compile. Execute the compiled algorithms using CPython interpreter. Delete _pycache_ folder between measurements. – **For Nuitka:** Compile with Nuitka and execute the algorithms.
Phases III and IV are repeated for each algorithm.	
V. Test Case Data Analysis	Analyse the energy consumption data for each test case. Check that the measurements are correct (outliers, wrong executions and so on) and eliminate the wrong measures if it is necessary.
VI. Software Entity Data Analysis	– **For each SE:** Calculate the mean of the energy consumption for each algorithm and for each component (DUT, processor, and graphic card). Calculate the mean of the energy consumption for each SE considering the mean of energy consumption of all the algorithms. State conclusions (see “Results” section).
VII. Reporting the result	This paper.

In the first phase of GSMP the scope of the study must be defined (see Fig. 2). As we mentioned previously, the purpose of this study was to examine programs written in Python to determine whether different execution forms and programming approaches (dynamically typed variables (DTV) or statically typed variables (STV)) have an impact on the energy consumption required to run the resulting application.

As modes of execution, we have selected CPython, the py_compile module and the Nuitka compiler for several reasons that contribute to the richness and representativeness of our study, which are described below:

- CPython, the reference implementation of Python [23], was chosen because of its wide adoption and widespread use in the Python development community. As the standard implementation, CPython provides a representative perspective of how Python programs run in most production environments.
- The py_compile module is also an essential tool in the Python development environment, as it is used to speed up the execution of programs. The inclusion of py_compile

in our measurement allows us to explore the energy impact associated with the compilation phase.
- The inclusion of the Nuitka compiler adds an interesting dimension to our research as it directly converts Python code to machine code through Ahead-Of-Time (AOT) compilation, thus avoiding the use of CPython for execution. This enables us to evaluate the energy efficiency and performance benefits of generating native executables.

As mentioned in the previous section, it was impossible to include the GraalPy compiler in our study. However, measurements have been performed with GraalPy and the results are available in the study repository at [22].

By comparing the energy consumption between CPython, py_compile and Nuitka, we can assess how choices in execution tools and technologies impact power consumption, thus providing valuable information for developers looking to optimise their processes. The aspects under study according to the research questions, result on eight different combinations (SEs), shown in Table 4.

Table 4. Combinations to be studied and compared respect to their energy consumption

	DTV	**STV**
Interpreted (CPython)	SE1	SE4
Pre-compiled (Py-compile)	SE2	SE5
Compiled (Nuitka)	SE3	SE6

SE1 is the 'usual' way to use Python. The CPython interpreter compiles Python code to bytecode before executing it, but it does not use Just-In-Time (JIT) compilation. CPython follows a traditional rendering approach, so it did not require any additional action beyond executing the code. Similarly, SE5 did not require any additional action to execute, but it was necessary to adapt the algorithms to declare variables statically.

Similarly, SE4 did not require any additional action for its execution, but it was necessary to adapt the algorithms to declare the variables statically. For SE2 and SE5, each algorithm was pre-compiled using the py_compile command and then the resulting file was executed using the CPython interpreter. The bytecode code is stored in a folder called 'pycache', which was deleted between measurements so as not to affect the results of the experiment.

For SE3 and SE6, each of the algorithms was compiled using Nuitka [1] and then run without using CPython.

Ten algorithms written in Python (Binary-trees, Fannkuch, Fasta, Mandelbrot, K-nucleotide, N-body, Pidigits, Reverse-complement, Regex-redux and Spectral-norm) selected from "The Computer Language Benchmarks Game" [20] were used to measure energy consumption. The CLBG Initiative has developed a framework for testing and comparison of multiple programming languages using a collection of general programming problems. Although there is a specific tool for performing experiments in Python called "The Python Performance Benchmark Suite" [47], we have considered performing our experiments using CLBG because it was the one used by [36] where Python resulted

as the most energy consumer.

Table 5 shows the selected algorithms together with their description and the size of the input used for their execution.

Table 5. Algorithm for test cases

Algorithm	Description	Data size
Binary-trees	Allocate and deallocate many binary trees.	21
Fannkuch-redux	Indexed-access to tiny integer-sequence.	12
Fasta	Generate and write random DNA sequences.	25000000
K-nucleotide	Hashtable update and k-nucleotide strings.	25000000
Mandelbrot	Generates a Mandelbrot set	16000
N-body	Double-precision N-body simulation.	50000000
Pi-digits	Calculates all digits in pi till the nth position.	10000
Regex-redux	Match DNA 8-mers and substitute magic patterns.	5000000
Reverse-complement	Converts a DNA sequence into its reverse-complement.	25000000
Spectral norm	Eigenvalue using the power method.	5500

The 10 algorithms chosen will be implemented with dynamically (as usual in Python) or statically typed variables. Moreover, these implementations will be executed in an interpreted (CPython), precompiled (Py_compile) and compile way (Nuitka). The experiment will be performed on the most updated CPython version to date, 3.11 and the compilation with Nuitka will be done using the –standalone option [40]

The software entity class was defined as the Python programming language, the software entities are the already described forms of execution, the test cases were the ten algorithms selected and the run test cases is defined as the combination of each way of executing and programming Python for each algorithm.

To answer the research questions, we will measure the energy consumption for each one of the combinations (see Table 4), what means a total of 80 data sets of energy consumption (10 algorithms vs. 8 SEs).

As a result of the second phase of the GSMP process, we selected EET [28] as the measuring instrument (the technological component of FEETINGS) and defined the specification of the Device Under Test (DUT) where the test cases were executed.

From the measurements provided by EET, we would analyse the ones of the DUT (i.e. the energy consumption of the entire PC), the graphics card and the processor. The data samples obtained by EET will be analysed with ELLIOT.

In the third phase of the process, the measurement environment was prepared. The versions of Python, and Nuitka needed to carry out our study were installed, and between measurements, other specific actions were applied.

In the fourth phase, measurements are carried out in EET. Each run test case corresponds to the execution of an algorithm on a SE given. And each test case run was repeated 30 times, our decision regarding the number of measurements is based on the recommendation of authors such as [24] for evaluations of software power consumption in a controlled environment. Generally, a sample of 30 measurements is sufficient for the

analysis of each intended test case, as it tends to produce a near-normal sampling distribution.

It is worth emphasizing that EET obtains 100 samples (instantaneous power values) per second, resulting in many values per test case that must be managed and analysed by ELLIOT. As an example, and to facilitate the reader's understanding, for this study, between 6800 and 1,435,535 samples (depending on the runtime of each algorithm) of instantaneous processor power have been obtained for the run test cases. Therefore, for each measurement, the average value of the 30 runs of the algorithm is obtained and the resulting data sets are analysed using ELLIOT being then possible to interpret them according to the test cases defined, to, at the end, answer the research questions (Phases V and VI). As a final remark, the consumption data analysed are the ones obtained after subtracting the baseline, i.e., the consumption of the operating system and the hardware devices in the background. All the information and data about the study can be found at [22].

5. Results and Discussion

Table 6 shows the time (s) and energy consumption (J) results for each one of the ten algorithms for Software Entities SE1, SE2 and SE3. Table 7 shows the corresponding to SE4, SE5 and SE6 (see Table 4 for details of the SEs).

Table 6. SE1-3 time and consumption results

	SE 1		SE 2		SE 3	
Algorithm	Time (s)	DUT (J)	Time (s)	DUT (J)	Time (s)	DUT (J)
Binary-trees	13.3242	2985.7902	12.9068	2860.2744	15.8172	3357.8365
Fannkuch-redux	99.5651	19140,0541	99.1313	19806.2811	107.9207	21544.9500
Fasta	29.1941	2480.1973	29.2924	2456.5435	35.9573	3312.2393
K-nucleotide	14.0083	3230.7826	13.7028	3370.1803	17.5856	3973.8959
Mandelbrot	57.0628	11576.6236	57.0012	11201.6928	53.0170	10462.8838
N-body	182.1192	17957.4216	180.9061	17493.4335	278.6027	29050.4140
Pi-digits	310.0411	30235.2839	312.7975	27114.8860	312.6554	28845.3751
Regex-redux	4.7917	549.2286	4.7883	554.7591	4.7883	563.2351
Reverse-complement	2.0077	101.4849	2.1446	90.8980	2.1456	107.6565
Spectral-norm	48.2717	9627.9359	47.9484	9084.3719	61.8417	11750.8619

To answer the research questions proposed, the results of the possible combinations of the SE will be compared. The idea is to analyse the influence of each factor separately. Therefore, nine different comparisons will be presented, as shown in Fig. 3.

To help to better interpret the comparisons, for each case, we will use the SEn with the best average results in terms of energy consumption for most of the 10 algorithms as basis and then we will calculate the percentage of increase on the energy consumption required by the other SEn of the comparison.

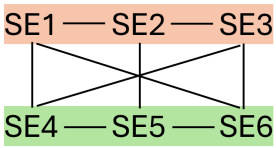


Fig. 3. Comparisons analysed

Table 7. SE4-6 time and consumption results

	SE 4		SE 5		SE 6	
Algorithm	Time (s)	DUT (J)	Time (s)	DUT (J)	Time (s)	DUT (J)
Binary-trees	13.3137	28549624	13.1434	2854.9970	15.6363	3293,1061
Fannkuch-redux	99.5693	1979.4,1939	99.7754	20306.4685	108.3222	21417,3473
Fasta	28.7633	2858.1650	29.3463	2522.0431	36.5822	3202,7499
K-nucleotide	13.8507	3309.3222	13.7778	3283.5624	17.5838	4049,7263
Mandelbrot	56.9158	11374.2202	56.9270	11493.3948	51.1548	10419,3156
N-body	182.1336	16834.5039	180.7359	15494.6057	283.2040	30017,9264
Pi-digts	310.1995	28704.6007	311.6585	27360.2596	312.8101	28983,7958
Regex-redux	4.7935	535.7790	4.7930	536.2314	4.7899	569,0573
Reverse-complement	2.1445	108.2303	2.1445	89.5619	2.1428	94,4505
Spectral-norm	50.8669	9316.0233	51.0056	9977.2440	58.9530	11701,2531

5.1. SE1, SE2 and SE3: Python Interpreted, compiled (py_compile) and Compiled (Nuitka) with DTV

This comparison aims to study the influence on consumption of the way in which the code to be executed is obtained when the variables are dynamically typed. In relation to the power consumption of the DUT, the best SE is SE2 (compiled (py_compile) and DTV), therefore, Table 8 shows the relative percentage increase in power consumption of SE1 (interpreted and DTV) and SE3 (compiled (Nuitka) and DTV) with respect to SE2.

Table 8. Time and energy consumption comparison: interpreted/compiled (Nuitka) vs. compiled (py_compile); with DTV

	SE1 vs. SE2		SE3 vs. SE2	
Algorithms	Time (%)	Energy consumption of DUT (%)	Time (%)	Energy consumption of DUT (%)
Binary-trees	3.2338	4.3882	22.5497	17.3956
Fannkuch-redux	0.4376	-3.3637	8.8665	8.7784
Fasta	-0.3357	0.9629	22.7529	34.8333
K-nucleotide	2.2295	-4.1362	28.3359	17.9135
Mandelbrot	0.1082	3.3471	-6.9897	-6.5955
N-body	0.6705	2.6524	54.0040	66.0647
Pi-digits	-0.8812	11.5081	-0.0454	6.3821
Regex-redux	0.0714	-0.9969	0.0281	1.5279
Reverse-complement	-6.3830	11.6470	0.0461	18.4365
Spectral-norm	0.6742	5.9835	28.9755	29.3525

As observed in Table 8 (left part), interpreting code with dynamically typed variables generally increases both time and energy consumption compared to compiling with the py_compile module.

From the above results, we can deduce that compiling Python using the py_compile module, with dynamically typed variables, seem to lead to a noteworthy improvement, pri-

marily in terms of energy consumption. However, compiling the code with Nuitka entails a significant increase in both runtime and energy consumption, making it an unfavourable choice.

Fig. 4 shows the average consumption of the graphics card and the processor for SE1, SE2 and SE3.

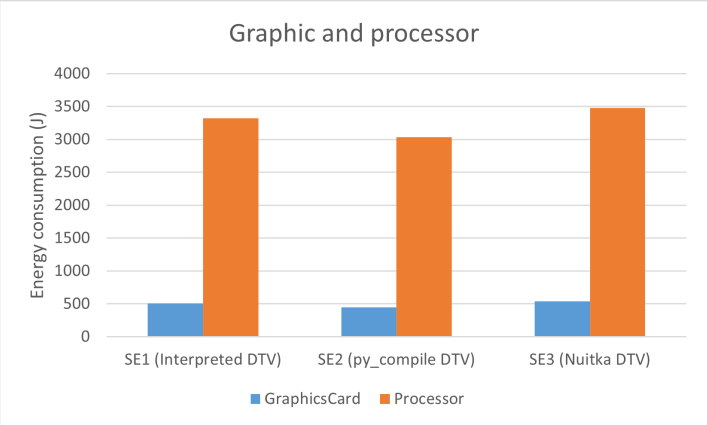


Fig. 4. Graphics and Processor analysis in DTV SEs

The consumption of the graphics card is minimal as it is hardly used in the algorithms analysed. However, both the graphics card and the processor show a slight decrease in consumption in SE2. The data for the graphics card and the processor can be found in the repository. Based on the obtained results, we can conclude that:

> When employing dynamically typed variables (DTV), the optimal choice is to compile the code using the pycompile module.

5.2. SE4, SE5 and SE6: Python Interpreted, compiled (py_compile) and Compiled (Nuitka) with STV

As in the previous section, we are going to compare again the consumption of the three possible options to obtain running software (interpreted, compiled (py_compile) and compiled (Nuitka)) but this time when variables are statically typed.

The best SE for the DUT consumption is SE5 (compiled, py_compile). Therefore, Table 9 shows the percentage of relative increase in consumption of SE4 (interpreted) and SE6 (complied, Nuitka) against SE5 consumption. In the comparison of SE4 (interpreted) and SE5 (compiled, py_compile) it can be observed that half of the algorithms obtain better results in the former and the other half in the latter.

So, it seems than in four of the algorithms, the use of compiling the code with py_compile is much more energy efficient, whereas in three of them there is not a big difference between interpreting or using py_compile. The other three algorithms have better energy

Table 9. Time and energy consumption comparison: interpreted and compiled (Nuitka) vs. compiled (py_compile); with STV

	SE4 vs. SE5		SE6 vs. SE5	
Algorithms	Time (%)	Energy consumption (%)	Time (%)	Energy consumption (%)
Binary-trees	1.2953	-0.0012	18.9672	15.3453
Fannkuch-redux	-0.2066	-2.5227	8.5661	5.4706
Fasta	-1.9867	13.3274	24.6566	26.9903
K-nucleotide	0.5296	0.7845	27.6245	23.3333
Mandelbrot	-0.0196	-1.0369	-10.1396	-9.3452
N-body	0.7733	8.6475	56.6949	93.7315
Pi-digits	-0.4681	4.9135	0.3695	5.9339
Regex-redux	0.0097	-0.0844	-0.0643	6.1216
Reverse-complement	0.0001	20.8442	-0.0786	5.4584
Spectral-norm	-0.2719	-6.6273	15.5815	17.2794

behaviour when are interpreted, but with scarce difference respect to compiling with py_compile.

In this case, there seem that there is no relationship between the energy consumption and the runtime.

Focusing now on the comparison between SE6 (compiled, nuitka) respect to SE5 (compiled, py_compile) the results show that the use of Nuitka is less energy efficient than the use of py_compile, arriving up to the 90% of increment in the N-body algorithm.

Fig. 5 shows the average consumption of the graphics card and the processor for SE4, SE5 and SE6.

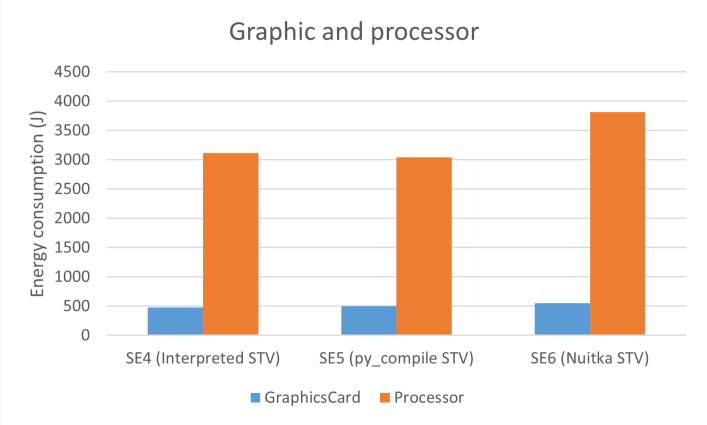


Fig. 5. Graphics and Processor analysis in STV SEs

The consumption of the graphics card is minimal, as it is hardly used in the algorithms analysed, and has no influence on the execution and development methods analysed in these three SEs. The consumption of the graphics card shows hardly any differences be-

tween the SEs. However, there is a slight decrease in processor consumption in SE5, with SE6 consuming the most. The data for the graphics card and the processor can be found in the repository.

Based on the obtained results, we observe that it is better to compile the code using the py_compile module instead of Nuitka. Although it is more difficult to draw a conclusion about the election between interpret the code or using py_compile, taking into consideration the differences in both cases and the results of the hardware components, SE5 also seems to be better option. So, we can conclude that:

> When employing statically typed variables (STV), the optimal choice is to compile the code using the py_compile module.

5.3. DTV vs. STV analysis

One of the reasons why Python might be so inefficient in terms of energy consumption is its great dynamism in typing variables. In this section we present the comparison that aims to check if there is any difference in energy consumption and runtime when using dynamically or statically typed variables, using an interpreted, compiled (py_compile) and compiled (Nuitka) versions of Python.

5.3.1 SE1 and SE4: Python Interpreted DTV and Python Interpreted STV About DUT consumption, the SE that gave us the best results in the interpreted version of Python is SE4, considering the averages of the algorithms. Therefore, Table 10 shows the percentage of relative increase in runtime and consumption of SE1 with respect to SE4.

Table 10. Time and energy consumption comparison: interpreted DTV and interpreted STV

	SE1 vs. SE4	
Algorithms	Time (%)	Energy consumption (%)
Binary-trees	0.0791	4.5825
Fannkuch-redux	-0.0042	-3.3047
Fasta	1.4977	-13.2241
K-nucleotide	1.1377	-2.3733
Mandelbrot	0.2583	1.7795
N-body	-0.0079	6.6703
Pi-digits	-0.0511	5.3325
Regex-redux	-0.0372	2.5103
Reverse-complement	-6.3787	-6.2324
Spectral norm	-5.1020	3.3481

Regarding energy consumption, most algorithms exhibit an increment in SE1 compared to SE4. However, in some algorithms (Fannkuch-redux, Fasta, K-nucleotide, and Reverse-complement), energy consumption improves in their STV version.

Looking at the runtime, there does not seem to be a relationship between run time and energy consumption.

So, we can conclude that:

> When using interpreted Python, the optimal choice is to use statically declared variables.

5.3.2 SE2 and SE5: py_compile DTV vs. py_compile STV When running the compiled code using the py_compile module, we obtained that half of the algorithms had better energy consumption for one of the SE and the other half for the other. So, to provide a simple interpretation of the results, Table 11 shows the percentage increase of the worst average (SE2) over the best average (SE5).

Table 11. Time and energy consumption comparison: compiled (py_compile) DTV and compiled (py_compile) STV

	SE2 vs. SE5	
Algorithms	Time (%)	Energy consumption (%)
Binary-trees	-1.8001	0.1848
Fannkuch-redux	-0.6456	-2.4632
Fasta	-0.1837	-2.5971
K-nucleotide	-0.5440	2.6379
Mandelbrot	0.1304	-2.5380
N-body	0.0942	12.9002
Pi-digits	0.3655	-0.8968
Regex-redux	-0.0989	3.4552
Reverse-complement	0.0047	1.4919
Spectral norm	-5.9938	-8.9491

As can be observed, there is a notable difference of energy consumption in two of the algorithms (N-body and Spectral-norm), each showing better performance in different scenarios. For the rest of the algorithms, there are no significant differences in energy consumption.

In terms of time, there does not seem to be a clear relationship between execution time and energy consumption.

So, we can conclude that:

> When using compiled code using the py_compile module, it does not seem to matter the choice between statically or dynamically declared variables.

5.3.3 SE3 and SE6: Python Compiled (Nuitka) DTV and Python Compiled (Nuitka) STV Finally, when we compile the code using Nuitka, SE3, which is the version with DTV, has yielded the best results, in contrast to the previous cases. Therefore, Table

12 presents the percentage increases of SE6 with respect to SE3.

Table 12. Time and energy consumption comparison: compiled (Nuitka) DTV and compiled (Nuitka) STV

	SE6 vs. SE3	
Algorithms	Time (%)	Energy consumption (%)
Binary-trees	-1.1437	-1.9656
Fannkuch-redux	0.3721	-0.5958
Fasta	1.7378	-3.4186
K-nucleotide	-0.0103	1.8725
Mandelbrot	-3.5124	-0.4181
N-body	1.6516	3.2231
Pi-digits	0.0495	0.4776
Regex-redux	0.0065	1.0231
Reverse-complement	-0.1293	-13.9819
Spectral-norm	-4.6711	-0.4240

In terms of algorithm runtime comparison, we find that only half of the algorithms (Fankuch-redux, Fasta, N-body, Pi-digits and Regex-redux) experience an increase. The rest of the algorithms decrease their execution time when running in SE3.

On the other hand, in terms of energy consumption, only four algorithms show an increase compared to SE3. If we make the opposite comparison (SE3 vs. SE6), most algorithms consume less in SE6.

In this case, again there does not seem to be a relationship between execution time and energy consumption.

So, we can conclude that:

> When using compiled code using Nuitka, the optimal choice is to use dinamically declared variables.

5.4. Comparison of opposites SEs

In this section we present the comparison of the most opposite versions of those included in the study. Concretely we have selected interpreted SEs and compiled by Nuitka SEs, because to run interpreted code is the opposite to run it as machine code.

5.4.1 SE1 and SE6: Python Interpreted with DTV and Python Compiled (Nuitka) with STV First, we are going to compare the most usual way of using the Python language (SE1-Interpreted and DTV) with the version compiled with Nuitka and STV (SE6). The version that has given us the best consumption results is Interpreted DTV (SE1), so Table 13 shows the percentage increase of SE6 with respect to SE1.

Looking at the consumption, 7 out of 10 algorithms increase their consumption when compiled with Nuitka and using statically declared variables. In fact, four of them (Fasta,

Table 13. Comparison of opposing SEs: Interpreted DTV and Compiled (Nuitka) STV

	SE6 vs. SE1	
Algorithms	Time (%)	Energy consumption (%)
Binary-trees	17.3530	10.2926
Fannkuch-redux	8.7954	11.8980
Fasta	25.3067	29.1329
K-nucleotide	25.5240	25.3482
Mandelbrot	-10.3536	-9.9969
N-body	55.5048	67.1617
Pi-digits	0.8931	-4.1392
Regex-redux	-0.0368	3.6103
Reverse-complement	6.7293	-6.9315
Spectral-norm	22.1275	21.5344

K-nucleotide, N-body and Sprectral.norm) increase more than a 20% the consumption. In the counterpart, the three algorithms that decrease their consumption on its version compiled with Nuitka and using statically declared variables (Mandelbrot, Pi-digits, and Reverse-complement) show a decrement less than a 10%.

Also, the execution time increases for most of the algorithms, being over the 55% for one algorithm and over the 25% in other two. In general, there seem to be a relationship between the energy consumption and the execution time.

So, we can conclude that:

> The use of interpreted code with DTV seems to decrease considerably the energy consumption and the execution time than compiled with STV code using Nuitka.

5.4.2 SE3 and SE4: Python Compiled (Nuitka) with DTV and Python Interpreted with STV In this case, we undertake a comparative analysis of two divergent configurations, with the aim of deriving further insights into the impact of using Python and the variable typing approach. Specifically, we contrast the effects of employing compiled Python (Nuitka) with DTV against interpreted Python with STV. Table 14 shows the percentage increase of SE3 with respect to SE4.

In terms of energy consumption, most of the algorithms (8 out of 10) consume more in SE3, being N-body the algorithm with the biggest increase (72%). Exceptions are Mandelbrot (which decreases an 8% the consumption in SE3) and Reverse-complement (but although being less consumer in SE3, the percentage of savings is minimal, around half point).

The execution time is, in general, proportionally related to the energy consumption, there seem to be a relationship between the energy consumption and the execution time.

Therefore, we can affirm that working with a compiled version with Nuitka that also has statically typed variables means a considerable increase in energy consumption and a longer execution time than working with its interpreted version and with dynamically typed variables.

So, we can conclude that:

Table 14. Time and energy consumption comparison: compiled (Nuitka) DTV and interpreted STV.

	SE3 vs. SE4	
Algorithms	Time (%)	Energy consumption (%)
Binary-trees	18.8047	17.6140
Fannkuch-redux	8.3875	8.8448
Fasta	25.0110	15.8869
K-nucleotide	26.9653	20.0819
Mandelbrot	-6.8502	-8.0123
N-body	52.9661	72.5647
Pi-digits	0.7917	0.4904
Regex-redux	-0.0805	5.1245
Reverse-complement	0.0507	-0.5302
Spectral-norm	21.5755	26.1360

The use of interpreted code with STV seems to decrease considerably the energy consumption and the execution time than compiled with DTV code using Nuitka.

And from analysis of sections 5.4.1 and 5.4.2. we can conclude that:

The use of interpreted code seems to decrease considerably the energy consumption and the execution time than compiled code using Nuitka.

5.5. Comparing the algorithms

Having presented the comparisons of the results of the SEs, in this section we show the energy consumption results grouped by algorithm. Fig. 6 shows an overall comparison of the DUT consumption of each algorithm in the different SEs.

As discussed in the previous sections, most algorithms show higher consumption in SE3 and SE6 (Nuitka). However, the Pi-digits algorithm shows an increase in consumption in SE1.

On the other hand, the Mandelbrot and Reverse-complement algorithms have a higher consumption in SE1 and SE4 respectively. It is also remarkable the behaviour of the Fasta algorithm, which shows a similar trend in SE1-SE5, however, in SE6 its consumption increases a lot.

Obviously, performance and resource consumption can vary depending on the specific algorithm, code characteristics, compiler optimisations and other factors. Without knowing specific details of each algorithm in question, some general reasons why algorithms perform better on some SEs than on others are as follows:

- Interpreter Optimisations: CPython, is the reference interpreter for Python, so it is highly optimised and can perform certain run-time optimisations. However, not in all algorithms, these optimisations result in lower power consumption.

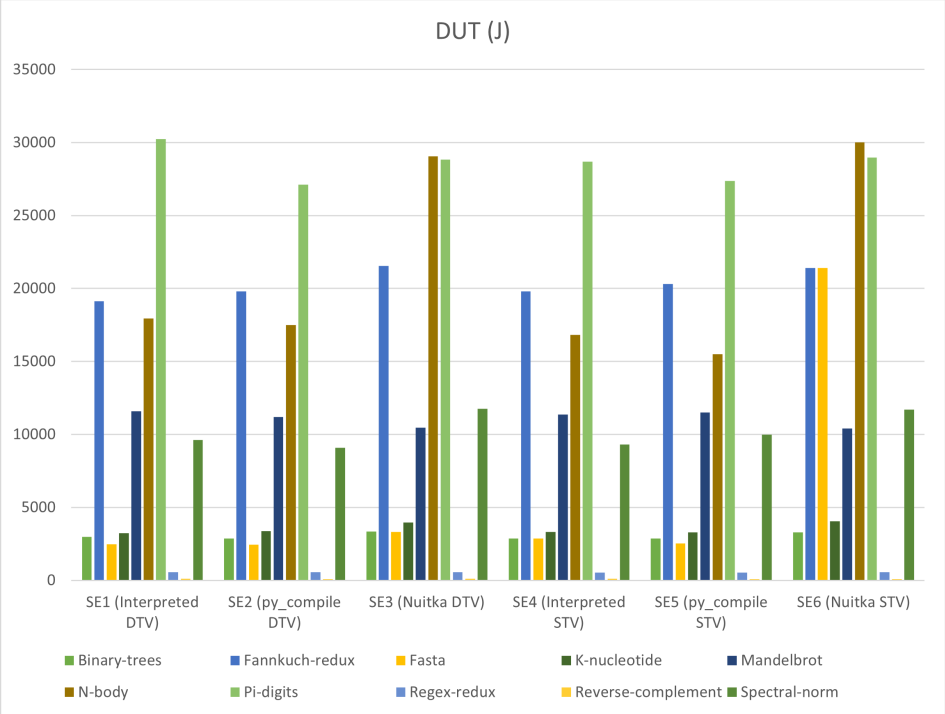


Fig. 6. Comparison of DUT consumption in the algorithms

- Code Characteristics: Code efficiency can vary between different algorithms. Some algorithms may benefit more from optimisations made by the CPython interpreter, while others may show improved performance when compiled.

Fig. 7 and Fig. 8 show the results of processor and graphics card consumption for the algorithms. Although the processor and the graphics card are two independent components, in some algorithms it can be seen how both components have a similar consumption tendency, as is the case of Fasta, N-body and Spectral-norm.

The energy consumption behavior of the algorithms is mainly due to the nature of each algorithm. However, the fact that the fasta, n-body and spectral-norm algorithms have similar resource consumption on both the processor (CPU) and graphics card (GC) may be due to several reasons related to the way the data is processed on both the CPU and the GPU of the graphics card, according to some sources as [21] and [31]:

- Nature of the Algorithm: They may have features that do not benefit significantly from the massive parallelization that a GPU could offer. Some algorithms, especially those with data dependencies or complex control flow structures, may not be as efficient on a GPU.
- Data Transfer Overhead: If algorithms involve large amounts of data transfer between the CPU and GPU, or if the data sets are small, the benefit of using a GPU may be offset by the data transfer overhead.

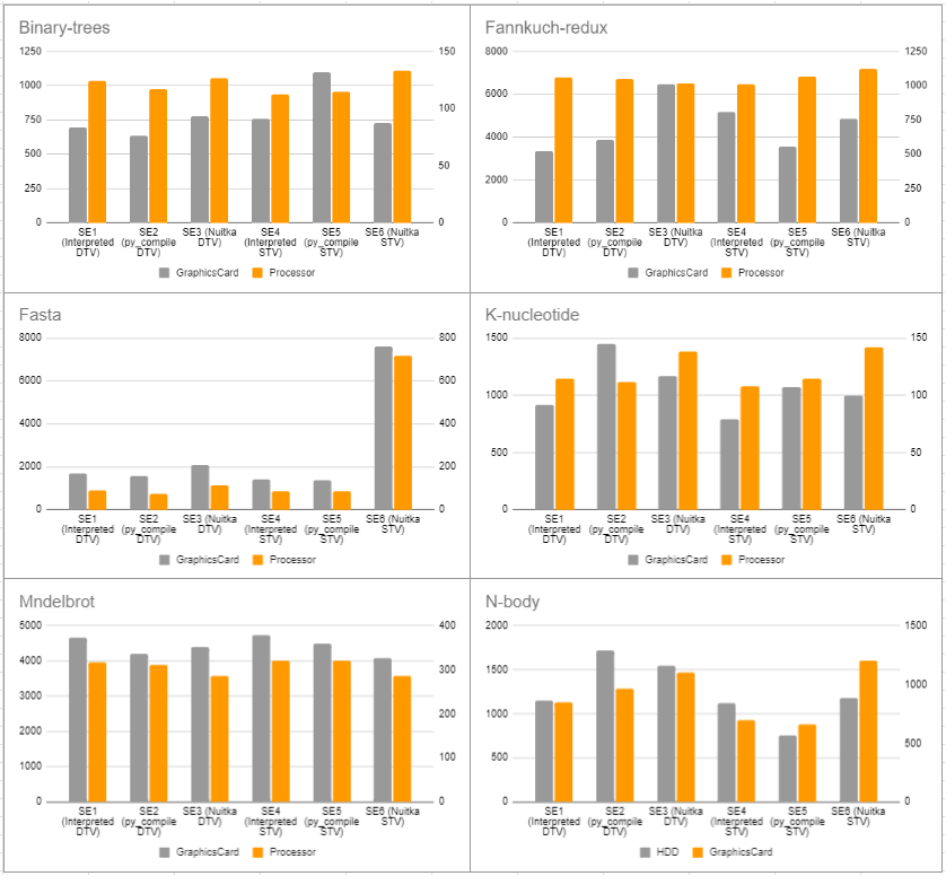


Fig. 7. Graphics and processor energy consumption results for Algorithms 1-6

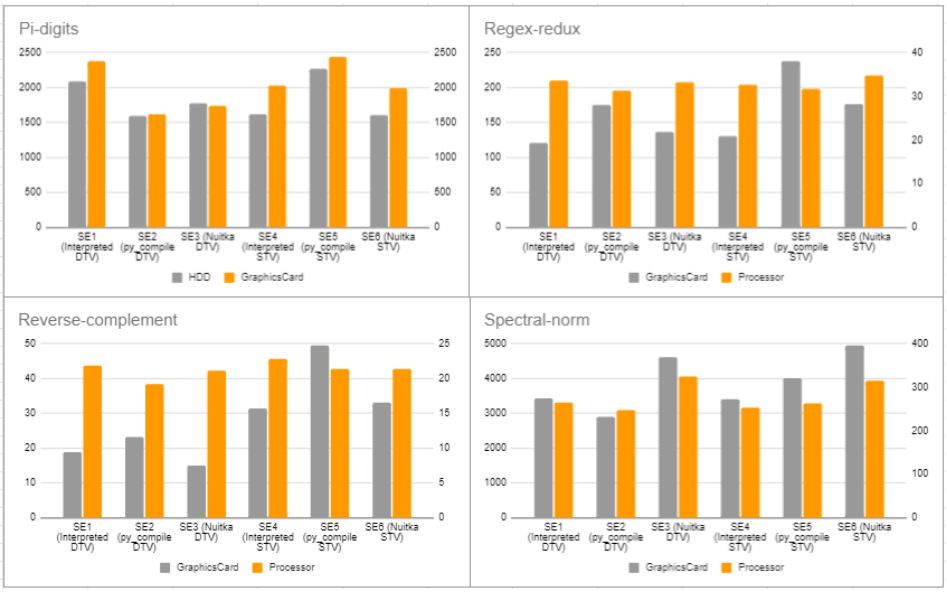


Fig. 8. Graphics and processor energy consumption results for Algorithms 7-10

- Implementation and Optimizations: GPU efficiency can depend heavily on how algorithms are implemented, and the specific optimizations made to take advantage of the GPU architecture. If the implementation has not been optimized to take advantage of GPU-specific features, performance may not differ significantly from CPU performance.
- GPU characteristics: Not all tasks are suitable for GPU execution. Some tasks, especially those involving intensive matrix operations or massively parallel computations, are more suitable for GPU execution.

On the other hand, it is noteworthy that the consumption of the graphics card and the processor in 8 of the 10 algorithms is higher in the SEn with statically typed variables. However, the algorithms are not the same in both cases. In the Pi-digits and Fasta algorithms, graphics consumption is higher with STV, but processor consumption is higher with DTV. In the Fannkuch and K-nucleotide algorithms the opposite is the case.

Regarding the way Python programs are executed, most algorithms (except Binary-trees and K-nucleotide) have a lower processor consumption in compiled SEs, either with the py_compile module or with Nuitka. Similarly, most algorithms (except for Regex-redux and Mandelbrot) also have lower graphic card consumption in compiled versions.

So, we can conclude that:

> The use of statically typed variables seems to considerably increase the power consumption of the graphics card and the processor.

And on how to run the Python code:

The use of compiled code, either with the Python module or with Nuitka, seems to decrease graphics card and processor energy consumption.

6. Answering the research questions and recommendations

Once the results obtained from the measurement have been analysed according to different comparison, we can answer the stated research questions, as follows:

RQ1. Is there any relationship between the energy consumption and time required (by a software application at runtime) and the use of Cpython, py_compile module or Nuitka?
After analysing the results obtained, we can affirm that yes, there is a significant relationship between energy consumption during software runtime and the use of interpreted or compiled Python.

The most efficient option in terms of energy consumption is to use Python compiled with the py_compile module, while the least efficient option is Python compiled with Nuitka, which shows an overwhelming increase in energy consumption. Throughout section 4.1, the increase in time and consumption can be clearly seen when comparing the use or not of py_compile, with an increase in consumption of up to 66%, while in time there are also differences, albeit smaller. Similarly, in section 4.2, very significant differences are found in all the algorithms, with the N-body algorithm standing out, whose difference when compiled with Nuitka represents a 93% increase in energy consumption and a 56% difference in execution time.

In addition, the use of interpreted code seems to significantly decrease energy consumption compared to code compiled with Nuitka. As for the way variables are declared, it does not seem to affect the power consumption of the software when compiling code using the py_compile module. However, when using code compiled with Nuitka, the best option in terms of variable declaration is to use dynamically typed variables.

RQ2. Is there any relationship between the energy consumption required (by a software application at runtime) and the use of variables dynamically or statically typed (during the development)?
After analysing the results obtained, we can conclude that it depends on the type of execution:

- When using interpreted Python, the optimal choice is to use statically declared variables.
- When using compiled code using the py_compile module, the choice between statically or dynamically declared variables does not seem to matter.
- Using code interpreted with DTV seems to considerably decrease energy consumption than compiling with STV code using Nuitka.
- Using code interpreted with STV seems to considerably decrease energy consumption over code compiled with DTV using Nuitka.

Finally, Fig. 9 ranks the software entities studied in the presented study according to the energy consumption measurements obtained in our study. Python compiled with py_compile module and declaring variables occupies the top of the list whilst the bottom is for Python compiled using Nuitka and declaring variables.

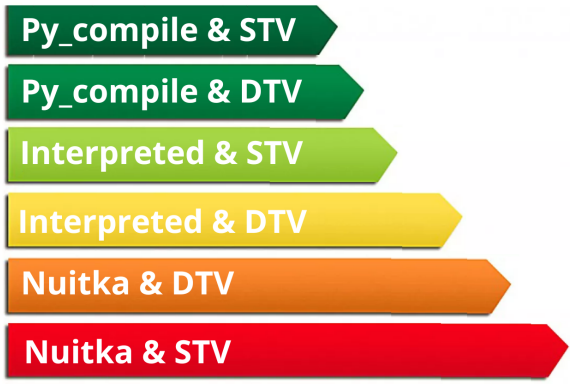


Fig. 9. Python energy-consumption classification

7. Threats to validity

This section tackles the threats to the validity of the study by following the recommendations in [49] and how we have tried to minimize their effects:

Construct validity. The first point is about the reliability of the measurements. We have used EET to measure consumption, which enables exact measurements of the energy consumed by the different hardware components in a very small interval of time (approximately 100 samples per second). Obviously, the measurements obtained are specific to EET and may differ if we use other mechanisms as an estimate, or if we employ other components (where they exist). However, EET has been already validated proving its reliability [34] and has been used previously in other measurements of this type.

Internal validity. With regards to those uncontrolled factors that may affect the results of the experiment, the most remarkable ones are related in the conditions in which the measurements were performed.

Firstly, the algorithms executed were the same in the six SEs (adapting the specific aspects on the way of using Python). Moreover, several executions were performed to mitigate the possible atypical values related with consumption. We used the same DUT to perform the executions and capture the energy consumption and measures have been taken to ensure that the DUT was always in the same conditions for the running of each different execution. To avoid the possible execution of background processes, before starting each measurement, all programs that could cause interference were closed, and the base consumption of the DUT was subtracted from the measurements. Regarding the number of measurements performed, there is no ideal number of measurements. Our decision is based on the recommendation of authors such as [24] for evaluations of software power consumption in a controlled environment. Generally, a sample of 30 measurements is sufficient for the analysis of each intended test case, as it tends to produce a near-normal sampling distribution. However, such as reported in the experimental package [22] statistical significance analysis did not obtain significant difference between most of the compared scenarios with resulting small effect sizes. New empirical studies will be conducted with more complex scenarios to confirm if the obtained scenarios differences in this study can be significant from statistical point of view in more complex settings.

External validity. Finally, related to the power of generalising the results obtained in this experiment. The results are based on a specific combination of algorithms and configuration of Python. Our experiment was performed on CPython interpreter version 3.11 and compilation with Nuitka was performed using the –standalone option.

Therefore, different interpreter versions and other compilation options could differ in the results. So, the study could be repeated using other interpreter versions, as well as extending it by considering other compilation options, other algorithms, the use of libraries, etc. A more exhaustive analysis could also be performed to report the parts of the code that involve higher consumption in the different components. Nevertheless, the current study is a good starting point on how to improve the large amount of power required by Python.

Throughout the document we have mentioned the existence of a new compiler, GraalPy, which promises to be very efficient. We have made some energy consumption measurements with the aim of including it in our study. However, its current state of development has prevented us from performing exhaustive measurements of all our algorithms due to its limitations.

Likewise, we are also aware of the existence of another compiler called Numba [35] which, due to its limitations with some libraries [41], we have not been able to include in our study either.

However, for both cases we have made some measurements. In the case of Numba, we performed the measurements on another version of the Mandelbrot algorithm compatible with this compiler. The same algorithm has also been measured on Interpreted and on Nuitka (DTV and STV) and the results obtained have been compared. In the case of GraalPy, we have measured all the algorithms. The results of the compatible algorithms together with the errors of the remaining algorithms can be found in the experimental package in our repository [22].

8. Conclusions

Sustainability has emerged as a paramount concern within contemporary society, prompting an increasing number of companies to integrate it into their product development practices. However, within the realm of software development companies, the consideration of sustainability remains an area with significant room for improvement. Despite the growing body of research addressing the sustainability of software development, substantial gaps persist. Part of the software sustainability is concerned by the obtaining of green software and the consumption required by software.

Numerous studies have contributed valuable insights into the green software development in general and in programming languages (focus of this paper) in particular. For instance, the work of [36] delves into the energy efficiency of various programming languages. Notably, the study's findings highlight Python, a language widely acclaimed by developers, as one of the most energy-intensive languages. Despite ongoing efforts to optimize Python, these endeavours have yet to yield comprehensive insights into the optimal methods for its development and execution.

Against this backdrop, the present research endeavours to gauge the real energy consumption associated with three distinct methods of executing Python programs, namely CPython, the Py_compile module, and Nuitka. Additionally, two different approaches to

Python development, involving dynamically typed variables and statically typed variables, are considered in this study.

Using the FEETINGS framework [30] and the EET [28], we measured the energy consumption of 10 algorithms written in Python. Each of these algorithms was adapted in two versions (Dynamically typed variables and statically typed variables) and executed 30 times in three different ways (Interpreted, compiled with the py_compile module and compiled with Nuitka). With the resulting consumption data, different comparisons were made to answer our research questions.

Based on the findings, a set of recommendations have been developed to make it easier for Python programmers to apply it in real life. Each of these recommendations is detailed below:

- R1. If you want to speed up loading and execution, it is better to compile using the py_compile module rather than Nuitka.
- R2. If the code is compiled using py_compile, it is better to declare the variables (STV).
- R3. When using interpreted Python is required, the best option is to use statically typed variables (STV).
- R4. When using dynamically typed variables (DTV) is required, the best option is to use compiled Python with py_compile module.
- R5. It is preferable to run the interpreted code with DTV, rather than to compile it with Nuitka.
- R6. A compiled version of Python with dynamically typed variables is better than an interpreted version with statically typed variables.
- R7. The use of statically typed variables does not always save energy, it depends on the algorithm executed.
- R8. It is better to use Python on its classical way (interpreted with DTV) than using STV and compile using Nuitka.
- R9. To reduce GC and processor consumption, it is better to use dynamically typed variables.
- R10. It is better to compile the code, either with the Python module or with Nuitka, to reduce the consumption of the graphics card and the processor.

Considering that Python is a very versatile language, several lines of future work are open. Python offers a multitude of libraries [48] that eliminate the need to write code from scratch, providing the programmer with various functionalities such as processing large amounts of data or image processing. It would therefore be interesting to check the influence of these libraries on energy consumption.

We are also aware of the existence of other types of benchmarks, for example "The Python Performance Benchmark Suite" [47], specifically for measuring the performance of programs written in this language. It would therefore be interesting to replicate our study using this benchmark to compare the results obtained and offer more specific recommendations.

Other compilers (Numba and GraalPy) were considered to be included in our study. Due to their limitations, mainly because they are yet under construction, they were not included, but they will be considered in future works. Work is also underway to carry out the study presented in this paper in other programming languages, with the aim of providing relevant information on the influence that the use of different compilers has

on the energy consumption of the software. Finally, we also consider of main interest to study the energy efficiency of compiler optimisations, being another line of future work.

Acknowledgments. This work has been supported by the following projects: OASSIS (PID2021-122554OB-C31/ AEI/10.13039/ 501100011033/FEDER, UE); EMMA (Project SBPLY/21/180501/000115, funded by CECD (JCCM) and FEDER funds); SEEAT (PDC2022-133249-C31 funded by MCIN/AEI/ 10.13039/501100011033 and European Union NextGenerationEU/PRTR); PLAGEMIS (TED2021-129245B-C22 funded by MCIN/AEI/ 10.13039/501100011033 and European Union NextGenerationEU/PRTR). Financial support for the execution of applied research projects, within the framework of the UCLM Own Research Plan, co-financed at 85% by the European Regional Development Fund (FEDER) UNION (2022-GRIN-34110).

References

1. Nuitka (2023), https://nuitka.net/index.html
2. Py_compile (2023), https://docs.python.org/es/3/library/py_compile.html
3. Python implementations (2023), https://wiki.python.org/moin/PythonImplementations#Other_Implementations
4. TIOBE Index (Feb 2024), https://www.tiobe.com/tiobe-index/
5. Abdelnabi, A.A.B.: An analytical hierarchical process model to select programming language for novice programmers for data analytics applications. pp. 128–132. IEEE (2019)
6. Abdulsalam, S., Lakomski, D., Gu, Q., Jin, T., Zong, Z.: Program energy efficiency: The impact of language, compiler and implementation choices. pp. 1–6. IEEE (2014)
7. Abdulsalam, S., Zong, Z., Gu, Q., Qiu, M.: Using the Greenup, Powerup, and Speedup metrics to evaluate software energy efficiency. In: 2015 Sixth International Green and Sustainable Computing Conference (IGSC). pp. 1–8 (2015)
8. and/or its affiliates, O.: GraalPy native-image, https://www.graalvm.org/latest/reference-manual/python/native-applications/
9. and/or its affiliates, O.: GraalPy (2024), https://www.graalvm.org/latest/reference-manual/python/
10. Anaconda, I.a.o.: AOT compilation (2020), https://numba.pydata.org/numba-doc/dev/user/pycc.html
11. Aycock, J.: A brief history of just-in-time. ACM Comput. Surv. 35(2), 97–113 (Jun 2003), https://doi.org/10.1145/857076.857077, place: New York, NY, USA Publisher: Association for Computing Machinery
12. BEREC: BEREC Report on Sustainability: Assessing BEREC's contribution to limiting the impact of the digital sector on the environment. Tech. rep. (Jun 2022), https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-sustainability-assessing-berecs-contribution-to_-limiting-the-impact-of-the-digital-sector-on-the-environment
13. Bhattacharya, P., Neamtiu, I.: Assessing programming language impact on development and maintenance: A study on C and C++. pp. 171–180 (2011)
14. Calero, C., Polo, M., Moraga, M.Ã.: Investigating the impact on execution time and energy consumption of developing with Spring. Sustainable Computing: Informatics and Systems 32, 100603 (2021), publisher: Elsevier
15. Carbonnelle, P.: PYPL (Oct 2023), https://pypl.github.io/PYPL.html
16. Cass, S.: Top Programming Languages 2022 (Aug 2022), https://spectrum.ieee.org/top-programming-languages-2022

17. Connolly Bree, D., Ó Cinnéide, M.: Inheritance versus Delegation: which is more energy efficient? pp. 323–329 (Jun 2020)
18. Delorey, D.P., Knutson, C.D., Chun, S.: Do programming languages affect productivity? a case study using data from open source projects. pp. 8–8. IEEE (2007)
19. Gordillo, A., Mancebo, J.: ELLIOT: GESTIÓN Y ANÁLISIS DE DATOS DE CONSUMO DE SOFTWARE. In: Calidad y sostenibilidad de sistemas de información en la práctica. 1 edn. (Jan 2022)
20. Gouy, I.: The Computer Language Benchmarks Game (2008), https://benchmarksgame-team.pages.debian.net/benchmarksgame/sometimes-people-just-make-up-stuff.html
21. Grama, A.: Introduction to parallel computing. Pearson Education (2003)
22. Jimenez, E., Gordillo, A., Calero, C., Moraga, M.Ā., Garcia, F.: Repository of energy consumption of Python results, https://github.com/GrupoAlarcos/PythonEnergyConsumptionStudy
23. Kaushik, S.: Best Python Interpreters: Choose the Best in 2023 (Nov 2022), https://hackr.io/blog/python-interpreters
24. Kern, E., Hilty, L.M., Guldner, A., Maksimov, Y.V., Filler, A., Gröger, J., Naumann, S.: Sustainable software products—Towards assessment criteria for resource and energy efficiency. Future Generation Computer Systems 86, 199–210 (2018), publisher: Elsevier
25. Krein, J.L., MacLean, A.C., Knutson, C.D., Delorey, D.P., Eggett, D.L.: Impact of programming language fragmentation on developer productivity: a sourceforge empirical study. International Journal of Open Source Software and Processes (IJOSSP) 2(2), 41–61 (2010), publisher: IGI Global
26. Lavazza, L., Morasca, S., Tosi, D.: An empirical study on the effect of programming languages on productivity. pp. 1434–1439 (2016)
27. Lima, L.G., Soares-Neto, F., Lieuthier, P., Castor, F., Melfe, G., Fernandes, J.P.: On Haskell and energy efficiency. Journal of Systems and Software 149, 554–580 (2019), https://www.sciencedirect.com/science/article/pii/S0164121218302747
28. Mancebo, J., Arriaga, H.O., García, F., Moraga, M.Ā., García-Rodríguez de Guzmán, I., Calero, C.: EET: A Device to Support the Measurement of Software Consumption. In: 2018 IEEE/ACM 6th International Workshop on Green And Sustainable Software (GREENS). pp. 16–22 (2018)
29. Mancebo, J., Calero, C., Garcia, F.: GSMP: Green Software Measurement Process. pp. 43–67 (Oct 2021)
30. Mancebo, J., Calero, C., García, F., Moraga, M.Ā., de Guzmán, I.G.R.: FEETINGS: Framework for energy efficiency testing to improve environmental goal of the software. Sustainable Computing: Informatics and Systems 30, 100558 (2021), publisher: Elsevier
31. Merrit, R.: What Is Accelerated Computing? (Sep 2021), https://blogs.nvidia.com/blog/what-is-accelerated-computing/
32. Meyerovich, L.A., Rabkin, A.S.: Empirical analysis of programming language adoption. pp. 1–18 (2013)
33. Montanaro, S.: Python dynamic language (Feb 2012), https://n9.cl/pyaxq8
34. Muna, A.: Assessing programming language impact on software development productivity based on mining oss repositories. ACM SIGSOFT Software Engineering Notes 44(1), 36–38 (2022), publisher: ACM New York, NY, USA
35. Oliphant, T.: Numba (2012), https://numba.pydata.org/
36. Pereira, R., Couto, M., Ribeiro, F., Rua, R., Cunha, J., Fernandes, J.P., Saraiva, J.: Energy Efficiency across Programming Languages: How Do Energy, Time, and Memory Relate? In: Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering. pp. 256–267. SLE 2017, Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3136014.3136031, event-place: Vancouver, BC, Canada

37. Pereira, R., Couto, M., Saraiva, J., Cunha, J., Fernandes, J.P.: The Influence of the Java Collection Framework on Overall Energy Consumption. In: Proceedings of the 5th International Workshop on Green and Sustainable Software. pp. 15–21. GREENS '16, Association for Computing Machinery, New York, NY, USA (2016), https://doi.org/10.1145/2896967.2896968, event-place: Austin, Texas
38. Phipps, G.: Comparing observed bug and productivity rates for Java and C++. Software: Practice and Experience 29(4), 345–358 (1999), publisher: Wiley Online Library
39. Pinto, G., Liu, K., Castor, F., Liu, Y.D.: A Comprehensive Study on the Energy Efficiency of Java's Thread-Safe Collections. In: 2016 IEEE International Conference on Software Maintenance and Evolution (ICSME). pp. 20–31 (2016)
40. public: Nuitka options (2023), https://github.com/Nuitka/Nuitka
41. python: Numba limitations (2020), https://numba.readthedocs.io/en/stable/reference/pysupported.html
42. Ralph, P., Nauman, A.: Empirical Standards ACM SIGSOFT (Mar 2021), https://github.com/acmsigsoft/EmpiricalStandards
43. Redondo, J.M., Ortin, F.: A Comprehensive Evaluation of Common Python Implementations. IEEE Softw. 32(4), 76–84 (Jul 2015), https://doi.org/10.1109/MS.2014.104, place: Washington, DC, USA Publisher: IEEE Computer Society Press
44. Reya, N.F., Ahmed, A., Islam, T.Z.M.M.: GreenPy: Evaluating Application-Level Energy Efficiency in Python for Green Computing. Annals of Emerging Technologies in Computing (AETiC) 7(3) (2023)
45. Shaw, A.: Why is Python so slow? (Jul 2018)
46. Spinellis, D.: Choosing a programming language. IEEE software 23(4), 62–63 (2006), publisher: IEEE
47. Stinner, V.: The Python Performance Benchmark Suite (2017), https://pyperformance.readthedocs.io/
48. Team, G.L.: Top 30 Python Libraries To Know in 2024 (Nov 2023), https://www.mygreatlearning.com/blog/open-source-python-libraries/
49. Wohlin, C., Runeson, P., Hst, M., Ohlsson, C., Regnel, Wessln, A.: Experimentation in Software Engineering. Springer (2012)
50. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in software engineering. Springer Science & Business Media (2012)

Elisa Jimenez is currently a PhD student in Advanced Information Technologies. She graduated in Computer Engineering in 2021 and obtained a Master's degree also in Computer Engineering in 2022. She is a member of the Alarcos Research Group, University of Castilla-La Mancha (UCLM), Ciudad Real, Spain. She works on software sustainability by performing energy consumption measurements among other research. Contact her at elisa.jimenez@uclm.es

Alberto Gordillo is currently a PhD student in Advanced Information Technologies. He graduated in Computer Engineering in 2020 and obtained a master's degree also in Computer Engineering in 2022. He is a member of the Alarcos Research Group, University of Castilla-La Mancha (UCLM), Ciudad Real, Spain. He works on process automation and software sustainability by performing energy consumption measurements among other research. He holds the ITIL4 professional certification. (Information Technology Infrastructure Library). Contact him at alberto.gordillo@uclm.es

Coral Calero is currently a Full Professor with the University of Castilla-La Mancha (UCLM), Ciudad Real, Spain. She is a member of the Alarcos Research Group, being responsible of the "Green and Sustainable software" line research. She received the M.Sc. degree in 1996 from the University of Seville and the Ph.D. degree in Computer science in 2001 from the UCLM. She holds the professional certifications PMP (Project Management Professional), Scrum Master, and Scrum Manager. She is member of the Spanish Committee on Research Ethics. Contact her at Coral.Calero@uclm.es

Mª Ángeles Moraga is currently an Associate Professor and a Member of the Alarcos Research Group, University of Castilla-La Mancha (UCLM), Ciudad Real, Spain. Mª Angeles Moraga received the M.Sc. degree in 2003, and the Ph.D. degrees in computer science in 2006, from the UCLM. She works on software quality and measures, and software sustainability. She holds the following professional certifications: PMP (Project Management Professional), Scrum Master I - PSM I, and Scrum Manager. Contact her at MariaAngeles.Moraga@uclm.es

Félix García is Full Professor at the University of Castilla-La Mancha (UCLM) and a member of the Alarcos Research Group, Ciudad Real, Spain. His research interests include software sustainability, business process management, software processes, software measurement, and agile methods. He holds the following professional certifications: PMP (Project Management Professional), CISA (Certified Information Systems Auditor), Scrum Master I - PSM I, and Scrum Manager. Contact him at Felix.Garcia@uclm.es.

Received: February 28, 2024; Accepted: September 25, 2024.

Computer Science and Information Systems 22(3):727–753 https://doi.org/10.2298/CSIS240229030F

Stages and Critical Success Factors in ERP Implementation: Insights from Five Case Studies

Sergio Ferrer-Gilabert, Beatriz Forés, and Rafael Lapiedra

Department of Business Administration and Marketing, Universitat Jaume I
Avinguda de Vicent Sos Baynat, s/n, 12006 Castelló de la Plana, Spain
sgilaber@uji.es

Abstract. This study proposes a conceptualization of the implementation stages of an ERP system and identifies the critical factors that ensure success. Data collection was done in two main stages. The first stage was aimed at identifying the Critical Success Factors (CSFs) for ERP implementation and was achieved by developing a structured, self-administered electronic questionnaire, which was sent to a panel of 31 Spanish experts in information systems. The aim of the second stage was to confirm the relevance of each of the identified CSFs in the different stages of the proposed ERP life cycle model. Specifically, this stage consisted of four semi-structured interviews with five Spanish firms, from different industries, which have implemented an ERP system. The results of our case studies offer an understanding of the dynamics and complexity of each case, highlighting the success factors, processes, critical issues, relevant agents and influences on the five ERP implementation stages.

Keywords: ERP systems; ERP life cycle; Critical success factors; Implementation; Information systems, Case Studies.

1. Introduction

Enterprise Resource Planning (ERP) systems are software packages for managing organizational information systems that integrate all business processes ([1]), sharing information and using a single common database ([2]). ERP systems combine business processes and information technology (IT) features ([3]). Companies worldwide of any type or size that have implemented or are implementing ERP systems consider their use a determining factor of their competitive advantage ([4]; [5]; [6]; [7]; [8]; [9]).

In general, these systems help companies manage their business more effectively and efficiently, by integrating process flows across functional areas ([1]; [10]; [8]); standardizing core activities to meet industry standards ([11]); improving the quality of data analysis for better strategic planning of assets, decision-making, and managerial control; reducing inventory levels; optimizing supply chain coordination ([12]) and enabling higher quality customer service ([13]; [11]; [14]; [15]; [8]). The adoption of these types of systems also has a significant effect on sustainability ([16]), helping to reduce costs, material use, and waste.

However, some authors claim that more than 50% of attempts to implement an ERP unfortunately end in failure or do not meet the expected objectives ([17]; [18]). The high failure rate in ERP implementation is also due to the multiple organizational, strategic, and

human factors involved in the process ([19]; [20]). In this context, it is essential to understand the structure and key factors in the ERP life cycle. This process includes the initial stages of needs and scope analysis, followed by the implementation and deployment of the solution, and its subsequent maintenance and updating. A better understanding of the stages of the ERP life cycle and its critical success factors is needed, not only by academia but also the companies themselves, who see the successful implementation of these information systems as an important guarantee of their current and future competitiveness ([5]; [6]; [7]; [8]; [9]).

Despite the growing number of theoretical works (e.g. [17]; [21]; [22]) and exploratory studies ([11]; [21]; [23]; [24]) that contribute to the related knowledge base, very few studies have sought to understand the dynamics and results of ERP implementation in business practice ([3]; [25]), and fewer still in a context marked by the COVID pandemic, which justifies the main focus of this study.

Through an in-depth study of five business cases, this work contributes to the existing research by clarifying the causes of problems that arise when implementing these information systems, the key success factors, and generally offering a better understanding of the ERP implementation process and the main agents involved.

To this end, the paper has been structured into five main blocks. After this introduction, the second section reviews the models in the literature that describe the stages of the ERP implementation process and the ERP life cycle. The following section details the critical success factors in the implementation process identified by the literature. The fourth section presents the methodology. Then, the fifth section details the main results of the quantitative research involving a panel of experts and the study of five business cases. The final section sets out the conclusions and implications, both for academia and for business practice, as well as future research lines.

2. Models of the ERP life cycle

Although there is a large body of literature on the ERP life cycle, there is no clear consensus on the number of stages that should be included. The academic community offers a wide variety of proposals with different numbers of stages. Table 1 compiles a sample of the models that are most widely recognized by the research community, based on the number of citations.

From the analysis of the Table 1 , it is clear that the most widely adopted ERP life cycle models by the research community are those proposed by Cooper and Zmud (1990) [30] and Markus and Tanis (2000) [27].

The model proposed by Markus and Tanis (2000) [27] consists of four phases: **"Chartering"**, **"Project"**, **"Shakedown"**, and **"Onward and upward"**. The **"Chartering"** phase involves crucial decisions related to funding an enterprise system, and the engagement of key players such as suppliers, consultants, executives, and IT specialists. Key tasks include developing a clear business model, selecting software packages, identifying a project manager, and approving timelines and budgets.

The **"Chartering"** phase in the Markus and Tanis (2000) [27] model is split into **"Initiation"** and **"Adoption"** in the model proposed by Cooper and Zmud (1990) [30]. The **"Initiation"** phase includes actively or passively searching for business opportunities and

Table 1. Models of the ERP life cycle

Stages	Authors	Article	Journal/Book	Citations
2	Plant and Willcocks (2007) [24]	Critical success factors in international ERP implementations: a case research approach	The Journal of Computer Information Systems	259
3	Loh and Koh (2004) [26]	Critical elements for a successful enterprise resource planning implementation in small-and medium-sized enterprises	International Journal of Production Research	542
4	Markus and Tanis (2000) [27]	The enterprise systems experience-from adoption to success	Framing the domains of IT research: Glimpsing the future through the past	2497
5	Ross and Vitale (2000) [28]	The ERP revolution: surviving vs. thriving	Information Systems Frontiers	959
5	Esteves and Pastor (2006) [29]	Organizational and technological critical success factors behavior along the ERP implementation phases	Enterprise Information Systems	55
6	Cooper and Zmud (1990) [30]	Information Technology Implementation Research: A Technological Diffusion Approach	The Institute of Management Science	4518
7	Shanks (2000) [31]	A model of ERP project implementation	Journal of Information Technology	997

Source: own elaboration

threats, as well as potential IT solutions, culminating in the selection of a software package. The **"Adoption"** phase involves negotiating the resources needed for implementing the selected IT solution.

In the **"Project"** phase, the aim of the activities conducted is to operationalize the information system across organizational units, involving the project manager, project team members (often non-technical staff from various business and functional areas), internal IT specialists, suppliers, and consultants. This phase includes software configuration, system integration, organizational adaptation, and training organization members. In terms of the objectives as well as the scope and composition, this phase corresponds to the Adaptation phase defined by Cooper and Zmud (1990) [30].

The **"Shakedown"** phase involves the organization coming to terms with the enterprise system. The project team may continue to be involved or may hand over control to operational managers and end-users. It ends when daily operations with the new application become normalized or if the organization decides to discontinue the ERP implementation process. This phase corresponds to the Acceptance and Routinization phases of the Cooper and Zmud (1990) [30] model.

Lastly, the **"Onward and upward"** phase entails the assessment of whether the investment has yielded benefits. It begins once daily operations stabilize and continues until the system is replaced by a disruptive update or a new system. This phase features the intense involvement of operational managers, end-users, IT support staff, and potentially the IT system provider and consultants. It focuses on continuous improvement, user skill development, and benefit evaluation. Finally, this phase is reflected in the Infusion phase of Cooper and Zmud (1990) [30].

Comparing and analysing the models of Cooper and Zmud (1990) [30] and Markus and Tanis (2000) [27] it can be seen that, despite the robustness and specificity of the second model regarding the definition of each stage of ERP implementation, the **"Chartering"** stage proposed in this model should be divided into two, as proposed by Cooper and Zmud (1990) [30].

This division is appropriate due to the nature of the processes that occur in each stage, as well as the agents involved and their strategic position. Thus, in the first stage, which Cooper and Zmud (1990) [30] call **"Initiation"**, our proposal includes the analysis and assessment of the suitability of incorporating an ERP system, either replacing the current one or installing one for the first time. This analysis must be based on a strategic planning vision. As emphasized by recent research, it is crucial to ensure the alignment of the ERP process implementation with the organizational strategic plans [32].

To carry out this analysis, it is essential to define the company's needs and assess different solution alternatives—not just from a technical or functional standpoint, but adopting a strategic perspective, exploring what opportunities an ERP system can offer and what problems it can resolve. This stage concludes with the identification and planning of project objectives and scope, and the selection of the project manager and their team. The outcome of this stage may necessitate seeking an ERP solution that aligns with the company's strategic goals, or the end result may be that the company decides against such a change. To emphasize the information gathering and analytical capabilities of project management and the project team, we propose naming this stage **"Analysis"**.

Then comes the second phase in our model, called **"Adoption"**, which involves other stakeholders beyond the steering committee, such as suppliers, consultants, executives, and IT specialists. This stage involves interactions with developers and implementers of IT solutions. Thus, in the **"Adoption"** stage, the company makes contact with vendors and determines the approach and specific resources allocated for project implementation. In this stage, the identification and planning from the previous stage are translated into a concrete plan of action. The subsequent stages of our model align with those proposed by Markus and Tanis (2000) [27]. Table 2 shows the two analysed models of Cooper and Zmud (1990) [30] and Markus and Tanis (2000) [27], as well as our proposed ERP implementation model.

Table 2. Proposed ERP Implementation Model

Cooper and Zmood (1990) [30]	Initiation	Adoption	Adaptation	Acceptance	Routinization	Infusion
Markus and Tanis (2000) [27]	Chatering		Project	Shakedown		Onward and Upward
Own Model	**Analysis**	**Adoption**	**Project**	**Delivery and stabilization**		**Continuity and Improvement**

Source: own elaboration

3. Critical Success Factors in ERP Implementation

The study of factors that determine the successful implementation of ERP systems has been a key research issue in the literature ([23]; [20]). However, some authors argue that research in this field has often been limited to simply identifying possible critical success factors (CSFs), without understanding their role and effective influence in a real-life business context ([33]).

Among all the analysed studies, that of Somers and Nelson (2001) [34] is the most cited and has informed many other studies, both theoretical and empirical (e.g. [17]; [35]).Based on the literature review, we consider it appropriate to add three CSFs to the 22 identified by Somers and Nelson (2001) [34].

Studies such as those by Osman (2018) [36] and Reitsma and Hilletofth (2018) [37] and Finney and Corbett (2007) [22] highlight the importance of "Software development, testing, and troubleshooting" as a critical component of ERP implementation. This involves testing organizational and production processes within the ERP, including a specific plan for such testing. Since the startup of an ERP system requires configurations, adaptations, and programming, a mechanism for checks and verifications is needed to ensure the proper functioning of the system.

Furthermore, Osman (2018) [36], Shatat and Dana (2016) [38] and Finney and Corbett (2007) [22] discuss the "Delegation of authority to workers" as a means to motivate employees and encourage them to make a greater effort to ensure a successful implementation. This delegation of authority enhances trust, productivity, proactivity, and leads to greater involvement in the process, thereby improving efficiency.

Additionally, there is the crucial role of "Internal and external benchmarking" processes, as considered necessary by Butarbutar et al. (2023) [39], Ahmed et al. (2017) [40] and Finney and Corbett (2007) [22]. These processes enable organizations to learn and incorporate new ideas and knowledge, particularly in information systems and, by extension, in ERP implementation. As such, they are directly linked to strategic organizational decisions. Table 3 presents the 25 factors considered in this study and their main antecedents in the literature.

These 25 factors are grouped into organizational-related, technological-ERP-related, project-related and individual-related factors, in line with studies such as those by Ram and Corkindale (2014) ([20]) and Ayat et al. (2021) ([41]). In addition, we test the association of these 25 factors with both the achievement of success in ERP implementation and post-implementation stages ([39]) and with performance improvements [20] (Table 3).

Table 3. Critical Success Factors

Organization-related	
F1	Top management support
F2	Management of expectations
F3	Vendor / customer partnerships
F4	Use of consultants
F5	Dedicated resources
F6	Change management
F7	Clear goals and objectives
F8	Interdepartmental communication
F9	Interdepartmental cooperation
F10	Ongoing vendor support
F11*	Empowered decision-makers
Technological/ERP-related	
F12	Use of vendors' development tools
F13	Careful selection of the appropriate package
F14	Data analysis and conversion
F15	Business process reengineering
F16	Defining the architecture
Project-related	
F17	Project champion
F18	Project management
F19	Steering committee
F20	Minimal customization
F21	Project team competence
Individual-related	
F22	User training and education
F23	Education on new business processes
F24*	Software development, testing and troubleshooting
F25*	Benchmarking, internal and external

Source: own elaboration

Note: * New factors added to the model proposed by Somers and Nelson (2001)

4. Research methodology

The research strategy primarily consisted of a multiple-case design with five Spanish companies, from different sectors, which had recently completed the implementation of an ERP solution. The rationale for the multiple-case design was that the focus could be directed at understanding the dynamics and complexities of each case; specifically, the processes, critical issues, agents and influences of the different stages — *"Analysis"*, *"Adoption"*, *"Project"*, and *"Delivery and Stabilization"* — in each organization's ERP implementation project. This approach proved to be particularly well suited for this study because it unveiled a multitude of factors, dimensions and stages that make the implementation of ERP software such a complex process.

4.1. Data collection

Data collection was done in two main stages. The first stage was aimed at identifying the CSFs with the greatest impact on the success of ERP implementation, which led to the development of a structured self-administered electronic questionnaire for a panel of 31 experts in information systems. These experts come from academic and/or professional backgrounds.

The survey consists of three parts; in the first part, we collect the experts' personal characteristics; in the second part, the experts assess the impact of the CSFs on the overall ERP implementation process (yes/no questions); and in the last part of the survey, the same group of experts also assess the degree of importance of each CSF in the different stages of our proposed ERP life cycle model. The question block in this second part is based on Likert-type scales with seven possible answers ranging from 1 to 7 (1 = Null, 2 = Quite Low, 3 = Low, 4 = Medium, 5 = Quite High, 6 = High, and 7 = Very High). The third part consists of a group of questions asking the experts to indicate in which stage or stages, according to our proposed model, they consider each of the 25 CSFs most relevant. Each factor is presented along with the option to select the stage or stages they consider relevant.

This questionnaire lasted approximately 20 minutes and was administered at the end of 2018, following the recommendations of Stanton and Rogelberg (2001) [42] for the planning and implementation of Internet-based research and to avoid possible technological risks. In the end, we obtained 29 responses. Table I in A. Appendix shows the percentage importance that experts assigned to each of the 25 factors for the development of the five stages of our ERP life cycle model. Although all the percentages are presented in the Table I in A. Appendix, only those greater than 50% are highlighted.

The second stage was aimed at corroborating the relevance of each of the CSFs mentioned in the literature and evaluated by the panel of experts in the different stages of our proposed ERP life cycle model. In each of the four stages, we analysed aspects related to how planning has been developed, time and resource management, the participation of people, both internal to the company and external, how training of staff has been carried out, and the degree of achievement of objectives, among others.

Specifically, this stage consisted of four semi-structured interviews. The interviews were conducted with one or two individuals from each of the five firms that had recently implemented the ERP project. Therefore, the interviews took place after the completion of each stage of the ERP life cycle in each of the studied companies. We did not conduct

interviews regarding the fifth and final stage of the ERP life cycle in our model (Continuity and Improvement) as we believe that companies require a period of work and development in this stage before it can be studied. Thus, we have left it for future research.

Each of the four interviews lasted around an hour and a half. The interviews related to the first stage, **"Analysis"**, all took place in 2018 except for Ventur, which was the first case analysed in 2016 (see Figure 1). However, the complexity of ERP implementation required different timings for the following interviews, as explained later (see Figure 1). The case study concluded in 2022.

All of the informants were directly involved in the ERP implementation process and were selected based on their roles in the project. Therefore, for the design of the interview, the selection of the most important aspects in the implementation of the ERP and the most relevant CSFs in each stage was based on the results obtained in the qualitative research with the panel of experts (see Table I in A. Appendix for the results of the panel of experts).

Open-ended questions were used throughout the interviews (see Table II in A. Appendix). They allowed for flexibility and provided the "possibilities of depth; and to make better estimates of the respondent's true intentions, beliefs, and attitudes" ([43]).

All interviews were audio-taped for subsequent transcription and for verification of accurate interpretation. Member checks were performed during which the informants were asked to review the transcription of their interviews for verification or amendment of the content. Follow-up questions were asked, when required, to further clarify ambiguities or discrepancies.

The data from this study were validated using a triangulation method ([44]). To this end, we kept in touch with the panel of experts, the individuals inside the firms (in some case studies we had the opportunity to talk with both the firm's CEO and a member of the top management team) and external suppliers of the software solution. The results show that while each of the five cases is different with regard to the type of software solution that was being implemented, the same process was developed, similar tasks were performed, and similar factors impacted the process. Although the generalizability of the findings has yet to be determined, there is no obvious reason to believe that the results would not apply to a larger population.

4.2. The cases

The information on the five organizations that participated in the study are presented in Table 4 and Table 5. Figure 1 presents the timeline over which the different companies addressed each of the four implementation stages analysed. As can be seen, the COVID-19 pandemic coincided with the project stage of two of the five analysed companies, whose **"Delivery and stabilization"** stages were extended until 2022.

5. Results

5.1. Analysis stage

In the first stage of our model, **Analysis**, our goal is to understand the strategic aspects or reasons why the organization has decided to replace the current information system with an ERP ([40]; [21]; [22]).

Table 4. Characteristics of the companies

Company	Number of Employees	Annual turnover 2018	Foundation	Sector or activity
FM Iluminación, S.L.U.	15	€1.7 million	1992	Manufacture of lighting fixtures, recessed and hanging lights
Logicus Engineering, S.L.U.	12	€600,000	2013	Project management services, technical and legal directions
Molcaworld, S.L.U.	27	€4 million	1998	Branding and visual communication
Visionis Distribución, S.L.	45	€5.5 million	2008	Distribution of optical products for the optical professionals' sector
Internacional Ventur, S.A.	52	€24 million	1988	Manufacture and distribution of dental products and dental logistics platform

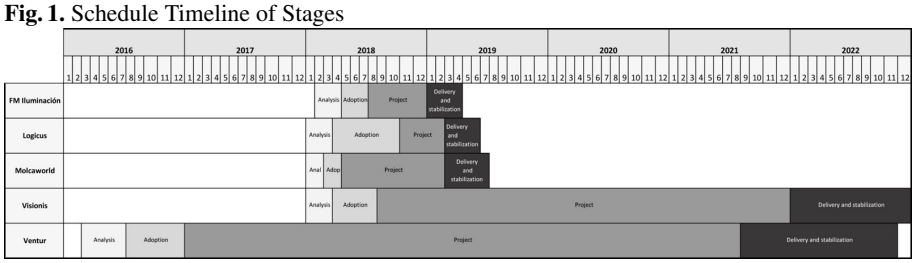
Source: own elaboration

Table 5. Characteristics of the respondents

Company	Position	Years of experience	Education
FM Iluminación, S.L.U.	CEO	20	Degree in Architecture
FM Iluminación, S.L.U.	Administration Manager	21	Degree in Business Adminitration
Logicus Engineering, S.L.U.	CEO & Founder	15	Industrial Engineer
Molcaworld, S.L.U.	Strategic Account Manager	12	Degree in Labour Relations
Visionis Distribución, S.L.	Deputy Director	3	Degree in Business Administration, MBA (Master of Business Administration) and Master of Innovation
Internacional Ventur, S.A.	IT and Logistics Manager	22	Degree in Computer Management

Source: own elaboration

Fig. 1. Schedule Timeline of Stages



Source: own elaboration

The administration manager of FM Iluminación states that:

"The technological changes of recent years mean that the company's ERP has to be able to evolve and adapt to these changes".

The need to adapt to new environmental changes and stakeholder needs is also the main driving factor for ERP adoption in the case of Visionis, according to the statements of its deputy director:

"We want the ERP not only to cover current needs, but also future needs, taking a long-term perspective".

Additionally, the company is committed to entering new marketing channels, all framed within a strategic plan for expansion and growth of the company. According to the CEO:

"We want to enter online marketing through our own platform. Having a new ERP with modules that allow us to do e-commerce is a fundamental aspect".

On the other hand, the use of unstructured tools generates problems for companies. This has been the main reason for ERP adoption, according to the CEO of Logicus:

"We have had problems such as not having the requested material from suppliers when we need them, due to errors in the delivery dates of these products, as a result of poor information management in the spreadsheet. With Excel sheets, more errors are made, and work is duplicated, unlike what would happen if we worked with an integrated tool".

Companies face new challenges that require the optimization of their processes, making it necessary to redefine business and process strategies. At the same time, they need to improve the way they collect and manage information, to ensure quicker, more efficient decision-making. This aspect is considered of vital importance by the IT manager of Ventur:

"We want to work to achieve a single data point, with a central, reliable, and efficient information centre for decision-making; this aspect has currently become a problem. The ERP can help us make decisions with a unique and irrefutable data point, a characteristic that should define a good ERP".

Moving on to analyse the important success factors in this first **Analysis** stage, the interviewees agree on the importance of having **"Clear goals and objectives"** (F7).

Specifically, the administration manager of FM Iluminación points out:

"With the information we gather from various work meetings, we analyse the needs that the new system should cover".

Along the same lines, the strategic accounts manager of Molcaworld explains:

"We must conduct a good analysis, properly understand what we need, in order to select the most appropriate ERP".

The deputy director of Visionis confirms the importance of this factor in his statements:

"During the last year, everything that has been modified in the current ERP has been compiled into a document, and all needs have been collected".

The IT and logistics manager of Ventur also emphasizes the importance of a good analysis of the company's situation and objectives to perform a proper diagnosis.

"Top management support" (F1), providing resources, global vision, and authority, is another determining factor in this first stage, as demonstrated by the statements of the interviewed companies' representatives. Specifically, the CEO of FM Iluminación notes:

"As the top executive of the company, I have directly participated in almost the entire project. I honestly believe that this should always be the case, and that the company's management should be fully involved for the benefit of the project".

The strategic accounts manager of Molcaworld also notes its importance:

"The CEO has participated in some of our meetings and has always been informed of all the aspects discussed. He has also agreed on the need to incorporate an ERP software solution and integrate a tool for the company's information system".

In terms of the human element, the role of the **"Steering Committee"** (F19) factor is also corroborated in these interviews. Specifically, the administration manager and CEO of FM Iluminación state:

"We have considered all of our employees' opinions, although the main core consisted of the two of us, along with the production manager, the purchasing manager, and the deputy director".

The IT and logistics manager of Ventur highlights the importance of having a steering committee formed by people who have a broad knowledge of the organization, its needs, and its strategic plan:

"The group of people who have carried out the analysis of needs and objectives are the members of the executive committee, that is, the heads of each department, plus the CEO of the company".

Other cross-cutting issues related to human resources highlighted by the interviewed representatives are the relationships among all the people in the organization, included in the **"Interdepartmental communication"** (F8) and **"Interdepartmental cooperation"** (F9) factors. The administration manager and CEO of FM Iluminación indicate:

"Numerous meetings have been held with the rest of the workers to see the relationships between them, as far as information system needs are concerned".

Additionally, the strategic accounts of Molcaworld points out that:

"It has been important for everyone to participate, in order to collect as much information as possible and thus share it among all of us".

Likewise, the IT manager of Ventur states:

"In order to gather the maximum amount of information, we held some meetings with department heads, as well as occasional meetings in which other department members participated".

The deputy director of Visionis also confirms the importance of these factors:

"Maximum participation and information from all company personnel is very important".

The final factor highlighted in this stage is the **"Project champion"** (F17). The administration manager and CEO of FM Iluminación speak to its importance:

"We were aware that we had to make the change sooner or later, and both the administration manager and I have led this process".

According to the strategic accounts manager of Molcaworld, it is another key element:

"The person in charge of leading the project has been the procurement manager. As with any project, it is important to have someone responsible for it, to guide and push it in the right direction".

The deputy director of Visionis declares that the project needs to be headed up by someone who is both a leader and an initiator:

"It is necessary for the person who leads and coordinates the project to know all the organization's processes very well and to rely on the rest of the people".

5.2. Adoption Stage

This process includes analysing the various potential ERP solutions, contacting companies that develop and implement these solutions, and making the final decision on which one to select.

In this second stage, the interviewees highlight the **"Careful selection of the appropriate package"** (F13) as crucial. Specifically, the two interviewees from FM Iluminación emphasize the need for thorough research and analysis:

"We started by searching for information and evaluating several programs, but we discarded many of them because they did not meet our needs. We had meetings with several software providers; we first explained our company and what we expected from the new ERP, and they gave us demonstrations of the computer applications they distribute".

Regarding the process of analysing and selecting the IT solution, the CEO and founder of Logicus emphasizes that:

"This process is long and delicate; the decision about which solution is the most appropriate is very important and can be considered strategic for the company".

The enormous effort dedicated to this search and analysis process is also evident in the words of the IT and logistics manager at Ventur:

"We contacted all the software suppliers, explained our needs, and told them about our business model and the internal workings of our organization. Each of them gave us a demonstration, showing us how their ERP could help us achieve our goals".

Closely related to the previous factor, we observe the relevance of the **"Vendor / customer partnerships"** (F3) factor, highlighted by the different interviewees as a fundamental factor in the development of any new information system implementation project. Thus, the IT manager of Ventur points out that:

"We value the partner very much, we are talking about ERPs that are all leaders in the field. We probably wouldn't have gone wrong with choosing any of the solutions; the really important thing is to choose a good travel companion".

A good understanding between the company that will implement the ERP and the ERP provider is also essential for the successful development of the process in Visionis. According to the deputy director:

"The physical proximity and good references we received about out partner Odoo were another determining factor in our decision, but above all, the close relationship and good rapport we have had from the beginning with the people of this company. For us, the human element is fundamental".

The strategic accounts manager at Molcaworld describes the relationship with the ERP provider as strategic and long-term:

"This is a long-term relationship. The ERP provider has to be a strategic supplier for us and will certainly help us become much more efficient".

As in the first stage, the interviewed companies consider factors related to human resources essential in this second stage.

The CEO of FM Iluminación highlights the importance of the factor **"Top management support"** (F1):

"I have always actively participated in the implementation of the ERP, as I understand that the company's management should do".

In addition, the deputy director of Visionis recognizes the prominent role of the **"Steering committee"** (F19) factor:

"The directors of all departments have participated in the entire process, attending internal meetings and meetings with ERP distributors".

At the same time, this interviewee recognizes the importance of the work done between departments and, consequently, the importance of the factors **"Interdepartmental cooperation"** (F9) and **"Interdepartmental communication"** (F8):

"The people who make up the different departments have participated in the entire process of selecting the new ERP; their contribution has been very valuable. A meeting was held with practically everybody, explaining what was going to be done, to make everyone aware of what the change of the ERP was going to entail. At the same time, we requested their collaboration and patience, preparing people for the change. It is good for people to be informed of what we are going to do".

The **"Project champion"** (F17) is also recognized as one of the outstanding factors by FM Iluminación. The two interviewees from the company agree that:

"It was the administration manager who led the project, participating in all meetings, coordinating, and encouraging all members of the company".

5.3. Project Stage

The **Project** stage comprises the activities aimed at putting the information system into operation in one or more organizational units; that is, it is the stage prior to the implementation of the ERP in the company as a whole. This is the stage where we find the largest number of factors deemed relevant by the interviewed companies, which is a reflection of its complexity.

The team designated to this stage and their associated responsibilities are critical aspects included in the **"Project management"** (F18) factor, the importance of which was corroborated by the CEO of Logicus:

"The team in charge of coordination and implementation is a fundamental piece for the success of the project".

"Software development, testing and troubleshooting" (F24) of the project is another crucial aspect in this stage, as indicated by the strategic accounts manager of Molcaworld:

"It is important to carry out scheduled monitoring on a weekly basis, depending on the degree of urgency of the needs".

"The use of consultants" (F4) is another factor that the same person interviewed recognizes as decisive:

"The help and guidance of a consultant has been very important for us, due to their knowledge of the tool".

The IT and logistics manager of Ventur also emphasizes the importance of having the support of expert ERP consultants to guide them in the development of the project. In fact, they consider that changing consultants during the **Project** stage was one of the causes of the delay in implementing the ERP:

"The ERP consultants are responsible for directing the project. They help us in the entire process of implementation, and are essential in this stage. From the beginning of

the project, we were working with a consultant with whom we had already established work guidelines; they had extensive knowledge of all our needs and the functioning of the company. The consulting company notified us of a change in the person who was going to take charge of our project, which had a negative impact on the dynamics of project management and made us feel we were starting over. Without a doubt, this change was one of the causes of the delay in implementing the ERP".

Logically, the process of implementing the new ERP also affects internal work processes. The company must ensure that its business processes and needs are aligned with the ERP system being implemented, adapting and/or restructuring the organization's processes with the new system to make them as efficient as possible. Hence the importance of the **"Business process reengineering"** (F15) factor. Specifically, the CEO and the administration manager of FM Iluminación point out that:

"At the same time as we wanted to adapt the software to our way of working, we wanted to see what possibilities the new system offered us to improve our procedures".

This intention fully coincides with that of Logicus, as indicated by its CEO— even more so in their case, as they had not previously worked with an ERP:

"We were not working with an ERP system, and we wanted to take advantage of the fact that we were going to incorporate such a system to review our procedures and make changes regarding our internal work processes".

According to the deputy director of Visionis, ERP implementation is inextricably linked with **"Business process reengineering"** (F15) factor.

"In this stage, we are reviewing all the processes we carry out in the organization and restructuring them, with the aim of simplifying them and making them more efficient".

Therefore, this process requires special attention to be paid to the elements intended for cultural and organizational change management, understanding it as a process that encompasses the entire organization, all of which is captured in the **"Change management"** (F6) factor. According to the statements of the CEO of FM Iluminación:

"We all have to be able to adapt to changes. There is always resistance to change, more from some people than from others, but we must overcome that resistance and adapt to better procedures for our tasks".

Visionis deputy directors also pointed out:

"Another noteworthy challenge is changing management. Addressing the scope of the changes to be made, their impact on the organization and on workers requires a considerable management effort".

The changes defined in organizations for the ERP implementation process must also be developed with a holistic vision of the organization itself. The strategic account manager at Molcaworld highlights the importance of this factor by pointing out that:

"We cannot focus our efforts on just one department, leaving aside any other areas; they are all interconnected".

During this stage, certain necessary adaptations and customizations must be made in the ERP. In addition, to ensure that its operation meets the company's requirements, numerous tests and continuous checks are necessary. The CEO of Logicus emphasizes the importance of the **"Software development, testing, and troubleshooting"** (F24) factor:

"Continuous testing has been carried out with the different adaptations and developments required. This continuous and constant process has allowed us to settle the new system well before making the final switch".

For his part, the IT and logistics manager at Ventur points out that companies agree to make use of the original functionalities of the ERP, as customization entails an increase in cost, time, and, in some cases, may incur other problems relating to the evolution and updating of the system. In this respect, the **"Minimum customization"** (F20) factor is valued as crucial by different interviewed companies. Specifically, the CEO of FM Iluminación states:

"We did not want to give into the temptation of starting to program everything; the less programming we had to do, the better, but I think it is necessary to be able to make certain adaptations that, due to our sector and our way of working, are essential for our efficiency".

Similarly, the CEO of Logicus also emphasizes the importance of this factor:

"Our intention has always been for the investment to be controlled and to minimize the economic impact. As much as possible, we have tried to avoid customized programming; we wanted to adapt, to a greater extent, to the standard".

"Education on new business processes" (F23) s also an essential component in this stage, referring to both the changes made to organizational procedures and the new information system. The administration manager of FM Iluminación makes this clear:

"Everyone has been trained on the changes in internal procedures, and this training has been carried out gradually".

The changes in internal processes and the reformulation of work procedures have involved specific training for this purpose, as the CEO of Logicus tells us:

"It has also been important to explain and train all employees of the company on the changes in internal processes that we have made".

The deputy director of Visionis fully agrees with the other interviewees:

"The staff requires appropriate training, as we have restructured many of the organization's processes".

The strategic accounts manager of Molcaworld comments that, in this stage, employees have been trained on both the operation of the new information system and the specific changes made to the processes resulting from the implementation. She also adds that:

"Everyone has participated in the training. Training has been made available to everyone and remains available and will continue to be so. We must train ourselves by getting to know all the devices that the program offers to make it more effective and efficient. Daily life consumes us. We must make sure that the program is not an obstacle but a very easy tool. I believe that training never ends; it is an evolution. Other changes will come, such as new versions, and that adaptive and evolutionary mentality must be up to date".

This same interviewee recognizes the importance of the factors **"Education on new business processes"** (F23) and **"User training and education"** (F22).

At Logicus, the CEO himself has led this stage, recognizing the importance of the factor **"Top management support"** (F1):

"I also performed the coordination function for everyone, both people from my own company and from the external consulting team. I understand that it is important to be directly involved".

The meetings held by the **"Steering committee"** (F19) at Molcaworld serve to ensure the project's smooth development, as the strategic accounts manager explains:

"The heads of each department created a list of objectives and functions, so that the program could be implemented through the work of those reporting to them".

The administration manager of FM Iluminación also confirms the importance of the factors **"Interdepartmental cooperation"** (F9) and **"Interdepartmental communication"** (F8):

"There are not many of us in the company, but we have all participated. Communication and collaboration between departments have been especially important since all the areas are interconnected".

In addition, this company states that the administration manager was the person in charge of leading the project, highlighting the importance of the **"Project management"** (F18) factor in this stage:

"In the meetings, both the CEO and I were present, and this is a very important aspect in project management. There should always be someone in charge of leading it, and that responsibility fell on me".

The IT and logistics manager at Ventur also point outs that:

"Those in charge of leading this type of project must be able to create expectations and generate the necessary motivation for the successful completion of the project, and to achieve this, the management of all the people involved is key".

5.4. Delivery and Stabilization Stage

This stage begins when the entire company starts working with the new ERP, that is, it begins to use the new system to manage the information of its business processes. The start of this stage is one of the most critical moments of the ERP implementation process, as all the work done in the previous stage is put into practice, and the success or failure of the project largely depends on this moment.

In this context, and following the work carried out in the previous stage, the importance of the **"Software development, testing, and troubleshooting"** (F24) factor is verified. Specifically, the strategic account sales manager at Molcaworld explains that the focus in this fourth stage is on resolving problems that were not detected in the previous stage, and then making adjustments or modifications to the software development:

"In this start-up stage, we have mainly been correcting detected issues, not so much developing the software itself, as that was the task we completed in the previous stage".

The support of consultants in this stage, as captured in the **"Use of consultants"** (F4) factor, is identified as a core factor by the Logicus CEO:

"The help and support of the Odoo technical consulting team has been essential for the launch of the new application".

The project manager at Ventur also confirms this:

"The start of operations with the new information system generates a lot of tension and uncertainty. The first few days were chaos. We had many fronts to turn to and crises to solve. The consultants determined the order and type of action in each case. Under their direction, we managed to prevent the project from failing".

Continuous learning, both about the new ERP system and the organizational procedure changes, are two aspects highlighted by the companies. Therefore, the **"User training and education"** (F22) and **"Education on new business processes"** (F23) factors are decisive in this stage, as evidenced by the statements of the top executive at Logicus:

"Training continues to be very important, even in this stage, for the good functioning of the system. We have made changes in our way of working and we need to know the new application well, as well as assimilate these changes properly."

The statements from the IT and logistics manager of Ventur endorse the importance of these two factors:

"We knew that training on the ERP and the changes we had introduced should continue during the implementation stage; in this way, the staff was able to adapt much better".

The deployment of an ERP tool requires the participation of top management and its team, making the importance of **"Top management support"** (F1) and **"Steering committee"** (F19) key factors in this stage. The strategic account manager of Molcaworld underlines this:

"Everything has remained the same as in the previous stage. There has been direct participation from top management and all departmental managers, as well as from the company's own personnel".

The deputy director of Visionis also confirms this:

"In this stage, all departmental managers, executive management, and company ownership have come together. It was the same in the other stagestages, but this one required greater dedication".

The administration manager and CEO of FM Iluminación also agree on the importance of **"Interdepartmental cooperation"** (F9) and **"Interdepartmental communication"** (F8):

"The project team was made up of the same people as in previous stages, but the work dynamics were much more intense than in the previous stage. Although we coordinated the work and tasks to be carried out, most of the time we were working with other people from our company, especially the procurement and manufacturing departmental managers".

The deputy director of Visionis notes the salient role of both factors, in his experience:

"During the start-up stage of the new ERP, an extra effort is required from all company personnel, who must work as a team".

The **"Project management"** (F18) also reveals his or her importance in this stage, as reflected in the words of the strategic account manager of Molcaworld:

"As in the previous stage, I was responsible for leading the project. It may be even more important in this stage, as problems emerge again and there is a lot of tension. In addition, time is of the essence when solving these problems, requiring an extra effort in decision-making and coordination of all teams."

For his part, the IT and logistics manager of Ventur declares:

"In this stage, I led the project, maintaining a constant relationship with the consultants and with the rest of my colleagues (of the steering committee). The project requires the figure of the leader to ensure that it develops smoothly".

6. Conclusions and future lines of research

6.1. Academic implications

The COVID-19 pandemic has raised awareness of the main social and economic roles of firms. Furthermore, the pandemic has motivated organizations to look for analytical tools that allow them to better understand the environmental forces affecting them and control

their impact on business operations ([16]; [45]). According to a recent report published by European (2022) [46], there has been a continuing rise in the number of companies, both large and small, adopting ICT solutions and specifically ERP systems to facilitate their management. This demonstrates the importance of strategic information management for the competitiveness of companies ([10]; [5]; [47]), especially in today's dynamic and globalized environment.

However, organizational processes in companies are becoming increasingly complex and interconnected, which makes ERP implementations more challenging ([10]). It is important to address the study of ERP systems holistically, considering the intangible resources that allow the company to successfully implement these systems, thus enabling progress towards more flexible and competitive business models. This study aims to analyse the factors that are critical to the success of each stage of an ERP implementation, through in-depth case studies of five companies. This work proposes a conceptualization of the stages of the ERP implementation life cycle by drawing on a comprehensive review of ERP life cycle models. This review confirms that there is no clear consensus on the specific stages that make up the life cycle, as demonstrated by the diversity of existing models. The lack of consensus hinders researchers' progress on identifying the activities and factors that ensure a successful ERP implementation.

Therefore, one of the objectives of this work has been to develop a model of the ERP life cycle in five stages, based on an in-depth analysis of the two most widely adopted, highly valued, and commonly cited models by the research community; namely, those of Markus and Tanis (2000) [27] and Cooper and Zmud (1990) [30].

This model proposed in this study highlights the strategic planning stage, since any information system must be integrated into corporate strategy ([48]; [49]), which may entail significant changes in corporate culture ([21]; [50]).

The proposed model of the ERP life cycle in five stages contributes to the academic field by providing a structured framework for understanding and analyzing ERP implementation. It addresses the lack of consensus in existing models, paving the way for more standardized research methodologies in this area. Researchers can utilize this model as a foundation for further studies on ERP systems, allowing for more comparable and consistent results.

Once these conceptual foundations of the ERP implementation model have been established, it is essential to understand and identify the Critical Success Factors (CSFs) in the implementation process. To this end, we reviewed and extended the study of the CSFs by Somers and Nelson (2001) [34] , eventually identifying a total of 25 factors considered relevant in the existing literature, comprising organizational, technological, project and individual factors.

In order to analyse the expected impact of each factor on the success of ERP implementation in each of its stages, a quantitative study was conducted, relying on a panel of information systems experts from the academic and professional fields. The results allowed us to understand the overall impact of each factor on the success of ERP implementation, as well as its contribution to each specific stage of its life cycle. The findings from the expert panel served as the basis for the subsequent qualitative in-depth study of five companies that were initiating the ERP implementation process. We were able to dynamically analyse the contingencies and CSFs of this process, thereby fulfilling the main research objective.

The identification of Critical Success Factors (CSFs) and their impact on each stage of the ERP life cycle adds valuable insights to the academic community. This detailed analysis provides a basis for future research on the specific factors influencing successful ERP implementation, facilitating a deeper understanding of the nuances involved.

The combination of quantitative and qualitative methods in this study offers a comprehensive approach that can serve as a model for future research in the field of ERP implementation. The integration of expert opinions with real-world case studies enriches the academic understanding of ERP dynamics, creating a robust basis for further exploration and refinement of ERP life cycle models. Below, we discuss the main practical implications for business management.

6.2. Practical implications

The interviews conducted with those responsible for the preparatory analysis and implementation of ERP in five companies underscore the importance of a **"Clear goals and objectives"** (F7) that the ERP must address. It is necessary to define this purpose in the first stage of the ERP life cycle and it must be understood strategically, aligning with the company's objectives ([48]; [49]).

The need to establish project objectives is clearly corroborated by the empirical study, where we can see that one of the reasons why two of the five companies have not yet been able to complete the implementation process is inadequate definition of objectives in the first stage, generating subsequent delays in the process and overrunning the initially estimated period.

Another element that we identify in the study as crucial to the success of ERP implementation projects is the relevance of the role played by **"Use of consultants"** (F4), and the need for a good **"Interdepartmental communication"** (F8) between them and the company's executives ([36]; [40]; [35]). This result confirms previous studies such as Osman (2018) [36], Maditinos et al. (2011) [3] and Finney and Corbett (2007) [22] on the importance of consultants' technical knowledge and expertise in helping users to implement and upgrade [39] a new ERP system. Therefore, companies should consider the attention paid to consultant selection in specific business contexts as a future investment or guarantee of ERP implementation success. Consultants not only have technical skills required but also genuine expertise on best practices in a business context.

Therefore, the role played by consultants, and the **"Interdepartmental cooperation"** (F8) between them and the company's employees that are involved in the ERP implementation, is crucial. Consequently, poor consultant performance or problems in the coordination between the two parties can generate serious issues that considerably hinder the ability to achieve ERP project success.

Effective **"Interdepartmental communication"** (F8) and **"Interdepartmental cooperation"** (F9) between different areas or departments of the company is also highlighted as a key factor in transferring knowledge and building solid internal capabilities related to ERP usage and post-implementation [39]. Building these capabilities will ensure that the how-how provided by consultants is properly institutionalized throughout the organization. This result is consistent with previous studies such as Cho and Lee (2024) [12], Maditinos et al. (2011) [3] and Wu and Wang (2007) [51].). In fact, the more dynamic the industry in which a company competes, the greater the need for it to enhance its

technological capabilities through internal learning and cooperation with other companies [12].

Another key practical implication of the work is the importance of adequate **"Dedicated resources"** (F5) by the company, mainly human capital resources. The company's employees who participate in the project play a substantial role in this process ([39]; [52]), and their satisfaction significantly influences the success of ERP implementation ([51]; [53]). Individual participation is relevant, but the participation of the **"Steering committee"** (F19), made up of top executives from different functional areas, project managers, and end-users of the ERP, is particularly significant, as demonstrated by the results of the empirical study. This result is consistent with previous studies such as Gill et al. (2020) [52] , Osman (2018) [36], and Nagpal et al. (2017) [35].

Companies need to ensure participative and proactive management of human resources, designing a **"User training and education"** (F22) plan for the personnel responsible for the implementation of the ERP, as well as promoting structures that encourage employee participation in knowledge exchange ([54]; [3]). In order to shorten the learning cycle and enhance productivity, new users should be taught both the technical aspects of the newly implemented ERP and their new organizational responsibilities ([3])), before, during and after ERP implementation [39].

In this respect, ERP implementation requires continuous support from the organization's management, departmental directors, and the leader of the strategic project itself. These findings highlight the importance of **"Top management support"** (F1) as key determinants of the success of ERP implementation. The support of top management is needed to drive a radical shift in the organization's culture, in order to encourage the investment in new technologies to share knowledge and expertise across all levels of the organization ([16]; [12]).

We have also been able to demonstrate the relevance of specific tasks related to **"Software development, testing and troubleshooting"** (F24) during the third stage of the ERP life cycle ([39]; [36]; [37]). Companies should understand the implementation process of an ERP as a dynamic, ongoing process since the system must be regularly reviewed and upgraded, and adapted to changes in user/system requirements [55]. The continuous deployment of ERP ensures that the system is aligned with both changes in business strategy and objectives, and changes in the competitive and technological arenas.

The acknowledgment of the impact of external shocks, such as the COVID-19 pandemic, underscores the importance of building resilience into ERP strategies. Businesses should, thus, proactively design ERP systems that can adapt to unforeseen disruptions and changes in the external environment.

The acknowledgment of the impact of external shocks, such as the COVID-19 pandemic, underscores the importance of building resilience into ERP strategies. These disruptions have highlighted the fragility of systems that fail to adapt quickly to unforeseen circumstances, leading to operational inefficiencies and even failure. Therefore, businesses should proactively design ERP systems that can adapt to such disruptions and changes in the external environment. A resilient ERP system must be agile, flexible, and scalable, enabling organizations to pivot swiftly when confronted with challenges such as supply chain interruptions, shifts in consumer demand, or changes in regulatory frameworks, because, as shown by the results of this study and previous research by Choo and

Lee (2024) [12], internal and external factors are equally important for a successful ERP implementation.

6.3. Future lines of research

Our study has focused on five SMEs located in the Valencian Community with different characteristics related to the sector of activity, size, age, and business volume, among others, which clearly condition the ERP requirements. We believe it is appropriate to extend this work to companies in other sectors, including service sector companies, as well as other geographical areas outside the Valencian Community.

Conducting a longitudinal study to track the evolution of companies post-implementation offers an opportunity to explore the long-term impacts and challenges associated with ERP systems. This could involve a more in-depth examination of the Continuity and Improvement stage, providing valuable insights into sustained success and areas for enhancement.

The interviews with the companies were conducted with people directly involved in the implementation projects, who held positions of responsibility and leadership in the organizations. In this respect, we believe it would be interesting to compare these responses with the opinions of other employees who have participated in the process, both those responsible for some areas or departments and end-users of the ERP.

Another central element, also reflected in our conclusions, is the importance of adequate human resource management, due to its impact on the success of ERP implementation ([52]). We believe it is necessary to analyse management and organizational leadership styles that allow the creation of collaborative and participatory environments conducive to ERP implementation. Understanding how leadership styles impact the management of human resources in the context of ERP implementation can contribute to best practices for successful project execution.

It would also be interesting to analyse how Industry 4.0 technologies—including IoT, artificial intelligence, big data analytics, and cyberphysical systems—influence the future of ERP systems and can help to create more efficient, innovative and interconnected business processes [56]. Additionally, future research on ERP systems should also focus on studying new fields that stand to benefit from the proper management and implementation of these systems, such as the circular economy, green supply chains and sustainability ([57]; [58]).

In line with the research by Kahraman and Bicen (2022) [45], future studies should delve deeper into the specific conditions and demands imposed by the COVID-19 pandemic on ERP systems, such as those related to the development of new technological competences by employees and new ways of organizing work. Exploring how organizations can adapt their ERP strategies to address the challenges posed by global disruptions enhances the relevance of ERP research in the context of rapidly changing external environments.

References

1. A. Ullah, R.B. Baharun, K. Nor, and M. Yasir. Overview of enterprise resource planning (erp) system in higher education institutions (heis). *Advanced Science Letters*, 24(6):4399–4406, 2018.

2. S.H. Chung and C.A. Snyder. Erp adoption: a technological evolution approach. *International Journal of Agile Management Systems*, 2(1):24–32, 2000.
3. D. Maditinos, D.and Chatzoudes and C. Tsairidis. Factors affecting erp system implementation effectiveness. *Journal of Enterprise Information Management*, 25(1):60–78, 2011.
4. A. Daviy. Does the regional environment matter in erp system adoption? evidence from russia. *Journal of Enterprise Information Management*, 36(2):437–458, 2023.
5. S.F. Wamba, A. Gunasekaran, S. Akter, S.J.F. Ren, R. Dubey, and S.J. Childe. Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70:356–365, 2017.
6. S. Mithas and R.T. Rust. To use or not to use: Modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system. *Mis Quarterly*, 40(1):223–246, 2016.
7. M. Al-Mashari, A. Al-Mudimigh, and M Zairi. Enterprise resource planning: A taxonomy of critical factors. *European Journal of Operational Research*, 146(2):352–364, 2003.
8. L.M. Hitt, D.J. Wu, and X. Zhou. Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems*, 19(1):71–98, 2002.
9. T.H. Davenport. *Mission critical: realizing the promise of enterprise systems*. Harvard Business Press, Boston, USA, 2000.
10. V.K. Ranjan, S.and Jha and 2016 Pal, P. Literature review on erp implementation challenges. *International Journal of Business Information Systems*, 21(3):388–402, 2016.
11. S. Rouhani and M. Mehri. Empowering benefits of erp systems implementation: empirical study of industrial firms. *Journal of Systems and Information Technology*, 20(1):54–72, 2018.
12. Y. Cho and C. Lee. The effects of process innovation and partnership in scm: Focusing on the mediating roles. *Computer Science and Information Systems*, 21(2):453–472, 2024.
13. P.J. Shyiu, K. Singh, J. Kokkranikal, R. Bharadwaj, S. Rai, and J. Antony. Service quality and customer satisfaction in hospitality, leisure, sport and tourism: an assessment of research in web of science. *Journal of Quality Assurance in Hospitality & Tourism*, 24(1):24–50, 2023.
14. M.F. Acar, M. Tarim, H. Zaim, S. Zaim, and Delen. Knowledge management and erp: Complementary or contradictory? *International Journal of Information Management*, 37(6):703–712, 2017.
15. C. Yang and Y.F. Su. The relationship between benefits of erp systems implementation and its impacts on firm performance of scm. *Journal of Enterprise Information Management*, 22(6):722–752, 2009.
16. N. Kumar, G. Kumar, and R.K. Singh. Analysis of barriers intensity for investment in big data analytics for sustainable manufacturing operations in post-covid-19 pandemic era. *Journal of Enterprise Information Management*, 35(1):179–213, 2022.
17. M. Ali and L. Miller. Erp system implementation in large enterprises–a systematic literature review. *Journal of Enterprise Information Management*, 30(4):666–692, 2017.
18. I. Mahmud, T. Ramayah, and S. Kurnia. To use or not to use: Modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system. *Information Systems*, 29:164–179, 2017.
19. Z. Shao, T. Wang, and Y. Feng. Impact of organizational culture and computer self-efficacy on knowledge sharing. *Industrial Management & Data Systems*, 115(4):590–611, 2015.
20. J. Ram and D. Corkindale. How "critical" are the critical success factors (csfs)? examining the role of csfs for erp. *Business Process Management Journal*, 20(1):151–174, 2014.
21. P. Garg and A. Garg. Factors influencing erp implementation in retail sector: an empirical study from india. *Journal of Enterprise Information Management*, 27(4):424–448, 2014.
22. S. Finney and M. Corbett. Erp implementation: a compilation and analysis of critical success factors. *Business Process Management Journal*, 13(3):329–347, 2007.
23. R.G. Saade and H. Nijher. Critical success factors in enterprise resource planning implementation: A review of case studies. *Journal of Enterprise Information Management*, 29(1):72–96, 2016.

24. R. Plant and L. Willcocks. Critical success factors in international erp implementations: a case research approach. *Journal of Computer Information Systems*, 47(3):60–70, 2007.
25. C. Doom, K. Milis, S. Poelmans, and E. Bloemen. Critical success factors for erp implementations in belgian smes. *Journal of Enterprise Information Management*, 23(3):378–406, 2010.
26. T.C. Loh and S.C.L. Koh. Critical elements for a successful enterprise resource planning implementation in small-and medium-sized enterprises. *International Journal of Production Research*, 42(17):3433–3455, 2004.
27. M.L. Markus and Tanis. *The enterprise systems experience-from adoption to success. Framing the Domains of IT Research: Glimpsing the Future through the Past*. Pinnaflex, Cincinnati, OH, 2000.
28. J.W. Ross and M.R. Vitale. The erp revolution: surviving vs. thriving. *Information Systems Frontiers*, 2:233–241, 2000.
29. J. Esteves and J.A. Pastor. Organizational and technological critical success factors behavior along the erp implementation phases. In *Enterprise Information Systems VI*, pages 63–71, Dordrecht, Netherlands, 2006. Springer.
30. R.B. Cooper and R.W. Zmud. Information technology implementation research: a technological diffusion approach. *Management Science*, 36(2):123–139, 1990.
31. G. Shanks. A model of erp project implementation. *Journal of Information Technology*, 15(4):289–303, 2000.
32. M. H. Pérez Álvarez, L. Parody, R. Gómez-López, M. T.and Gasca, and P. Ceravolo. Decision-making support for input data in business processes according to former instances. *Computer Science and Information Systems*, 18(3):835–865, 2021.
33. M. Haddara. Erp selection: the smart way. *Procedia Technology*, 16:394–403, 2020.
34. J. Esteves and J.A. Pastor. The impact of critical success factors across the stages of enterprise resource planning implementations. In *In Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, page 10, Hawaii, 2001. IEEE.
35. S. Nagpal, A. Kumar, and S.K. Khatri. Modeling interrelationships between csf in erp implementations: total ism and micmac approach. *International Journal of System Assurance Engineering and Management*, 8:782–798, 2017.
36. 2018 Osman, N. A software requirement engineering framework to enhance critical success factors for erp implementation. *International Journal of Computer Applications*, 180(10):32, 2018.
37. E. Reitsma and 2018 Hilletofth, P. Critical success factors for erp system implementation: A user perspective. *European Business Review*, 30(3):285–310, 2018.
38. A.S. Shatat and N. Dana. Critical success factors across the stages of erp system implementation in sohar university: A case study. *International journal of management and applied research*, 3(1):30–47, 2016.
39. Z. T. Butarbutar, P. W. Handayani, R. R. Suryono, and W. S. Wibowo. Systematic literature review of critical success factors on enterprise resource planning post implementation. *Cogent Business and Management*, 10(3), 2023.
40. A. Ahmed, A.A. Shaikh, and M. Sarim. Critical success factors plays a vital role in erp implementation in developing countries: an exploratory study in pakistan. *International Journal of Advanced Computer Science and Applications*, 8(10), 2017.
41. M. Ayat, M. Imran, A. Ullah, and C. W. Kang. Current trends analysis and prioritization of success factors: a systematic literature review of ict projects. *International journal of managing projects in business*, 14(3):652–679, 2021.
42. J.M. Stanton and S.G Rogelberg. Using internet/intranet web pages to collect organizational research data. *Organizational Research Methods*, 4(3):200–217, 2001.
43. F.N. Kerlinger. *Foundation of Behavioural Research*. Holt. Rienhart & Winston, New York, USA, 1986.
44. C. Robson. *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*. Blackwell, Oxford, England, 1993.

45. A. Kahraman and H. Bicen. The impact of digital transformation in teachers' professional development during the covid-19 pandemic. *Computer Science and Information Systems*, 19(3):1565–1582, 2022.
46. Union European. E-business integration. Eurostat Statistics Explained, 2022. [Online]. Available: ($https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-business_integration\&oldid=570394\#Enterprise_resource_planning.28ERP.29$).
47. T.M. Somers, Nelson, K., and J. Karimi. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13):1319–1350, 2007.
48. N.C.B. Amar and R.B. Romdhane. Organizational culture and information systems strategic alignment. *Journal of Enterprise Information Management*, 33(1):95–119, 2020.
49. L. Anaya, M. Dulaimi, and S. Abdallah. An investigation into the role of enterprise information systems in enabling business innovation. *Business Process Management Journal*, 21(4):771–790, 2015.
50. M.E. Porter and M.R. Kramer. The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12):78–92, 2006.
51. J.H. Wu and Y.M. Wang. Measuring erp success: The key-users' viewpoint of the erp to produce a viable is in the organization. *Computers in Human Behavior*, 23(3):1582–1596, 2007.
52. A.A. Gill, S. Amin, and A. Saleem. Investigation of critical factors for successful erp implementation: an exploratory study. *Journal of Business and Social Review in Emerging Economies*, 6(2):565–575, 2020.
53. T.M. Somers, Nelson, K., and J. Karimi. Confirmatory factor analysis of the end-user computing satisfaction instrument: replication within an erp domain. *Decision Sciences*, 34(3):595–621, 2003.
54. S.A. Gawankar, A. Gunasekaran, and S. Kamble. A study on investments in the big data-driven supply chain, performance measures and organisational performance in indian retail 4.0 context. *International Journal of Production Research*, 58(5):1574–1593, 2020.
55. S. Jayatilleke and R. Lai. A method of assessing rework for implementing software requirements changes. *Computer Science and Information Systems*, 18(0):32, 2021.
56. G. U. Ebirim, I. F. Unigwe, O. F. Asuzu, B. Odonkor, and U. I. Oshioste, E. E.and Okoli. A critical review of erp systems implementation in multinational corporations: trends, challenges, and future directions. *International Journal of Management and Entrepreneurship Research*, 6(2):281–295, 2024.
57. M. Pohludka, H. Stverkova, and B. Ślusarczyk. Implementation and unification of the erp system in a global company as a strategic decision for sustainable entrepreneurship. *Sustainability*, 10(8):2916, 2018.
58. C.J.C. Jabbour, A.B.L. De Sousa Jabbour, J. Sarkis, and M. Godinho Filho. Unlocking the circular economy through new business models based on large-scale data: an integrative framework and research agenda. *Technological Forecasting and Social Change*, 144:546–552, 2019.

A. Appendix

Table I. Results of the panel of experts

Factor Clave Éxito	Analysis	Adoption	Project	Delivery and stabilization	Continuity and improvement
1. Top management support	**62,10%**	**65,50%**	*48,30%*	*27,60%*	*3,40%*
2. Project team competence	*48,30%*	*44,80%*	**82,80%**	**58,60%**	*0%*
3. Interdepartmental cooperation	*34,50%*	*31,00%*	**51,70%**	**51,70%**	*3,40%*
4. Clear goals and objectives	**86,20%**	**51,70%**	*41,40%*	*13,80%*	*3,40%*
5. Project management	*17,20%*	*34,50%*	**89,70%**	*48,30%*	*3,40%*
6. Interdepartmental communication	*41,40%*	*31,00%*	**69,00%**	*48,30%*	*3,40%*
7. Management of expectations	**58,60%**	*37,90%*	*37,90%*	*48,30%*	*3,40%*
8. Project champion	**65,50%**	**62,10%**	**86,20%**	**51,70%**	*3,40%*
9. Ongoing vendor support	*13,80%*	*17,20%*	**55,20%**	**62,10%**	*3,40%*
10. Careful selection of the appropriate package	**72,40%**	**55,20%**	*31,00%*	*6,90%*	*3,40%*
11. Data analysis and conversion	*31,00%*	*24,10%*	**65,50%**	*48,30%*	*0%*
12. Dedicated resources	*44,80%*	**58,60%**	**62,10%**	*20,70%*	*3,40%*
13. Steering committee	**72,40%**	*44,80%*	*48,30%*	*31,00%*	*0%*
14. User training and education	*3,40%*	*24,10%*	*37,90%*	**79,30%**	*3,40%*
15. Education on new business processes	*13,80%*	*24,10%*	*44,80%*	**51,70%**	*3,40%*
16. Business process reengineering	**51,70%**	*24,10%*	**62,10%**	*24,10%*	*3,40%*
17. Minimal customization	*31,00%*	*6,90%*	**62,10%**	*41,40%*	*3,40%*
18. Defining the architecture	**62,10%**	*44,80%*	*44,80%*	*6,90%*	*0%*
19. Change management	*37,90%*	*31,00%*	**55,20%**	**58,60%**	*3,40%*
20. Vendor / customer partnerships	*41,40%*	*37,90%*	**55,20%**	**65,50%**	*3,40%*
21. Use of vendors' development tools	*17,20%*	*27,60%*	**65,50%**	*31,00%*	*0%*
22. Use of consultants	**58,60%**	*41,40%*	**62,10%**	*44,80%*	*0%*
23. Empowered decision-makers	*27,60%*	*48,30%*	**65,50%**	*48,30%*	*3,40%*
24. Software development, testing and troubleshooting	*10,30%*	*10,30%*	**79,30%**	**75,90%**	*0%*
25. Benchmarking, internal and external	*24,10%*	*17,20%*	*44,80%*	*31,00%*	*3,40%*

Source: own elaboration

Table II. Questionnaire

Analysis	Adoption	Project	Delivery and Stabilization
What **aspects or strategic reasons** motivate adopting or changing the existing ERP?	Have different software **solutions been evaluated?** What aspects have been key when deciding the type of ERP and the **software provider**?		How has the work of **incorporating the new ERP** into the daily operation of the company been developed? Have **problems** been found? If so, what kind of problems and how have they been resolved? Has there been any problem that could not be resolved?
In what **aspects do you think** the company will improve when the new ERP is implemented? What **improvements or benefits** will the new ERP bring regarding the following aspects? Economic, Social, Organizational, Environmental, etc. Did the company establish any type of **business plan** on the **strategic impact** of the ERP implementation? What **mechanisms of analysis or diagnosis** of the needs have been carried out to identify the strategic needs of the company?			What is the **level of company satisfaction** regarding the implementation of the new ERP? Which areas/departments have experienced more/less satisfactory implementation? What has been the **biggest challenge** in implementing the new ERP system? What would you change or improve about the whole implementation process? Has the company **reached the objectives** defined in the Analysis stage? If not, do you think these objectives were realistic and achievable? What were the objectives that have not been (fully) achieved? What are the reasons why they have not been achieved? Has the scope been changed during the implementation process? If so, what changes have been made and why?
Has **the work team** been defined for this stage? If so, which people and from which areas / departments make it up? What has the contribution of these people consisted of? How much of your working day have you dedicated to this project? Has a work schedule been defined for this task?	Has **the work team** been defined for this stage? If so, were the same group of people as in the previous stage? If not, which people and from which areas/departments make it up? How much of your working day have you dedicated to this project? Has a work schedule been defined for this task?		
Have there been any people in charge of **leading or coordinating** this stage? If so, what have been the requirements or characteristics for your designation? What roles and responsibilities have they performed? Has the person or people in charge of leading in this stage been assigned a specific role or authority?	Have there been any people in charge of **leading or coordinating** this stage? If so, were the same group of people as in the previous stage? Otherwise, what have been the requirements or characteristics for your designation? What roles and responsibilities have they performed? Has the person or people in charge of leading in this stage been assigned a specific role or authority?		
What has been the **role of the ownership and the management** of the company? What have been their functions and level of participation in this stage?			

(Continued)

Have regular meetings been held? If so, what have these meetings consisted of, what aspects have been discussed and how often have they been held?		
How and why was the **decision made to continue** to the Adoption stage?	Has the company's management **motivated and/or granted authority to workers** to make decisions in relation to the ERP implementation process? If so, what have been the motivating and authoritative elements?	
Has a **group of experts been defined** to make the decision to continue to the next stage? If so, which people have composed it? What have been their roles and responsibilities?	Will the implementation process be carried out only with **your own resources** or will you have an **external consultant**? If external resources are going to be used, have different software providers/consultants been assessed? What have been the decisive aspects when selecting the consulting company for the implementation of the ERP?	Has this stage been carried out only with your own resources or has an **external consultant been used**? If an external consultant has been used, was the consultant selected the same as in the previous stage? If the consultant has changed, what have been the reasons for the change? How was the work process with the external consultant? What have been the functions of the consultants?
Has the ERP change been promoted by the needs/demands of any ***stakeholder*** (customers, suppliers, public administration, etc.)?	Have **planning methodology and/or tools** to be used in the implementation process been assessed? Does the software provider or the implementing consultant work with any specific methodology and/or tools for its planning	**Have planning methodologies and/or tools** been used during the implementation process in this stage?
Have **internal factors,** such as the design of the organizational structure (level of centralization, departmentalization and formalization) and competitive strategy been taken into account in the strategic analysis of the ERP?	Has **the degree of customization of the ERP** been analyzed with respect to the standard that will be necessary?	**Have implementation milestones** been defined based on the Analysis stage? If so, were the objectives realistic and achievable? Were these milestones or objectives followed during this stage? Was their scope changed during the stage? If so, why?
		Has there been an adequate **allocation of resources** to this stage? Why? To which items or tasks were the financial resources allocated? Were there deviations from the budget? If so, why, and what percentage of deviation have you reached?
	Has the **information** that needs to be transferred from the current system to the new ERP been analyzed? If so, what areas/departments of the company is this information from? What volume and type of information do you want to transfer?	How has the **information been transferred** from the current system to the new ERP? If so, from what areas/departments of the company has the information been transferred and what has the process been like? Have problems been found? If so, what kind of problems and how have they been resolved? Has there been a problem that could not be resolved?
		What have been the **key departments and areas** in this stage of development? Why? What are those departments or people less involved in this stage of ERP implementation? What have been the reasons for their lesser participation?
	Has the need **to integrate the ERP** with other systems been identified?	Has the **ERP** been **integrated** with other systems?

(Continued)

	If so, what kind of integration and with what scope?	If so, has the type of integration and scope been the same as that expected and identified in the previous stages? How has the process been developed? Have problems been found? If so, what kind of problems and how have they been solved? Has there been a problem that could not be resolved?	
	Have **communication and coordination** channels been defined between all the people/departments participating in this stage? If so, what type, frequency and for what purpose?		
	Have there been **changes in the organizational / production processes** because of the adoption of the ERP? If so, which areas/departments have been affected? What have these changes consisted of? Has **training** on these new business processes been carried out? **Has a training** calendar been defined at the beginning of this stage? If so, for what period, and what people and contents made it up?		
Have the **Data Warehouse** structures influenced the strategic implementation of the ERP?	Have the **Data Warehouse** structures been considered?	Have there been important changes in the **necessary data structures** (Data Warehouse) between the old and the new ERP? If so, to what extent has the difference in data structure affected the implementation process?	
	Have **benchmarking** tasks been carried out to learn and incorporate improvement ideas, new knowledge, and best practices for ERP adoption? If so, what did these tasks consist of? Which areas of the organization and people have been responsible for carrying these tasks out?		
		Has a specific plan for **custom programming** of the software been developed? If so, what has been the process followed, the communication and coordination channels adopted, people/departments involved, type of tests and verifications done, and results obtained? Were deadlines defined? Were they met?	**Has a custom programming** of the software been adequately developed? Has more custom programming of the software been carried out than planned in the previous stage? If so, what was the reason? In which areas or departments has more custom programming been developed? If so, what has been the process followed, the communication and coordination channels adopted, people/departments involved, type of tests and verifications done, and results obtained?

Source: own elaboration

Sergio Ferrer Gilabert is associate lecturer of the Business Administration and Marketing Department of the Universitat Jaume I of Castellón. He holds a PhD in Economics and Business with an Outstanding Cum Laude qualification. His teaching lines are related to the management of business information systems. He also brings substantial experience as an expert consultant in the information technology sector, offering practical insights into the implementation of information systems in diverse organizational contexts.

Beatriz Forés is a full-time assistant professor of the Business Administration and Marketing Department of the Universitat Jaume I of Castellón. Her fields of expertise are strategic management, specifically the sources of dynamic capabilities, tourism, family firms, and industrial districts. Her work has been published in Tourism Management, International Journal of Tourism Research, Journal of Business Research, Corporate Social Responsibility and Environmental Management and Scandinavian Journal of Management, among others.

Rafael Lapiedra is Professor of Management at the University Jaume I of Castellón (Spain). He holds a PhD in Business Administration and a Master in European Business Management and Information Systems. He has been visiting professor at the Universidad Tecnológica Metropolitana of Santiago in Chile and at the London School of Economics and Political Science. He has held the position of Dean for the Faculty of Business and Law at University Jaume I. His research interests lie in information systems management and strategic alliances.

Received: February 29, 2024; Accepted: January 10, 2025.

Computer Science and Information Systems 22(3):755–781 https://doi.org/10.2298/CSIS240628016K

Image clustering using Zernike moments and self-organizing maps for gastrointestinal tract

Parminder Kaur[1], Avleen Malhi[2], and Husanbir Pannu[3]

[1] Durham University, UK
parminder.kaur@durham.ac.uk
[2] Aalto University, Finland
avleen.malhi@aalto.fi
[3] Thapar University, India
hspannu@thapar.edu

Abstract. Typically, the image features are compared to find the similarity among the images in a content-based image clustering system. However, images with high feature similarity may be different from each other in terms of semantics. Hence, this paper proposes a novel algorithm based on unsupervised neural classifier systems for in-vivo image clustering to address the semantic gap issue. The visual features are represented using Wavelet transform and Zernike moments, and a self-organizing map is utilized for the clustering of images. The algorithm-based prototype system is trained for categorizing gastral images in the respective clusters as per the similarity. The system can be used to segment images with automatic noise reduction and rotation invariances for given images. Experiments are performed on the real gastrointestinal images obtained from a known gastroenterologist, and the results using Daubechies Wavelet Transform + Zernike Moments on LUV color scheme yield 88.3% accuracy.

Keywords: Machine learning, Self-organizing maps, Zernike moments, Wavelet transforms, Gastroenterology.

1. Introduction

Automatic image analysis and segmentation is a skilled task carried out by experienced professionals. Features in an image are used to decompose and analyze the underlying anatomy by defining a mechanical and systematic procedure. Given the explosive growth of visual information, partly due to the expansion of the Web and partly due to the introduction of sophisticated and inexpensive image capture systems, there is an urgent need to develop programs that can learn to segment and annotate. Automatic segmentation and annotation systems are among the critical areas of research and development for the next decade and beyond, and machine learning will be a vital technology in developing such systems [54], [53]. Self-organizing maps (SOM) incorporated with extended fuzzy c-means clustering have been a popular method for image segmentation as studied in [3]. It has used a discrete wavelet transform for image description for edges and lines involved in contrast variation.

The objective of the proposed study is to analyze, segment, and cluster the endoscopy images such that the trained system can be helpful for gastroenterologists in problem diagnosis of the gastrointestinal tract. The *motivation* behind the current problem selection

is its complexity in terms of image feature distribution. An example of in-vivo gastral images has been shown in Fig. 1 in which an image has been analyzed using two segmentation algorithms:

1. *Region Growing* (It has been applied in [7] to segment 2D microscopy digital images of freshwater green microalgae. In this approach, the image is segmented into multiple disjoint regions (sub-regions), and then they are merged with their nearest neighboring seeded region (to grow regions) that satisfies a predefined homogeneity criterion.);
2. (b) *2D Otsu algorithm* [47] (which employs the gray level information of each pixel and its spatial correlation information within the neighborhood). The algorithm has failed to capture the region of interest in both the cases, which is bleeding and not the dark spot.

It can be observed that it is pretty challenging to accurately segment blood due to the obscure nature of the color distribution and irregular region boundary. The red and green boundaries have captured the wrong dark region instead of the red spot ROI (region of interest). Moreover, the underlying images are dynamic, involving continuous movements of the camera in the drifting capsule, body organs, insufficient light conditions to capture texture at the region of interest, and varying luminance and noise due to food particles and body fluid. In addition, complementary metal-oxide semiconductor (CMOS) image sensors involve noise, high compression ratio, and low resolution of 256×256. If a segmentation method can enhance the classification accuracy in this confounding case, then inherently, it would also contribute to other applications of image processing. This is the reason for the underlying case study about image segmentation for gastral images. Challenges involved in image retrieval have been discussed in Table 1.

Table 1. Summary of challenges of image representation and learning

Challenge	**Elaboration**
Image invariance	Yields same image, when rotated, scaled or moved.
Noise	The 'lens' of the camera is never perfect; surrounding environment may contribute to the noise, noise could be Gaussian or distributed differently.
Representation	In terms of the optical properties of the (individual) pixels of an image – mean intensity, x-tilt, y-tilt, focus astigmatism @ 0 degree & focus astigmatism @ 45 degrees, coma & x-tilt, coma & y-tilt, spherical & focus.
Learning	For recognizing the contents of a new image having "see" similar images before.

We have used wavelet resolution which helps to remove noise and makes images scale invariant. Zernike moments have been used for image vectorization and self-organizing maps based on unsupervised learning is used to cluster images for sick and healthy classes.

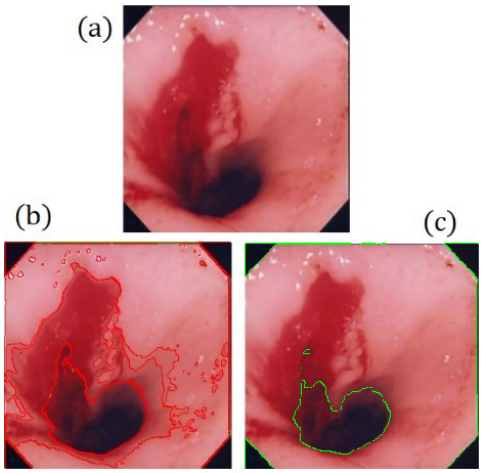


Fig. 1. (a) Original gastral image, (b) Region growing segmentation [7], and (c) 2D Otsu segmentation algorithm [47]. The red and green boundaries have captured the wrong (dark) spot instead of red region of interest (ROI) showing that problem is complex for image segmentation

Overview

The overview of the proposed research is as follows:

1. A novel algorithm for medical image clustering has been proposed which is based on unsupervised neural classifier systems.
2. The characteristic visual features are obtained from the images using Wavelet Transforms (WT), Zernike Moments (ZM), and Kohonen self-organizing feature map algorithm has been applied for clustering.
3. The proposed image clustering approach has been applied to the real capsule endoscopy images obtained from a known gastroenterologist and data distribution has been carefully studied using PCA and LDA plots to motivate the application of advanced machine learning techniques.
4. Performance analysis of the proof-of-concept model has been compared with both traditional and contemporary methods to support the belief.

This paper introduces an efficient image segmentation algorithm using Wavelet Transforms, Zernike Moments, and Linear Discriminant Analysis due to their characteristic visual feature extration and then unsupervised clustering algorithm – Kohonen self-organizing feature maps have been used to categorise the bleeding regions. For the performance comparison, ten different techniques have been executed on the dataset to justify the choice of the proposed technique.

The paper is organized as follows: Section 2 is about related works, Section 3 discusses Wavelet Transforms and Zernike moments for image vectorization, Section 4 explains single SOM, Section 5 presents the proposed method, Section 6 shows the experimental analysis, and Section 7 concludes this research study and talks about future work.

2. Literature review

This section summarizes the miscellaneous works by various researchers related to the proposed work. In [34], a comprehensive survey of computer vision techniques for wireless capsule endoscopy (WCE) has been studied. Information regarding various publicly available datasets of WCE has also been provided along with challenges and future scope. A survey has been presented in [45] for including deep learning to automate the process of WCE examination. Deep learning applications for WCE such as detecting polyps, bleeding, ulcers, hookworm, and celiac disease are discussed. A computer-aided diagnosis technique has been proposed in [10] for identifying and categorizing the abnormalities in vision-centered endoscopy detection. A novel deep sparse SVM feature selection model with group sparsity has also been incorporated, which assigns an appropriate weight to the feature dimensions and also removes the inadequate features from the feature pool. In [40], authors have utilized Zernike moments (ZM) to authenticate online signatures, and ZM represents the shape of the acceleration plot.

A novel recurrent framework has been proposed in [49] for joint unsupervised learning of deep representations and image clusters. The sequential tasks in the clustering algorithm are expressed as steps in the recurrent process, stacked on top of convolutional neural network (CNN) representations output. The research is inspired by the fact that good representation benefits image clustering, and clustering output gives supervisory indications to representation learning. Authors in [55] have proposed a Nonlinear Subspace Clustering (NSC) technique for image clustering that exposes the multi-cluster nonlinear structure of data instances using a nonlinear neural network. The technique introduced in [50] quantifies the clusterability of a dataset and is based on the probability density of a measure (S) of clusterability (in 1D) of projection of data onto a random line. After comparing the clusterability of image datasets with synthetically created clusters, it has been inferred that the structures we discover in image datasets do not fit the notion of clusters in the traditional sense. Moreover, the authors introduced a fast approach to hierarchically clustering high-dimensional data. In [8], the Deep Adaptive Clustering (DAC) approach has been proposed to represent the clustering problem as a binary pairwise classification framework for identifying whether pairs of images belong to the same cluster. The cosine distance metric has been utilized for calculating the similarities between label features of images produced by a deep convolutional network.

A novel technique, Robust learning for Unsupervised Clustering (RUC), has been introduced in [38] that is motivated by robust learning and overcomes the issues of faulty predictions and overconfident results in the case of unsupervised image clustering. This approach utilizes the pseudo-labels of existing image clustering models as noisy data that may comprise misclassified instances. In [41], the authors have proposed a two-stage deep density-based image clustering (DDC) framework to address the issue of selecting an appropriate number of clusters in advance. A pseudo-supervised joint approach has been proposed in [19] for image clustering, named Discriminative Pseudo Supervision Clustering (DPSC). Authors have resolved two significant issues in image clustering problems: appropriate image representation and lack of supervision. The main idea is to determine and use the pseudo supervision information for providing supervisory guidance for discriminative representation learning.

An improved version of ZM has been introduced in [21], which has been utilized for face recognition. In addition to the basic orthogonal and intrinsic characteristics, this ver-

sion is also invariant to noise, illumination, translation, in-plane rotation, and scaling. A hybrid similarity measure has also been proposed in this by integrating Jaccard similarity with L1 distance. Fractional-order Zernike moments, an improved version of ZM, have been presented in [23] for analyzing the grape leaf images. Multi support vector machine classifier is utilized to classify grape leaf diseases. In [25], Daubechies complex wavelet transform (DCxWT) and ZM have been used in combination for image representation. The multi-class support vector machine is used for object classification. To denoise image sequences using nonlocal means extended by ZMs, is proposed by [44]. It is found to be faster due to a reduction in weight computations, and block matching has been discounted. Similarity distance is found using photometric distance in consecutive images. A local ZM based spatio-temporal feature is proposed in [14] in the spatial domain exploiting motion change frequency for recognizing facial expressions. In [48], a study of modified principal component analysis has been performed to extract image features from the ORL face database and has been named image projection PCA (IMPCA). Sparse coded features are introduced for identifying bleeding in wireless capsule endoscopy images in [39]. These features are obtained after computing Scale-Invariant Feature Transform (SIFT) and uniform Local Binary Pattern features for WCE images. SVM is utilized for classifying the images. In [18], authors have proposed an automated system for detecting focal electroencephalogram (EEG) signals by using differencing and flexible analytic wavelet transform (FAWT) techniques. K-nearest neighbor and least squares support vector machine are applied as classifiers for automatic diagnosis.

In [4], automatic quality assessment of sperm quality (damaged or intact) has been predicted using ANN and KNN. Co-occurrence matrix and discrete wavelet transforms have been calculated from the underlying images for texture features and have been found to outperform moment-based descriptors in the study. A probability density function (PDF) based approach has been proposed in [29] for automatic detection of bleeding in WCE images. After determining the pixels of interest, local spatial features are extracted from the images by employing a linear separation scheme. In [30], an image retrieval system based upon semantic features has been studied. It uses ontological terms to define the image using multi-scale Reisz wavelets to analyze their annotation similarity. Liver lesions in CT images have been experimented with to validate the proof-of-concept. Normalized discounted cumulative gain (NDCG) score and AUC have been calculated and compared for the real-time decision-making capabilities of the model. For the robust representation of WCE images, the study given in [51] provides the assistance and discriminated definition for polyp images using a deep learning technique utilizing sparse auto-encoder. It uses a nearest neighbor graph to define inherent image manifold characteristics. A summary of the motivational literature review has been given in Table 2. In [32] a survey of large language models (LLM) have been studied for gastroentrology and semi-supervised variational models in [13].

Table 2. Summary of literature survey: pre-processing and noise removal, image representation, and learning

Sr	Method	Purpose	Outcome	Study
Pre-processing and noise removal				
1	DCxWT, ZM and multi-class SVM	Object classification	Better precision and accuracy values	Khare 2021 [25]
2	Nonlocal means extended by ZMs	Faster computation	Denoising and faster computation	Singh 2017 [44]
3	Differencing and FAWT	Automatic detection of focal EEG signals	94.41% accuracy	Gupta 2017 [18]
4	Riesz wavelets	image retrieval for a hemangioma, liver lesions	NDCG score = 0.92, AUC = 0.77	Kurtz 2014 [30]
Image representation				
5	Zernike moments	Online signature authentication	4% of False Rejection Rate, 2% of False Acceptance Rate	Radhika 2011 [40]
6	Local modified Zernike moment per unit mass	Face recognition	Higher recognition accuracies on two datasets	Kar 2020 [21]
7	Deep sparse SVM	Computer aided endoscopy diagnosis	New endoscopy dataset, Computation reduction and improved robustness	Cong 2015 [10]
8	Image principal component analysis	To analyse IMPCA is better than PCA, FDA	Better accuracy and reduced time	Yang 2002 [48]
Learning				
9	Local ZM, SVM	Facial expression recognition	Improved recognition rate	Fan 2017 [14]
10	Survey of computer vision methods for WCE	Determining major challenges of WCE and future scope	Comparative analysis	Muhammad 2020 [34]
11	Survey of deep learning for WCE	Systematic review and meta-analysis of deep learning methods for WCE	Comparative analysis	Soffer 2020 [45]
12	Sparse coded features, SVM	Detect bleeding in WCE	accuracy = 98.18%	Patel 2021 [39]
13	DWT, Invariant moments, ANN, KNN	veterinary field, spermatozoa healthy or sick	accuracy = 95%	Alegre 2012 [4]
14	Local spatial features, Rayleigh PDF model	Automatic bleeding detection in WCE images	Improved performance with less complexity	Kundu 2019 [29]
15	Stacked sparse autoencoder with image manifold constraint	polyp recognition	Overall accuracy = 98%	Yuan 2017 [51]

3. Image feature vectors

Image features involve color, texture, and shape metrics based upon the contrast-related discontinuities in the image. For this study, Wavelet Transforms [52] and Zernike moments [12] have been used due to their efficiency and power to capture the inherent characteristics.

3.1. Wavelet Transforms (WT)

These mathematical functions divide a signal (image) into different frequency components. The goal is to study each component with a resolution with a matching scale. WT

is better than Fourier Transforms (FT) or Short-Time Fourier Transform (STFT), which cannot analyze both frequency and time components [22]. Wavelet transform is composed of wavelet function $w(.)$, defined in finite time and normalized. The formula for WT is:

$$W_{f(\mu,\sigma)} = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{\sigma}} w\left(\frac{x-\mu}{\sigma}\right) dx \tag{1}$$

where (μ, σ) are translation and scaling parameters, respectively. To see lower frequency components of the signal, increase the value of σ for instance. Some prominent mother wavelets have been shown in Fig. 2. In our study, Daubechies 4 wavelet has been used (details in [11]). Spatial information comprises the image pixel positions (x, y) that act as the time axis and changes in pixel intensity $f(x, y)$ that serve as the frequency axis. Thus, edges have a higher frequency as compared to smooth areas. For Discrete WT (DWT), an image is decomposed into four components: approximation, horizontal, vertical, and diagonal. As shown in Fig. 3, the image is decomposed into one level using DWT (3a) with an example of a face image.

In our study, the image has been decomposed on three levels using WT, as shown in Fig. 4. It explains about horizontal, vertical and diagonal edges being detected in the original image. Ten components have been calculated as $\{(H_i, V_i, D_i, A_i) \mid i = 1, 2, 3$ for H, V, D and $i = 4$ for $A\}$. In expanded form, we get $H_1, V_1, D_1, H_2, V_2, D_2, H_3, V_3, D_3$, and A_3. Then for these 10 components, 12 Zernike moments have been calculated for $n = m = 5$ which are listed as $Z_{00}, Z_{11}, Z_{20}, Z_{22}, Z_{31}, Z_{33}, Z_{40}, Z_{42}, Z_{44}, Z_{51}, Z_{53}$, and Z_{55} or in the set notation $\{Z_{ij} \mid i \geq j$ and $i - |j|$ is even$\}$. After compiling all that information from LUV channels, the image feature vector has $12 \times 3 = 36$ dimensions. For example, Fig. 5 shows the results from a sample picture's approximation, horizontal, vertical, and diagonal edge detection decompositions.

Wavelet Transformation (WT) is quite useful for noise removal, image compression [52], and zooming capabilities for local characteristics of an image. It is also an efficient technique for texture characterization while preserving local and global spatial/spectral information. For instance, the noise removal feature of WT is shown in Fig. 5 with four decompositions levels, and image denoising has been illustrated in Fig. 6 for an image with a considerable amount of Gaussian noise.

3.2. Zernike moments (ZM)

Image moments are the weighted average of the intensity values of the image pixel (or a similar image function) to get the scalar quantities for image interpretation. Moments of different order yield varying information about the image, such as area, center of mass, and orientation. Zernike Moments (ZM) [16] of an image are similar to Discrete Cosine Transform (DCT) coefficients in their derivation and properties. ZM are projections of an image function along the real and imaginary axes (x-axis and y-axis), which are convolved by an orthogonal function. They represent an image in various frequency components which are referred to as the orders (along the radial) and repetitions (along the angular direction). Thus, Z_{00} represents the average intensity, Z_{11} represents the first-order moment, Z_{20} is similar to variance, and so on. Zernike polynomials are orthogonal functions that generate an orthogonal set over the unit circle in a complex plane. The center of the image stays the same as the center of the circle. Hence, a square image can be

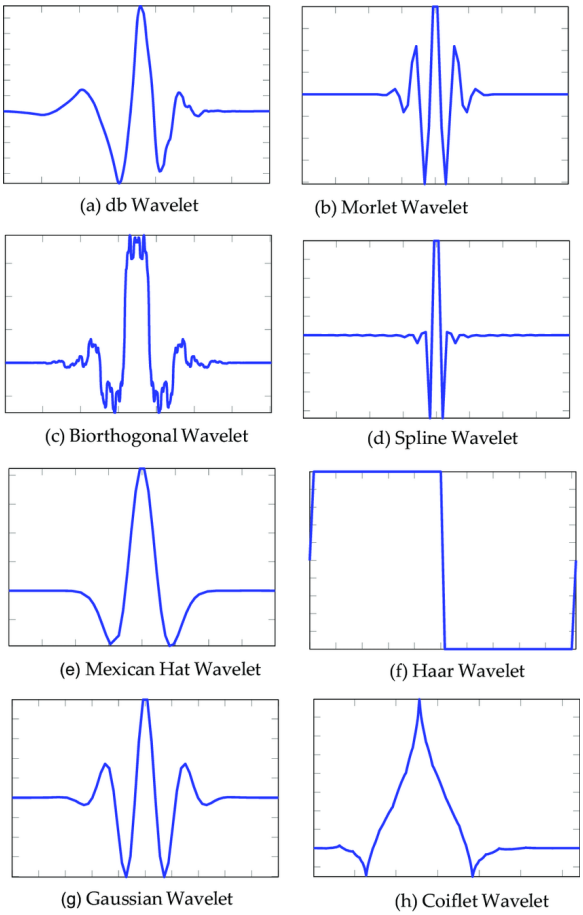
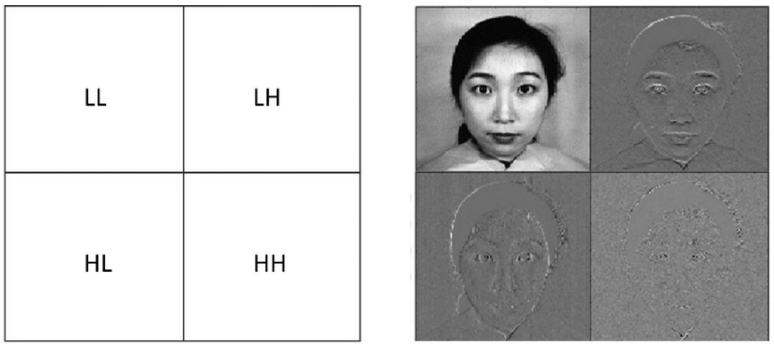


Fig. 2. A few popular mother wavelet functions [15] $w(.)$. Daubechies 4 wavelet have been utilized in the experimentation

mapped inside or outside an image [1]. In the case of inner mapping, the pixels which fall outside the unit disc must be discarded. So, to avoid the information loss from the edges, we have utilized outer mapping for our experimentation which is shown in Fig. 7.

Formula for Zernike polynomials is $V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta}$. Here n, m are whole numbers such that $n - |m| = \text{even}$, $n \geq 0$, $0 \leq \mid m \mid \leq n$, $\theta = arctan(y/x)$ and $j = \sqrt{-1}$. (ρ, θ) are radius and angle of the pixel from origin which simply means the polar coordinate of a pixel at (x, y). Formula for radial polynomial $R_{nm}(\rho)$ is given as follows:

$$R_{nm}(\rho) = \sum_{k=0}^{(n-|m|)/2} (-1)^k \times \frac{(n-k)!}{k!(\frac{n+|m|}{2}-k)!(\frac{n-|m|}{2}-k)!}\rho^{n-2k} \tag{2}$$



(a) DWT decomposition (b) DWT decomposition of face image

Fig. 3. Image decomposition using DWT with an example of face image [20]

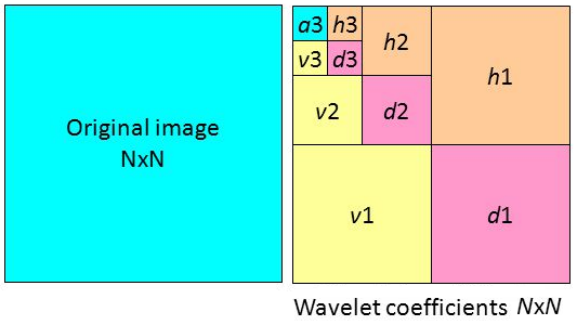


Fig. 4. Daubechies wavelet transformations are used in the experiments. a, v, h, d stands for approximation, vertical, horizontal, and diagonal details. Diagonal (low/low), horizontal (high/low), vertical (low/high), approximation (high/high)

$$Z_{nm} = \frac{n+1}{\pi} \sum_{x=1}^{N-1} \sum_{y=1}^{N-1} f(x,y) R_{nm}(\rho) e^{jm\theta} \tag{3}$$

Z_{nmx} and Z_{nmy} are cosine and sine values of Z_{nm} (Zernike moments). The corresponding value if ZM can be calculated as $Z_{nm} = \sqrt{Z_{nmx}^2 + Z_{nmy}^2}$. Rotational and scale invariance can be obtained in ZM by normalizing the image using Cartesian moments before the ZM calculation [26]. Moreover, if the center of mass of image is moved to origin then translation invariance can also be achieved.

4. Single SOM

Our method involves definitions for creating a set that associates the most active neuron for the set of the output layer of SOM, with a set of input vectors presented to the input layer of SOM as defined in Equations 6 and 7. It applies to a single SOM or can be

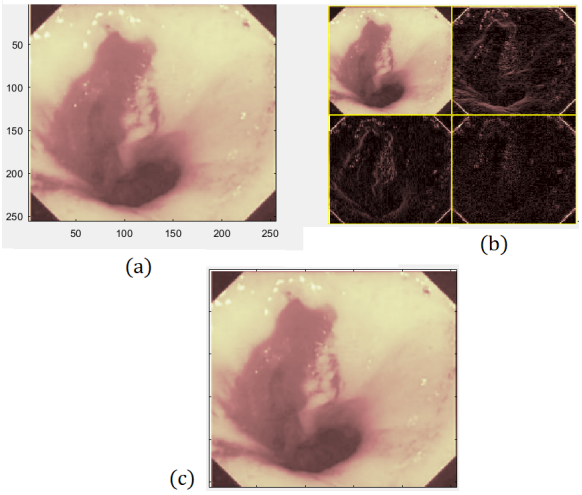


Fig. 5. The four decompositions explained with example: approximation, horizontal, vertical, and diagonal details to detect the corresponding edges. Fig. (a) is original image, (b) is the view of four decompositions, and (c) is denoised image

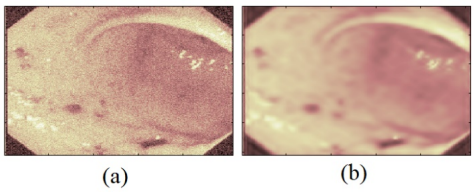


Fig. 6. (a) Noisy image and (b) denoised image with Daubechies wavelets (DB-4)

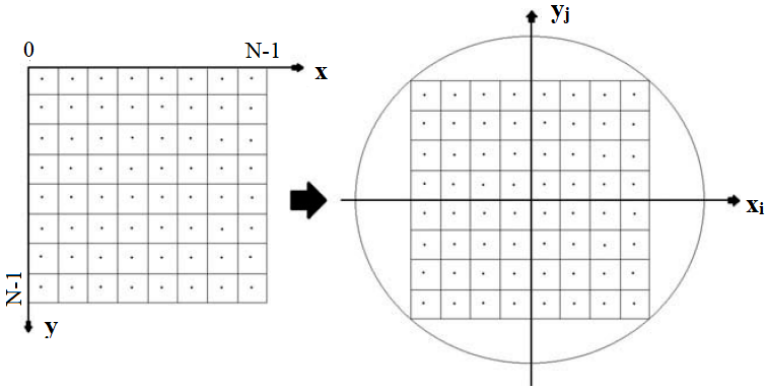


Fig. 7. Outer mapping which maps a given image inside the unit disc

extended as the collateral SOM for hybrid SOMs. Follow Algorithm 1 for creating single modal information systems for image clustering.

Algorithm 1 Algorithm for retrieving information from a single SOM

1: Identify the best match node $\vec{w}_k$.
2: Form a totally ordered set of the n nodes in the SOM, such that:

$$(W, \leq) = \left\{ \begin{array}{c} \vec{w}_k, k = 1..n \ \mid \ \vec{w}_i \leq \vec{w}_j \Leftrightarrow \\ \|\vec{x}_k - \vec{w}_i\| \leq \|\vec{x}_k - \vec{w}_j\| \end{array} \right\} \tag{4}$$

where $\vec{w}_i, \vec{w}_j \in W, 1 \leq i, j \leq n$ and $i \neq j$
3: Retrieve a totally ordered set R, of all p pre-stored items used in training, in response to the input vector $\vec{x}$

$$R_{single} = \{\vec{x}_l, l = 1..p \ \mid \ \exists \ \vec{w}_k \in W : (\vec{x}_l, \vec{w}_k) \in P_{single}\} \tag{5}$$

A Self-organizing Map (SOM) [28], also called a Kohonen Map, associates a multidimensional input space, comprising a set of feature vectors, onto a 2-dimensional surface (output map). The end of training leads to an association between an input vector $\vec{x}$ and a specific output node that 'wins' the input, known as the Best Matching Unit (BMU) for that input vector. If $\vec{w}$ represents the weight vector of an output node, then BMU for input vector $\vec{x}$ can be calculated as:

$$\|\vec{x} - \vec{w}_m\| = min\{\|\vec{x} - \vec{w}_m\|\} \tag{6}$$

where m depicts the index of SOM output node which is a BMU. One node may 'win' over more than one input forming a set. Let P_{single} be the pair set of q input vectors and the corresponding winning node is $\vec{w}_m$, then P_{single} is defined as:

$$P_{single} = \left\{ \begin{array}{c} (\vec{x}_k, \vec{w}_m), k = 1..q \\ \|\vec{x}_k - \vec{w}_m\| = \min\limits_{i=1}^{n}\{\|\vec{x}_k - \vec{w}_i\|\} \end{array} \right\} \tag{7}$$

Information retrieval from a SOM involves the presentation to the trained SOM of a set W. The mapping of the input vector from higher dimensional nodes in the output layer forming a space to the winning node in 2-D neuron space has shown in Fig. 8. The length of input vector X_i and neuron weight vector W_i must be the same. The retrieving information from a SOM has been depicted in the Algorithm 1. The following section explains image vector creation using Zernike Moments and Wavelet transformation for denoising.

During the initial stages of the SOM training, the weight vectors are initialized with random weights and then, together with the input vectors, are normalized. The learning and neighborhood rates are reduced exponentially during training following established practice in the SOM literature. Our testing regimen relies on the notion of best matching unit(s): the node(s) in the output layer that responds with the highest activation value to a given input vector. Note that if one or more neurons can be activated in response to the input vector, then the activated neurons are ordered according to their activation levels (Algorithm 1). If the category of the input vector matches the most activated neuron in the output layer, then we have a best-matching unit (BMU). If there are multiple activated nodes for a specific input then we are considering the two highly activated nodes only.

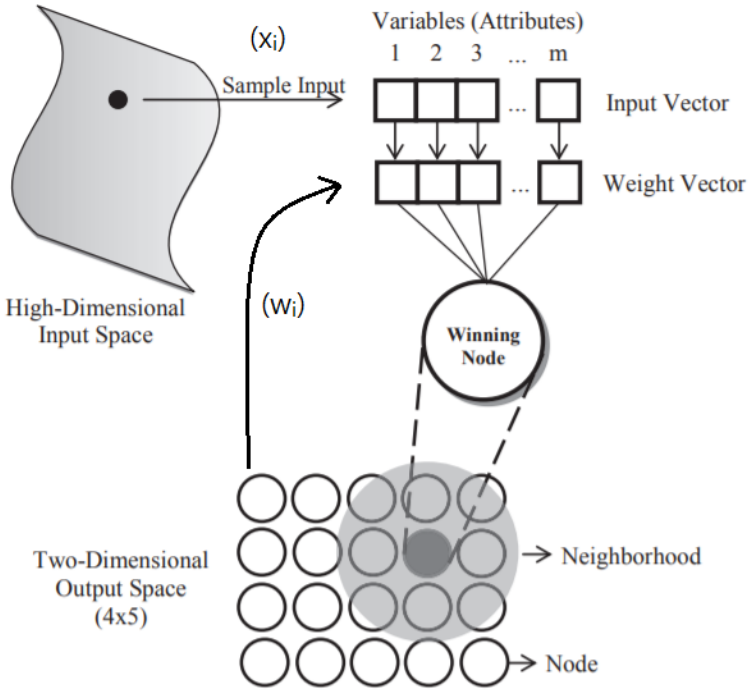


Fig. 8. Overview of single self-organizing map (SOM) model. X_i are input vectors with same length as weight vectors W_i. Each X_i is connected to every (winning) neuron

A matching matrix was created to analyze how an input vector may activate neurons that were trained to respond to one or more categories of keywords or images. If the winner or BMU in the output layer has the same category as the stimulus, and the stimulus did not excite any other neurons, then the match will be perfect. However, if a given stimulus activates neurons of various other categories, the match will be minimal. We define accuracy as the number of correctly clustered items (based upon the majority of similar items in the cluster as the test instance) divided by the total number of items in the category.

5. Proposed Methodology

The proposed research aims to effectively cluster the in-vivo gastrointestinal images based upon their similarity by carefully considering the image semantics. Let I be the training set of images that is an input to the proposed algorithm. The expected output is the trained self-organizing map and the image cluster sets (C_i) constructed as per the image similarity. The first step is to denoise the images using Daubechies wavelets with four decomposition levels: approximation, horizontal, vertical, and diagonal. The next step is the conversion of RGB to LUV channels. Wavelet transforms implementation details are

given in *Section 3.1*. Subsequently, 12 Zernike moments are calculated for each of the L, U, and V channel with $n = m = 5$, creating a total of $12 \times 3 = 36$ image vector dimensions. The ZM calculation steps and equations are mentioned in detail in *Section 3.2*. In the end, 4×4 SOM is trained using image vectors and constructs the image clusters. Algorithm 2 shows the steps for segmentation and clustering of the images using SOM.

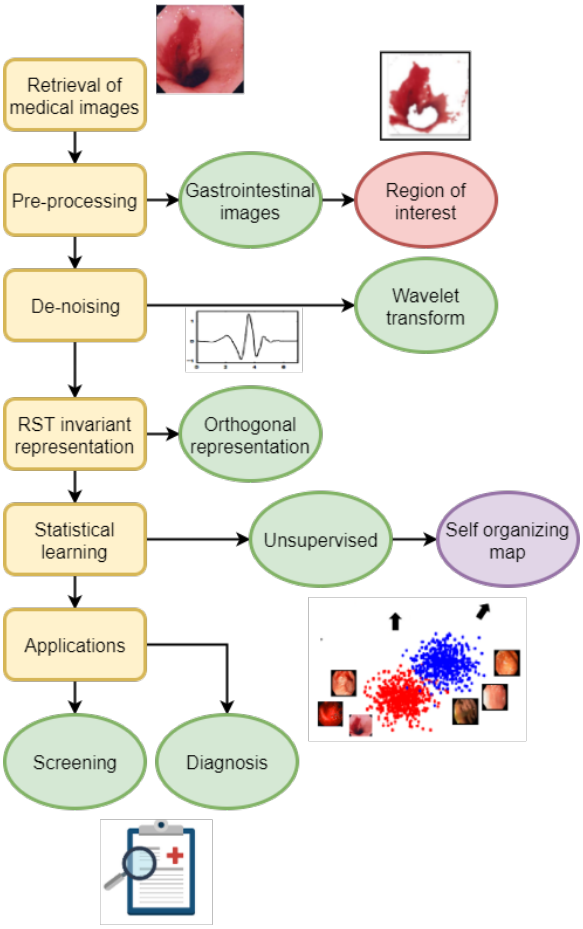


Fig. 9. Pipeline diagram for the proposed methodology

As per [37], the total number of multiplications required for computing a radial polynomial ($R_{nm}(\rho)$) using Equation 2, is almost $(n/2 + 1) \times (2n - 3) \times (n - 1)$. So, computational complexity of calculating a single $R_{nm}(\rho)$ value of order n and repetition m is $O(n^3)$. If the image dataset size is D, then the total complexity of ZM calculation becomes $O(Dn^3)$. The processing time of SOM is $O(D^2)$ [42]. So, the computational complexity of the proposed algorithm is $O(Dn^3 + D^2)$.

Algorithm 2 SOM image clustering

INPUT: Training set of images I
OUTPUT: Image cluster sets (C_i where i = number of SOM clusters and $C_i \subseteq I$)

procedure SOM TRAINING FOR ENDOSCOPY IMAGES
  Image denoising - Daubechies wavelets ($DB-4$) using four decomposition levels (a, v, h, d)
  RGB to LUV transformation
  Calculate ZM for each of L, U, and V channels with n=m=5
  (i) Calculate radial polynomial $R_{nm}(\rho)$ using eq. 2
  (ii) $V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta}$
  (iii) Z_{nmx} and Z_{nmy} are real and imaginary values of Z_{nm}
  (iv) $Z_{nm} = \sqrt{Z_{nmx}^2 + Z_{nmy}^2}$
  (v) Calculate 12 Z_{nm} for each L, U, and V channel, so total 36 elements in image vector
  Train SOM with 4×4 grid size using the obtained image vectors
  Required image clusters (C_i) are obtained after SOM training
end procedure

Fig. 9 is the pictorial representation of all the steps involved in implementing the proposed approach. Initially, we have a set of 300 raw endoscopy images. The images are pre-processed and the region of interest is identified. Afterward, the denoising of the images is performed using wavelet transforms and the RGB images are converted to LUV format. Subsequently, ZM features are extracted from the images, which are rotation, scaling, and translation invariant. The unsupervised self-organizing map is trained using the extracted image features, and the image clusters are formed based on the similarity. Now the trained system can be utilized by gastroenterologists for screening and diagnosis purposes for endoscopy images.

6. Experiments

This section includes the information regarding dataset, its analysis and results obtained using proposed approach. The configuration of the system used for experiments is: Desktop System is Dell Inc. with Model XPS 8930 with Windows 10 ProVersion 10.0.17763, Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz, 3192 Mhz, 6 Core(s), 12 Logical Processor(s), with 16GB RAM.

6.1. Dataset

The dataset comprises 300 real gastrointestinal images obtained from a known gastroenterologist with a ratio of 180:120 for healthy and sick cases. All the images are of size 256×256 and are from both upper (esophagus and stomach) and lower (small bowel and colon) gastrointestinal tract. The sample images from the dataset are demonstrated in Fig. 10 and Table 3 provides the information regarding the dataset.

6.2. Data distribution analysis

The data distributions of healthy and sick image vectors have been examined from various aspects (to analyze an appropriate learning model), which are as follows:

Fig. 10. Sample gastral images for bleeding detection. Total 300 in number with ratio of 180:120 for healthy and sick. Image size is 256×256 pixels

Table 3. Description of images in dataset

Category	**Healthy**	**Sick**
Image ratio	180	120
Size	256×256	256×256
Redness	Overall	Spots or saturation

- Relative red intensity in healthy/sick images.
- Distribution of RGB intensities in all images.
- Thresholding to crop the red color (for example, $R > 100, G < 60, B < 50$).

In Fig. 11 the average red color in the sick and healthy classes has been sorted and plotted. Although all gastral images are reddish brown in color, the sick images are more saturated with redness. The intensity plots of all the images have been illustrated in Fig. 12: (a) shows that the left half has more dispersion, especially in red color and R values are relatively higher. The other three plots (b-d) show the R versus G, R versus B, and B versus G plots. There is tremendous overlap, so a simple linear regression may not be sufficient for the bleeding analysis. Thus, there is a need for a non-linear learning system (such as SOM). The red color segmentation has been experimented with using MATLAB to further analyze the problem complexity, which is illustrated in Fig. 13. The threshold values $R > 100, G < 60, B < 50$ have been chosen for best human eye subjective red color cognition through the experiments. Again, it seems quite difficult to distinguish the healthy red versus the sick red spots for confounding cases. The middle two images are healthy in these four images, and the left/right extremes are bleeding cases. Therefore, simple thresholding is also insufficient to spot the bleeding even with various threshold values.

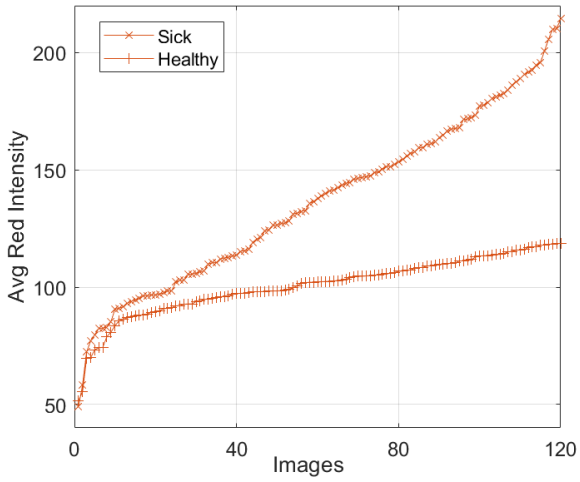


Fig. 11. Sorted average red pixel intensities for normal and abnormal images. The upper line is redness in sick images which is relatively higher as compared to healthy images

6.3. ZM extraction and SOM application

For each L, U, and V component, 12 ZM have been calculated, making a total of 36 ZMs to extract the luminance and color attributes as shown in Fig. 14. Zernike moments are rotation and noise invariant as studied in [26], which can be seen in the Figures 15 and 16. Furthermore, feature transformation techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied to these 300×36 image vectors as shown in Fig. 17. Both of these transformations are used for dimension reduction, but PCA focuses on maximizing the variance among mutually orthogonal transformed dimensions, and LDA focuses on the separability of the data concerning the labels [33]. Linear separability in the case of PCA has been found to be 74% and 84% in case of LDA. Therefore, linear separability becomes possible after extracting ZM on LUV image components. Self-Organizing Maps have been involved further to analyze the accuracy with the hope of improvement.

Fig. 18 shows the results when the model was trained using only healthy (180) points and tested for 120 sick images. It can be observed that there is some overlap of the tested sick images with healthy images due to the obscure nature of the images. But still, there is a complementary saturation between these two images showing that sick images have different data distribution on a broader scale.

Wavelet transforms applied to all the images before extracting their Zernike moments (ZM) for noise removal. In Table 4, the results of accuracy for ZM versus WT+ZM on 300×36 image vectors have been compared. Table 5 shows the difference between WT+ZM and ZM accuracy is positive on the average of 5 trials, confirming the advantage of WT application before ZM extraction. Fig. 19 is the visual illustration of Table 5. Finally, Table 6 presents the confusion matrix obtained after experiments and the best accuracy obtained using WT+ZM and Kohonen self-organized feature maps has been found to be approximately 88.3%. In the table, P_Healthy, P_Sick and A_Healthy, A_Sick are

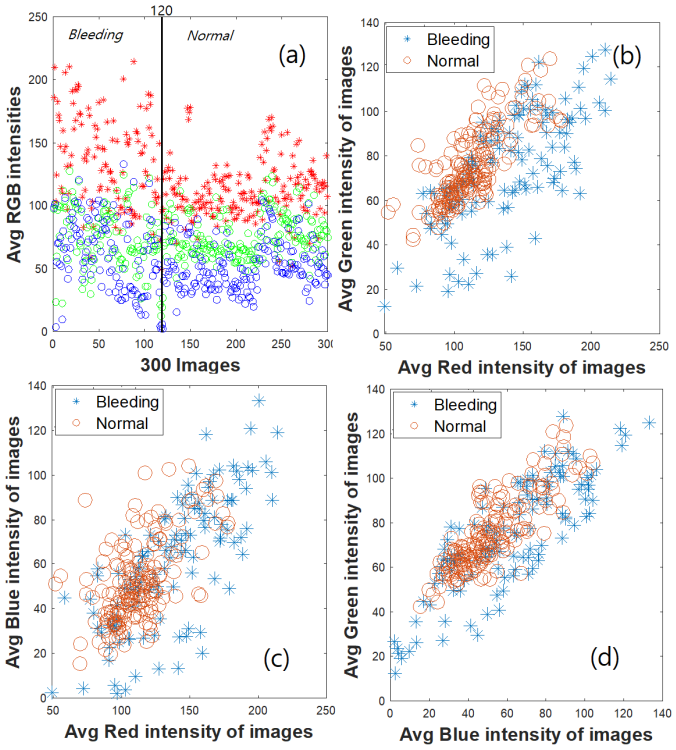
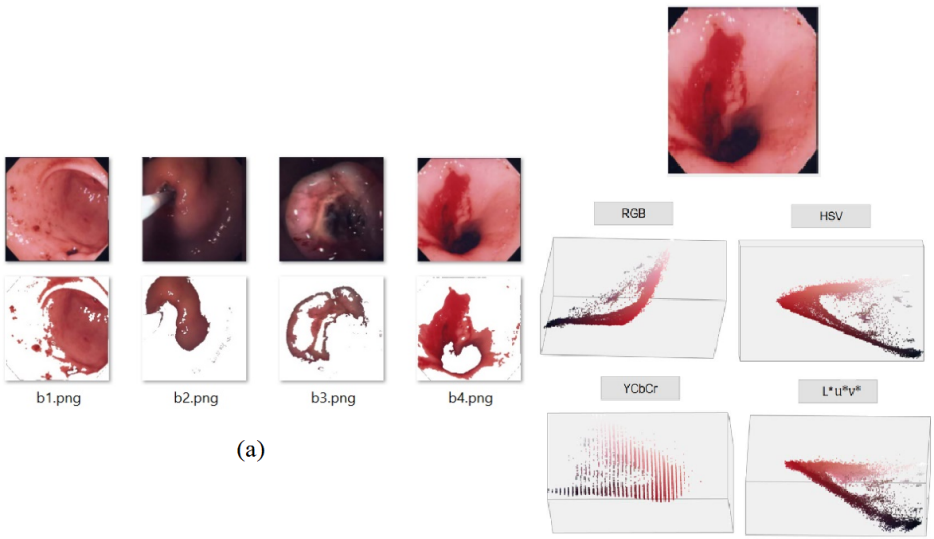


Fig. 12. (a) RGB intensity of 300 images (180 : 120 for healthy:sick cases) (b) Red vs Green average intensities (c) Red vs Blue avg. intensities (d) Blue vs Green intensities plotted for all the images. It is clear from (b-d) that none of the color intensities are easily separable for healthy and sick cases

Table 4. Accuracy given by ZM versus WT+ZM on 300 × 36 images. TR = training data, TS = testing data, VAL = validation data, and AVG = average.

	ZM				**WT+ZM**			
Trials	**TR**	**VAL**	**TS**	**All**	**TR**	**VAL**	**TS**	**All**
1	82.4	78.3	71.7	78.7	83.8	86.7	75.6	83
2	79.5	84.4	77.2	80	81.4	86.7	77.8	81.7
3	77.6	82.2	77.8	78.3	82.4	81	76.9	81
4	75.2	80.2	77.8	76.7	82.4	80	80	81.7
5	80.5	82.2	80.2	81	83.3	88.9	79.6	83
AVG	79	81.5	76.9	78.9	82.9	84.5	79.8	82.4



(a)

(b)

Fig. 13. (a) Red color cropping using subjective threshold RGB values (using MATLAB) $R > 100, G < 60, B < 50$. Hence, in order to classify the images, robust features are required to extract such as ZM. Firstly, RGB values have been converted to LUV color representation to separate the luminance component from the color composition as shown in Fig. (b)

	1	2	3	4	5	6	...	34	35	36
n1	512.4	165.4	204.8	6.6	144.3	36.6		34.5	17.8	0.2
n2	533.6	165.3	243.5	4.6	157.9	31.0		10.0	5.7	0.9
n3	457.5	141.7	217.3	1.3	143.5	19.6		31.7	11.1	1.1
n4	444.7	139.3	204.2	1.2	136.0	21.8		25.7	12.3	0.0
n5	520.9	157.2	239.8	11.5	154.9	30.0		30.1	16.6	0.6
n6	518.1	158.5	252.4	3.0	162.1	21.2		39.3	18.4	0.3
n7	476.4	150.6	218.2	5.1	146.0	22.5		30.1	8.3	0.1
n8	555.1	170.9	263.1	0.9	171.3	25.1		19.5	8.1	0.3
n9	470.4	143.6	215.9	3.8	131.2	28.2		22.5	9.8	1.1

Fig. 14. Snapshot of ZM for 9 healthy images. Rows corresponds to the image vectors and 36 columns are the Zernike moments of image. n means normal/healthy.

predicted and actual values of healthy and sick classes respectively. Table 7 shows the comparison of the other approaches with the proposed technique based on obtained accuracy. The comparison is performed with both traditional methods such as PCA and LDA (used for linear separability of data), and contemporary methods such as deep learning based approaches.

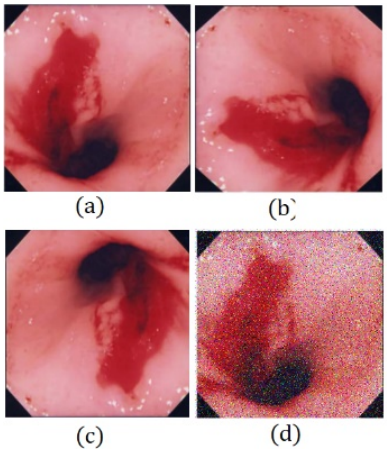


Fig. 15. (a)-(d) are original image, rotations of 90 and 180 degrees, Gaussian noise introduced

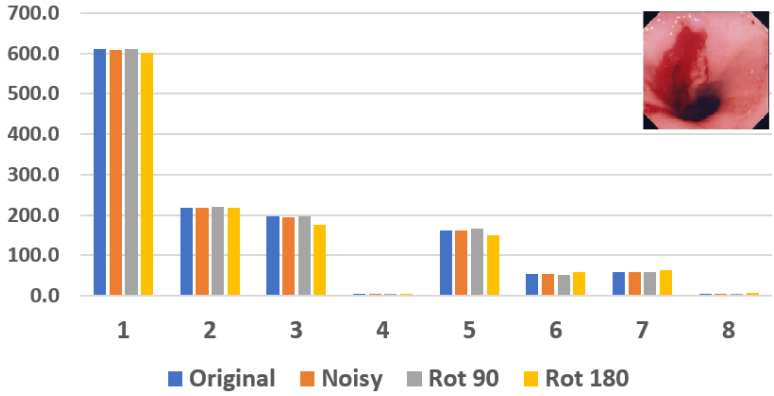


Fig. 16. ZM for all images in Fig. 15 are nearly same. Only 8 out of 36 ZM have been shown to simplify the demonstration. Slight differences are due to 3 types of errors involved in ZM calculation as studied in [1]

6.4. Incorporating Image Captions for Multi-modal Learning

In this subsection, the freely available descriptions with the gastral images are incorporated into the system to improve learning accuracy. In real-world medical diagnosis, image features are just not enough to yield the required information. For example, in Fig. 20, two confounding images have been considered which look identical; however, Fig. (a) involves bleeding, and Fig. (b) has an air bubble in case of a healthy image. Therefore, linguistic cues (provided by experts) can be associated with images to handle these kinds of cases.

Information from multiple modalities, such as images and collateral text, can be utilized simultaneously for different tasks, including classification, clustering, or object de-

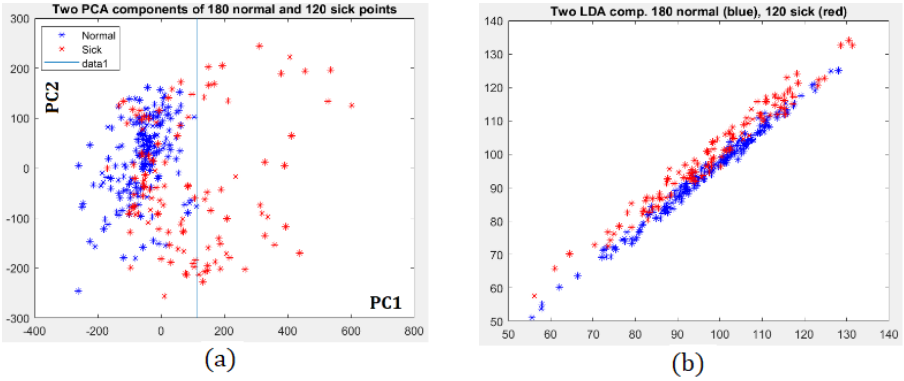


Fig. 17. PCA and LDA draws of (300) image vectors. PCA yields 74% and LDA yields 84% linear separability of 120 bleeding and 180 normal image vectors. *Blue* for healthy and *Red* for sick.

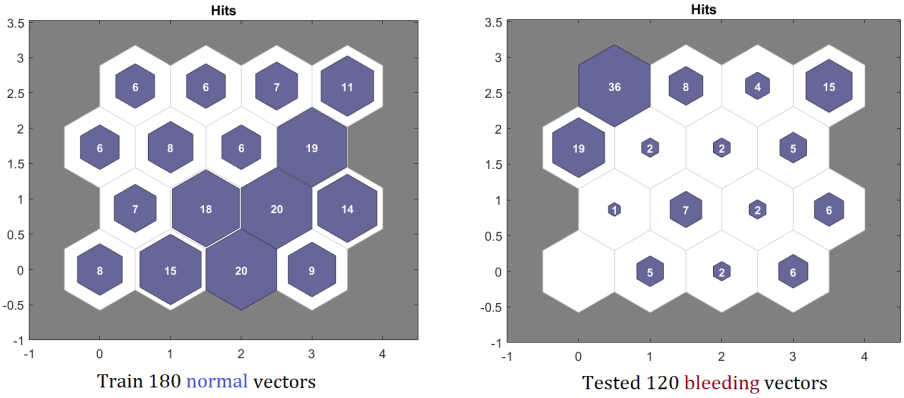


Fig. 18. SOM maps obtained upon training with 180 normal points and testing with 120 bleeding images using MATLAB. It can be observed that the mapping is different and thus distribution of 180×36 and 120×36 image vectors are different.

tection, which is known as multi-modal learning [24]. Sometimes, the clusters constructed by SOM are pretty small, which acts as the outliers. These clusters can be merged with similar bigger clusters using related textual information. The endoscopy images are accompanied by corresponding labels or small text that provides some description of them. A SOM is trained with this linguistic information similar to the SOM trained with image data. The raw text is pre-processed to remove the noise, and then it is represented as Bag-of-Words (BoW) for vectorization before training SOM. The small image clusters can now be merged based on the corresponding textual features associated with these images. The new accuracy obtained with this technique is 90.12% which is better than the previously achieved accuracy. From the results, it can be inferred that the system performance has improved with the inclusion of the second modality. The performance of the approach

Table 5. Difference in the accuracy values of after and before WT while extracting ZM. Positive values in the last row signifies the benefit of WT application prior to ZM.

	Accuracy of WTZM - ZM			
Trials	**TR**	**VAL**	**TS**	**All**
1	1.4	8.4	3.9	4.3
2	1.9	2.3	0.6	1.7
3	4.8	-1.2	-0.9	2.7
4	7.2	-0.2	2.2	5
5	2.8	6.7	-0.6	2
AVG	3.8	3	2.9	3.4

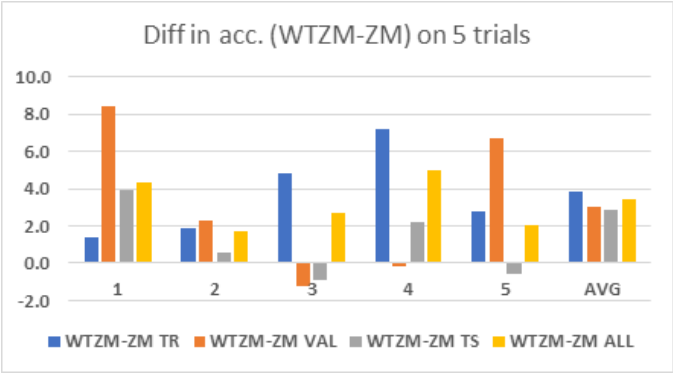


Fig. 19. Difference in the accuracy of results by using WT+ZM versus only ZM as in Table 5.

Table 6. Average test results for 5-trials with the proposed method on 300 image vectors. P_Healthy, P_Sick and A_Healthy, A_Sick are predicted and actual values of healthy and sick classes respectively.

	A_Healthy	**A_Sick**	**Total**
P_Healthy	161	16	177
P_Sick	19	104	123
Total	180	120	300
Accuracy	89.4%	86.7%	88.3%

can be further boosted if the quality of image captions is good and they are noise-free, having no stop words and more meaningful information about the corresponding image.

Table 7. Comparative analysis of techniques on the underlying dataset.

Sr.	Reference	Technique	Accuracy
1	[48]	Principal component analysis	74%
2	[49]	Agglomerative clustering and CNN	80.5%
3	[55]	Nonlinear subspace clustering	83.3%
4	[33]	Linear discriminant analysis	84%
5	[38]	Robust learning for unsupervised clustering	85.7%
6	[41]	Deep density-based image clustering	86.8%
7	[50]	Hierarchical clustering using 1D random projections	87.1%
8	[8]	Deep adaptive image clustering	87.6%
9	[19]	Discriminative pseudo supervision clustering	87.9%
10	**Proposed**	**WT+ZM on LUV**	**88.3%**

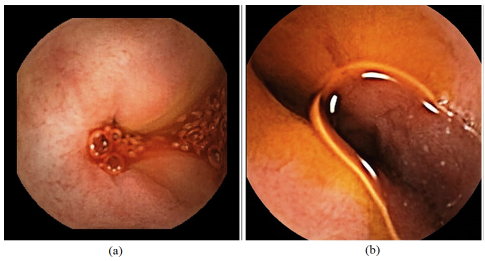


Fig. 20. (a) Active bleeding in small bowel (b) False positive (air bubble), images from [6]

6.5. Discussion

The primary goal of the proposed research is to present the importance of understanding and analyzing the data to find the appropriate methods for its processing as per the essence of the data and the underlying problem. The final values chosen for all the tuning parameters for experimentation have been decided based on multiple trials of experiments. We have also tested the proposed technique by increasing the size of the endoscopy data to 1200 and observed that the performance and accuracy are almost similar to the smaller data. The proposed approach outperforms the traditional as well as contemporary image clustering approaches due to following reasons:

1. An appropriate medical image representation is an important task for which the combination of Wavelet transform and Zernike moments have been utilized to retrieve noise-free, least redundant, and rotation, scaling, and translational invariant features. ZM captures the global features of an image and also effectively describe the shape characteristics [2].

2. Self-organizing map provides the robust medical image clustering as it works similar to the brain neurons [27]. In addition, SOM has been quite effective in diverse recent applications such as mental stress detection [46], coronary heart disease diagnosis [35], speech recognition [31], and in feature extraction as an add-on for better network intrusion detection [9].
3. SOM provides easy data interpretation, and understanding [27]. It provides potent data visualization and has the capability of clustering even complex datasets [28].
4. Deep learning requires a colossal dataset to perform well [5], which is unavailable in the proposed research; this may cause the over-fitting issue [43].
5. Generally, a deep learning approach (such as CNN) cannot directly outperform a machine learning approach as its performance mostly depends upon the design comprising training strategies, layer depth, and size [17].
6. To use transfer learning and retraining the deep learning model on a new dataset requires understanding various model parameters and the layer modifications, which is computationally expensive [36].

7. Conclusion

This paper introduced new ways of intelligently segmenting and analyzing image collections by training neural computing systems with images having obscure color and texture contrasts. The characteristic visual features of the image collection are derived from Wavelet Transforms, Zernike Moments, and Linear Discriminant Analysis. The images are categorized using an unsupervised clustering algorithm – Kohonen self-organizing feature maps. The proposed system can classify sick and healthy in-vivo images effectively without the labeled data, which is hard to get in reality, specifically medical data. It is often expensive to manually label the data by an expert in the related field. The system is beneficial in clustering vague color distribution, asymmetrical region boundary, and noisy image data. It is rotation, scaling, and translation invariant due to the use of ZM for image representation. The system efficacy improved by incorporating the second modality, i.e., free text with the gastral images in the experiments.

7.1. Limitations

There are three types of errors involved in the calculation of ZM: (a) Geometric error due to mapping of a digital image into a unit circle with pixels, (b) Discretization error due to the computer's digital representation of continuous variables, and (c) Numerical integration error due to the calculation of double integration through double summations while the center of a grid is used to calculate the basis function. The size of the data considered for experimentation in this study is small, which is a drawback.

7.2. Future scope

Grid size for SOM is a parameter for subjective tuning. The overall accuracy of the proposed system is encouraging, although image semantics need to be considered more carefully to improve the automatic learning system. Future experimentation can be performed

on ample data from diverse fields, and miscellaneous noise removal and appropriate feature extractors can be considered (as per the underlying data), which can further improve the accuracy. Various ways of integrating multi-modal data can be explored to extend this work further and improve the clustering process.

Acknowledgments. The authors are thankful to (a) Gastroenterologists Dr. Sunil Arya at Leela Bhawan Patiala & Dr. G.S. Sidhu at Max Hospital Mohali, India, for the dataset and technical feedback, and (b) Professor Khurshid Ahmad, Trinity College Dublin, Ireland for his valuable advice.

References

1. Aggarwal, A., Singh, C.: Zernike moments-based gurumukhi character recognition. Applied Artificial Intelligence 30(5), 429–444 (2016)
2. Aggarwal, H., Vishwakarma, D.K.: Covariate conscious approach for gait recognition based upon zernike moment invariants. IEEE Transactions on Cognitive and Developmental Systems 10(2), 397–407 (2017)
3. Aghajari, E., Chandrashekhar, G.D.: Self-organizing map based extended fuzzy c-means (seefc) algorithm for image segmentation. Applied Soft Computing 54, 347–363 (2017)
4. Alegre, E., González-Castro, V., Alaiz-Rodríguez, R., García-Ordás, M.T.: Texture and moments-based classification of the acrosome integrity of boar spermatozoa images. Computer Methods and Programs in Biomedicine 108(2), 873–881 (2012)
5. Bekhouche, S., Dornaika, F., Benlamoudi, A., Ouafi, A., Taleb-Ahmed, A.: A comparative study of human facial age estimation: handcrafted features vs. deep features. Multimedia Tools and Applications 79(35), 26605–26622 (2020)
6. Boal Carvalho, P., Magalhães, J., Dias de Castro, F., Monteiro, S., Rosa, B., Moreira, M.J., Cotter, J.: Suspected blood indicator in capsule endoscopy: a valuable tool for gastrointestinal bleeding diagnosis. Arquivos de gastroenterologia 54, 16–20 (2017)
7. Borges, V.R., de Oliveira, M.C.F., Silva, T.G., Vieira, A.A.H., Hamann, B.: Region growing for segmenting green microalgae images. IEEE/ACM transactions on computational biology and bioinformatics 15(1), 257–270 (2016)
8. Chang, J., Wang, L., Meng, G., Xiang, S., Pan, C.: Deep adaptive image clustering. In: Proceedings of the IEEE international conference on computer vision. pp. 5879–5887 (2017)
9. Chen, Y., Ashizawa, N., Yeo, C.K., Yanai, N., Yean, S.: Multi-scale self-organizing map assisted deep autoencoding gaussian mixture model for unsupervised intrusion detection. Knowledge-Based Systems 224, 107086 (2021)
10. Cong, Y., Wang, S., Liu, J., Cao, J., Yang, Y., Luo, J.: Deep sparse feature selection for computer aided endoscopy diagnosis. Pattern Recognition 48(3), 907–917 (2015)
11. Daubechies, I.: Different perspectives on wavelets, vol. 47. American Mathematical Soc. (2016)
12. Deng, A.W., Wei, C.H., Gwo, C.Y.: Stable, fast computation of high-order zernike moments using a recursive method. Pattern Recognition 56, 16–25 (2016)
13. Du, W., Rao, N., Yong, J., Wang, Y., Hu, D., Gan, T., Zhu, L., Zeng, B.: Improving the classification performance of esophageal disease on small dataset by semi-supervised efficient contrastive learning. Journal of Medical Systems 46, 1–13 (2022)
14. Fan, X., Tjahjadi, T.: A dynamic framework based on local zernike moment and motion history image for facial expression recognition. Pattern recognition 64, 399–406 (2017)
15. Faust, O., Acharya, U.R., Adeli, H., Adeli, A.: Wavelet-based eeg processing for computer-aided seizure detection and epilepsy diagnosis. Seizure 26, 56–64 (2015)

16. Fredo, A.J., Abilash, R., Femi, R., Mythili, A., Kumar, C.S.: Classification of damages in composite images using zernike moments and support vector machines. Composites Part B: Engineering 168, 77–86 (2019)
17. Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S.R., Tiede, D., Aryal, J.: Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing 11(2), 196 (2019)
18. Gupta, V., Priya, T., Yadav, A.K., Pachori, R.B., Acharya, U.R.: Automated detection of focal eeg signals using features extracted from flexible analytic wavelet transform. Pattern Recognition Letters 94, 180–188 (2017)
19. Hu, W., Chen, C., Ye, F., Zheng, Z., Du, Y.: Learning deep discriminative representations with pseudo supervision for image clustering. Information Sciences 568, 199–215 (2021)
20. Indolia, S., Nigam, S., Singh, R.: A self-attention-based fusion framework for facial expression recognition in wavelet domain. The Visual Computer pp. 1–17 (2023)
21. Kar, A., Pramanik, S., Chakraborty, A., Bhattacharjee, D., Ho, E.S., Shum, H.P.: Lmzmpm: local modified zernike moment per-unit mass for robust human face recognition. IEEE Transactions on Information Forensics and Security 16, 495–509 (2020)
22. Karim, S.A.A., Kamarudin, M.H., Karim, B.A., Hasan, M.K., Sulaiman, J.: Wavelet transform and fast fourier transform for signal compression: A comparative study. In: 2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA). pp. 280–285. IEEE (2011)
23. Kaur, P., Pannu, H.S., Malhi, A.K.: Plant disease recognition using fractional-order zernike moments and svm classifier. Neural Computing and Applications 31(12), 8749–8768 (2019)
24. Kaur, P., Pannu, H.S., Malhi, A.K.: Comparative analysis on cross-modal information retrieval: a review. Computer Science Review 39, 100336 (2021)
25. Khare, M., Khare, A.: Integration of complex wavelet transform and zernike moment for multiclass classification. Evolutionary Intelligence 14(2), 1151–1162 (2021)
26. Khotanzad, A., Hong, Y.H.: Invariant image recognition by zernike moments. IEEE Transactions on pattern analysis and machine intelligence 12(5), 489–497 (1990)
27. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biological cybernetics 43(1), 59–69 (1982)
28. Kohonen, T.: Essentials of the self-organizing map. Neural networks 37, 52–65 (2013)
29. Kundu, A.K., Fattah, S.A.: Probability density function based modeling of spatial feature variation in capsule endoscopy data for automatic bleeding detection. Computers in Biology and Medicine 115, 103478 (2019)
30. Kurtz, C., Depeursinge, A., Napel, S., Beaulieu, C.F., Rubin, D.L.: On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. Medical image analysis 18(7), 1082–1100 (2014)
31. Lokesh, S., Kumar, P.M., Devi, M.R., Parthasarathy, P., Gokulnath, C.: An automatic tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map. Neural Computing and Applications 31(5), 1521–1531 (2019)
32. Maida, M., Celsa, C., Lau, L.H., Ligresti, D., Baraldo, S., Ramai, D., Di Maria, G., Cannemi, M., Facciorusso, A., Cammà, C.: The application of large language models in gastroenterology: A review of the literature. Cancers 16(19), 3328 (2024)
33. Martinez, A.M., Kak, A.C.: Pca versus lda. IEEE transactions on pattern analysis and machine intelligence 23(2), 228–233 (2001)
34. Muhammad, K., Khan, S., Kumar, N., Del Ser, J., Mirjalili, S.: Vision-based personalized wireless capsule endoscopy for smart healthcare: Taxonomy, literature review, opportunities and challenges. Future Generation Computer Systems 113, 266–280 (2020)
35. Nilashi, M., Ahmadi, H., Manaf, A.A., Rashid, T.A., Samad, S., Shahmoradi, L., Aljojo, N., Akbari, E.: Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. International Journal of Fuzzy Systems 22(4), 1376–1388 (2020)

36. Pannu, H.S., Ahuja, S., Dang, N., Soni, S., Malhi, A.K.: Deep learning based image classification for intestinal hemorrhage. Multimedia Tools and Applications 79, 21941–21966 (2020)
37. Papakostas, G., Boutalis, Y., Karras, D., Mertzios, B.: Efficient computation of zernike and pseudo-zernike moments for pattern classification applications. Pattern Recognition and Image Analysis 20(1), 56–64 (2010)
38. Park, S., Han, S., Kim, S., Kim, D., Park, S., Hong, S., Cha, M.: Improving unsupervised image clustering with robust learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12278–12287 (2021)
39. Patel, A., Rani, K., Kumar, S., Figueiredo, I.N., Figueiredo, P.N.: Automated bleeding detection in wireless capsule endoscopy images based on sparse coding. Multimedia Tools and Applications 80(20), 30353–30366 (2021)
40. Radhika, K., Venkatesha, M., Sekhar, G.: An approach for on-line signature authentication using zernike moments. Pattern Recognition Letters 32(5), 749–760 (2011)
41. Ren, Y., Wang, N., Li, M., Xu, Z.: Deep density-based image clustering. Knowledge-Based Systems 197, 105841 (2020)
42. Roussinov, D., Chen, H.: A scalable self-organizing map algorithm for textual classification: A neural network approach to thesaurus generation. Communication Cognition and Artificial Intelligence 15(1-2), 81–111 (1998)
43. Siddiqui, S.A., Salman, A., Malik, M.I., Shafait, F., Mian, A., Shortis, M.R., Harvey, E.S.: Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES Journal of Marine Science 75(1), 374–389 (2018)
44. Singh, C., Aggarwal, A.: An efficient approach for image sequence denoising using zernike moments-based nonlocal means approach. Computers & Electrical Engineering 62, 330–344 (2017)
45. Soffer, S., Klang, E., Shimon, O., Nachmias, N., Eliakim, R., Ben-Horin, S., Kopylov, U., Barash, Y.: Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. Gastrointestinal endoscopy 92(4), 831–839 (2020)
46. Tervonen, J., Puttonen, S., Sillanpää, M.J., Hopsu, L., Homorodi, Z., Keränen, J., Pajukanta, J., Tolonen, A., Lämsä, A., Mäntyjärvi, J.: Personalized mental stress detection with self-organizing map: From laboratory to the field. Computers in Biology and Medicine 124, 103935 (2020)
47. Xue-guang, C., et al.: An improved image segmentation algorithm based on two-dimensional otsu method. Information Sciences Letters 1(3), 2 (2012)
48. Yang, J., Yang, J.y.: From image vector to matrix: A straightforward image projection technique—impca vs. pca. Pattern Recognition 35(9), 1997–1999 (2002)
49. Yang, J., Parikh, D., Batra, D.: Joint unsupervised learning of deep representations and image clusters. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5147–5156 (2016)
50. Yellamraju, T., Boutin, M.: Clusterability and clustering of images and other "real" high-dimensional data. IEEE Transactions on Image Processing 27(4), 1927–1938 (2018)
51. Yuan, Y., Meng, M.Q.H.: Deep learning for polyp recognition in wireless capsule endoscopy images. Medical physics 44(4), 1379–1389 (2017)
52. Zhang, D.: Wavelet transform. In: Fundamentals of Image Data Mining, pp. 35–44. Springer (2019)
53. Zhang, J., Wu, Q., Zhang, J., Shen, C., Lu, J., Wu, Q.: Heritage image annotation via collective knowledge. Pattern Recognition 93, 204–214 (2019)
54. Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., Nguyen, H.T.: Automated segmentation of soybean plants from 3d point cloud using machine learning. Computers and Electronics in Agriculture 162, 143–153 (2019)
55. Zhu, W., Lu, J., Zhou, J.: Nonlinear subspace clustering for image clustering. Pattern Recognition Letters 107, 131–136 (2018)

Parminder Kaur is a postdoc fellow at Durham University UK. She did her PhD from Thapar Institute Patiala India. Her research interests are machine learning, cross-modal learning using image and text, and image segmentation.
https://www.durham.ac.uk/staff/parminder-kaur/
https://scholar.google.co.in/citations?user=qvk7yvAAAAAJ&hl=en

Avleen Malhi is faculty at WMG, University of Warwick UK. She was a postdoc fellow at Aalto University Finland and did her PhD from Thapar Institute Patiala India. Her research interests are machine learning, explainable AI, wireless ad hoc networks.
https://ieeexplore.ieee.org/author/37085517161
https://scholar.google.co.in/citations?user=bMA1WcMAAAAJ&hl=en

Husanbir Singh Pannu is an assistant professor at Thapar Institute Patiala India. He was a postdoc in University of Liverpool UK and Trinity College Dublin Ireland. He did his PhD from University of North Texas USA. His research interests are machine learning and data analysis.
https://scholar.google.co.in/citations?user=DNPuK98AAAAJ&hl=enoi=ao

Received: June 28, 2024; Accepted: December 17, 2024.

Computer Science and Information Systems 22(3):783–814 https://doi.org/10.2298/CSIS240701031R

An MDA-based Requirements Analysis Process for Service-Oriented Computing Applications

Laura C. Rodriguez-Martinez[1], Hector A. Duran-Limon[2], Francisco Alvarez-Rodriguez[3], and Ricardo Mendoza-González[1]

[1] Tecnológico Nacional de México/IT Aguascalientes, Av. Adolfo López Mateos 1801 Ote., Bona Gens, CP 20256, Aguascalientes, Ags., Mexico
laura.rm@aguascalientes.tecnm.mx, mendozagric@aguascalientes.tecnm.mx
[2]CUCEA, University of Guadalajara, Periférico Norte 799, Zapopan, Jalisco, Mexico
hduran@cucea.udg.mx
[3] Autonomous University of Aguascalientes, Av. Universidad 940, Ciudad Universitaria, CP 20131, Aguascalientes, Ags., Mexico
fjalvar@correo.uaa.mx

Abstract. We propose an MDA-based requirements analysis process for Service-Oriented Computing Applications (SOCA). Our process is based on an analysis that identifies the most relevant elements of previous proposed requirements-processes. From the reviewed requirements-processes we identify such elements in terms of phases, activities, products, and roles/viewpoints. We reviewed proposals that include or emphasise the *process definition*, the *definition of products and models*, and *service-oriented modeling issues*. Also, we selected proposals within different research areas, namely Software Engineering (SE), Model-Driven Architecture (MDA), and Service-Oriented Computing (SOC). We carried out such analysis of previous requirements-processes by employing a comparative framework. We also studied some surveys about new proposals that define processes in MDA-based approaches. The main contribution of this work is a general *requirements analysis process for SOCA* called SOCA-rap that includes its activities and products allocated and grouped over a general development process. This general development process is structured in two dimensions where the first dimension involves four general activities, namely Requirements, Design, Construction, and Operation. The second dimension includes the three MDA models, namely the Computational Independent Model, the Platform Independent Model, and the Platform Specific Model. Additional contributions of this paper include (i) the identification of the phases, activities, products and roles/viewpoints of the processes of previous approaches of requirements analysis, (ii) a comparative framework of such elements, and (iii) the identification of the products included in the MDA models of the general development process.

Keywords: Service-Oriented Computing (SOC), Requirements Analysis, Architectural Design, Requirements Engineering, Model-Driven Architecture (MDA), Rational Unified Process (RUP), Service-Oriented Computing Applications (SOCA), Service-Oriented Software Engineering (SOSE).

1. Introduction

In the Software Engineering (SE) research area there are many traditional process-oriented software development models (or methodologies). As the term "process-oriented" indicates, each model describes the way to execute the development process by at least defining their phases / activities and the order to execute them. Also, such traditional process models are better structured when they define most of their elements (i.e. phases, activities, products, and roles [19]). On the other side, a special non-process-oriented development model, specifically Model-Driven Architecture (MDA) is a framework that proposes carrying out software development as a set of model refinements. The transformations between models become first class elements of the development process; therefore, a great deal of work takes places in these transformations [5]. In the process-oriented approaches the transformations between models "are implicitly defined by skilled architects in a software project" [5]. Thus, the MDA framework eases both the modelling and the model-to-model transformation. Although the MDA framework has been employed in industry and academia for software development, no specific process development model has been adopted. Also, MDA does not provide a "best practice guidance on how to maintain synchronized models on a large-scale development effort" [5]. Currently, when a practitioner uses MDA, he/she must define -in an explicit or implicit way-, a process for working (i.e., the practitioner must define a process-oriented MDA approach, commonly for a domain specific use) [5].

A process-oriented approach and an MDA framework complement each other into a ***process-oriented MDA approach*** -i.e., a hybrid approach-, by describing or capturing the model transformations explicitly integrated with a process definition. Specifically, the MDA approach for software development facilitates the analysis and transformation of the system models from the problem domain model to the executable model. Then, if we address the MDA models following an order, i.e., first the Computation Independent Model (CIM), next the Platform Independent Model (PIM) and finally the Platform Specific Model (PSM), the first problems to solve are related to the CIM.

Such an integration improves the way to develop software by merging the explicit way for modelling and performing model transformation of the MDA approaches with the explicit way for guiding the process of traditional software development process (SDP) models in SE. However, some works such as [20] recommend such an integration in both senses by either aggregating models -MDA issues- in a life cycle of software development or defining processes -SE issues- in an MDA framework for developing software. This is challenging because the integration requires great expertise in processes and modelling. As explained in [5] "*The spirit and intent of MDA is an approach that encourages use of models as the basis for software development. The ultimate vision of MDA is that models are used through the life cycle, with formal transformations between model refinements. In practice, that vision can only be realized by a very small percentage of software development organizations: the visionaries.*"

Also, Service-Oriented Software Engineering (SOSE), a parallel research area of SE, regards the study of the service-oriented paradigm [25]. In SOSE, several process models for developing Service-Oriented Computing Applications (SOCA) have been defined. Some of them focus on business modelling, system design or the use of new emergent technology [22], and a few of them cover either all the SOC applications-

development life cycle or the process details [25, 26]. Specifically, the CIM modelling of a SOCA presents two big problems: 1) what to model? i.e., what are the products that represent a CIM for SOCA; and 2) how to model it?, that is, the process to develop a SOCA. On the other side, there exists a big problem to solve when we use MDA to develop software, specifically, the issue that face practitioners for defining an integration of a process-oriented approach with an MDA approach. We believe that a SOSE methodology can be defined as a process-oriented MDA approach. Hence, the general problem of defining a SOSE methodology implies several specific problems like the definition of the detailed processes of each development phase -requirements, design, construction, and operation- [22] for SOCA. Also, the general problem of defining an MDA-based SOCA development approach has specific problems such as defining the products of each MDA model and determining when such models must be developed throughout the development process.

This paper reviews scientific literature that, in the last years [20, 3], reports some surveys and proposals that study how to integrate the definition of the processes in new MDA approaches. This paper also studies the elements of Requirements-processes covering approaches of MDA, SOSE, and SE reported during the years from 1976 to 2019. And finally, we propose an MDA-based Requirements Analysis Process for SOCA that emphasizes the activity of modelling as well as defining the process that is needed to construct the products, and a way to iterate over the process activities. Model-Driven Software Development (MDSD) is a more generic approach than MDA, which involves a process. This process includes the following steps [31]: requirements modelling, model transformation (PIM to PSM), code generation (PSM to code), testing, deployment, and maintenance and evolution. However, the steps conceived for this process are rather general and do not provide detailed activities of each of the steps or the input and output artifacts of each activity. Thus, we adopt the transformation phases of MDA and propose a detailed Requirements Analysis Process for SOCA. Furthermore, the way to iterate over the proposed process activities is through the use of a generic software development process. This development process is structured in two dimensions. The first dimension involves four general activities, namely Requirements, Design, Construction, and Operation. The second dimension includes the three MDA models, namely the Computational Independent Model, the Platform Independent Model, and the Platform Specific Model.

The remainder of this paper is structured as follows. Section 2 defines a conceptual framework for the analysis of elements (i.e. phases, activities, products, and views) of the process requirements models. In Section 3, the framework of the previous section is applied to identify the elements of the ten selected requirements-processes -specially their products-; resulting in the classification and generalization of the products of the requirements processes under study. Section 4 proposes the products of the requirements processes grouped into the MDA models (CIM, PIM, PSM). Section 5 proposes a new MDA-based Requirements Analysis Process for SOCA. Section 6 presents an example of using the new MDA-based Requirements Analysis Process for SOCA. Section 7 presents an evaluation of the proposed process for SOCA. Finally, some concluding remarks are given in Section 8.

2. Framework for the Requirements-Analysis Process

Our framework for the ***Requirements-Analysis Process*** includes the following. First, it defines one criterion that includes a way for the *Classification of requirements process models.* Next, our framework describes how we can classify the *Elements of the software-system development process* in a template (see Table 1); this for the analysis of the elements of each specific requirement process under study. Finally, the framework proposes *three big viewpoints for classifying the products of the requirements processes,* and a template (see Table 2) based on such viewpoints to compare the products of the proposals.

2.1 Classification of requirements process models

The selected proposals focus on CIM modelling and offer requirements and architectural-design processes. Each proposal is classified as **RA**, **RE**, **AD**, **SOSE** and **MDA**, depending on its research area focus, respectively: *Requirements Analysis*[1], *Requirements Engineering*, *Architecture Design*, *Service-Oriented Software System*, *and Model-Driven Architecture*. Then the criterion to classify the requirements models is explained as follows.

The **RA (***Requirements Analysis***)** term refers to the methodological proposals for requirements analysis prior to the appearance of the requirements engineering research area. The term has to do with the process of analysing requirements to (i) detect and resolve conflicts between requirements; (ii) discover the links within the software and how they interact with its organizational and operational environment; and (iii) define systems requirements to specify software requirements. This has the traditional view of requirements analysis (or requirements process), which boils down to conceptual modelling using an analysis method, such as the structured analysis method. The **RE (***Requirements Engineering***)** term is widely used in the field to denote the systematic handling of requirements. This term represents a more complete vision of the requirements process that involves requirements management, an engineer of requirements as a role, and the classification and negotiation of the requirements. The **AD (***Architectural Design***)** term refers to the methodological proposals where software design activities are involved in the requirements process. Such proposals implicate that it is impossible the total separation of the two tasks -i.e., requirements and design processes are not disjoint-. In most cases the software engineer acts as a software architect too, since the processes of analysis and elaboration of requirements demands the identification of components of architecture/design that implements the requirements specification [4]. The **SOSE** (Service-Oriented Software Engineering) term refers to the methodological proposals where a requirements process is included as part of the Service-Oriented Software Systems development process.

1 Requirements Analysis -that was rather an activity-, exists before Requirements Engineering. RE emerged later on as a process that includes RA as an activity and other activities to complete the process. Afterwards, either RE substituted RA or RA was absorbed and included in RE as an activity.

The **MDA** (Model-Driven Architecture) term refers to the methodological proposals that considers or studies the requirements process as part of the Model-Driven Development approaches.

2.2 Elements of the software-system development process

For the analysis of the components of the proposals, we consider three components of the software-system development processes (SSDP) [19]. The first two components are the ***phases*** and the ***activities*** of the SSDP, which properly define the process. The third component regards the ***products*** that are constructed and interchanged during the SSDP through the phases and activities. Also, four components, defined by [4, 6, 12, 13], that are widely interrelated are considered in the analysis. These components are *view, viewpoint, stakeholder, and abstraction level*. Where, a ***view*** is made up of all the models that have to do with a specific viewpoint (point of view), represented at a certain level of abstraction or limited to a user perspective (or stakeholder). A ***viewpoint*** is a technical abstraction that focuses on a particular set of concerns that are part of the system, removing irrelevant details -and also, a viewpoint can be represented by one or more models-. A ***stakeholder*** is a person who has to do with the System in some way, as a user, the owner, the administrator, etc. The ***abstraction level*** refers to the level of detail in which a System model is represented.

Furthermore, a view can consist of one or more products. Also, each product can be represented by a specific point of view or stakeholder at different levels of abstraction. They are important components since they need to model the requirements model, in a way that facilitates its transformation into a design model and maintain the traceability between both models [2]. Table 1 presents a template for the analysis of the requirements processes under study. This template includes some numbered characteristics or aspects that are (1) the type of proposal, that is classified in RA, RE, AD, MDA, SOSE; (2) the name of the proposal; (3) the bibliographical reference of the proposal; (4) the components of the proposal (phases, activities, products) regarding the requirements process that the proposal defines in an implicit or explicit way; and (5) viewpoints, stakeholders, views, and abstraction levels that the proposal considers for representing and/or dividing the requirements.

Table 1. Framework of component-analysis of processes

Type of proposal: (1)			
Name of the proposal: (2)			
Reference: (3)			
Phase of the software-development process (4)	Products (4)	Activities (4)	Viewpoint /Stakeholder / view / abstraction level (5)
	-	-	

2.3 Three big viewpoints for classifying the products of the requirements processes

To classify, compare, and normalise the products of the requirements processes under study, we use the framework presented in Table 2. This framework is based on three big viewpoints that we consider cover the typical concerns of the products of the requirements processes. These big viewpoints are the *business viewpoint*, the *system viewpoint* and the *project viewpoint*. Where the ***business viewpoint*** involves elicitation & analysis and covers the viewpoint of the user. The ***system viewpoint*** regards modelling and covers the viewpoint of the system analyst. And the ***project viewpoint*** regards project planning and risk management issues. Then for the normalisation of the products of the requirements processes, the products are classified into these three viewpoints.

Table 2. Comparative framework to compare and normalise the products of requirements-processes

Viewpoint	P1	…	Pn
Business (elicitation & analysis, user viewpoint)			
System (modelling, system analyst viewpoint)			
Project (project planning, risk management)			

We present the complete framework, although the scope of this study does not include products and activities of the project viewpoint since they are not aligned with any MDA model.

3. Analysis of Ten Requirements-Processes

First, we present some relevant information of each proposal, highlighting the products proposed and presenting examples of how the identified products can be classified into the three viewpoints: business, system and project. Table 3 presents the list of the requirements proposals (or requirements-process models) that we analysed, referred to as P1..P10. First, we present some relevant information of each proposal, highlighting the products proposed and presenting examples of how the identified products can be classified into the three viewpoints: business (elicitation & analysis, viewpoint of the user), system (modelling, viewpoint of the system analyst) and project (project planning and risk management).

The proposals reviewed covers the software development processes for modelling a CIM. These proposals are commonly either requirements or architectural design processes. Then, through the analysis of such proposals, we identify and classify their activities, products, and the views for the representation of the product models.

Table 3. Proposals of CIM modeling and of requirements processes

Proposal	Reference	Type of proposal	Name of the proposal
P1	Leite, 1988 [12]	RA	"Viewpoint Resolution in Requirements Elicitation"
P2	Boehm, 2004 [27]	RA	(MBASE)
P3	Getchell, 2002 [29]	RE	"MSF Process Model v. 3.1"
P4	Hofmeister, 2005 [8]	AD	"Global Analysis"
P5	Hofmeister, 2007 [9]	AD	"General Model of Software Architecture Design Process"
P6	Péraire, 2007 [21]	SOSE & RE	(RUP-SOA)
P7	Habe, 2013 [7]	AD	"The Missing Link -between Requirements and Design"
P8	Bourque, 2014 [4]	RE	(general model proposed in SWEBOK V3.0)
P9	Asadi, 2008 [3]	MDA	"An MDA-based System Development Lifecycle"
P10	Rodríguez-Martínez, 2019 [22]	MDA & SOSE	(SOCA-DSEM)

Below we present a brief description and analysis of the elements of each reviewed proposal.

P1. "Viewpoint Resolution in Requirements Elicitation" [12]. In [12] Leite (1988) proposes a process-oriented definition of RA: "*Requirements analysis is a process which '**what is to be done**' is elicited and modeled. This process has to consider different viewpoints, and it uses a combination of methods, tools and actors. The product of this process is a model, from which a document, the requirements is produced*". This work proposes two documents as products: a "Requirements analysis model" and "Requirements document". The first document is defined from the viewpoint of the analyst, is internal to the development process, is written in a machine-processable form, and must be readable by analysts and designers to understand the application to build. The second document reflects the viewpoint of the user, is in agreement with the user of what he/she demands from the system and must be readable by both users and analysts [12]. Such documents correspond to two viewpoints, specifically the "business" and "system" viewpoints.

P2. "MBASE" [27]. It is an approach that integrates the process-models, products, properties, and the success criteria to develop a software system. The essence of the approach is to develop some elements of systems definition in a concurrent and iterative way (or by refinement) by using the Win-Win Spiral model of Boehm [27]. Such elements are the following: Operational Concept Description (OCD), System and Software Requirements Definition (SSRD), System and Software Architecture Description (SSAD), Life Cycle Plan (LCP), Feasibility Rationale Description (FRD), Construction Transition Support (CTS) plans and reports, and Risk-driven prototypes. Issues of risks are implicit in MBASE since it is a risk-driven approach. The risk control is implicit in the way to proceed, i.e., it does not indicate products or documents of risk planning or management, e.g., it uses risk-driven prototypes.

P3. "MSF Process Model V.3.1" [29]. In [29], Getchell proposes several products in two phases: (i) the Envisioning Phase, is where the project is defined and planned; and (ii) the Planning Phase, is where requirements are specified and the software development is planned. In such phases several products are proposed, for the three big viewpoints. For example, for the single activity in the Envisioning Phase it has the following products for each viewpoint: (1) for the business viewpoint: proposes the "vision/scope document", (2) for the project viewpoint: proposes the product "risk

assessment document and project structure document", and (3) for the system viewpoint: proposes the product of "prototypes". Also, in the case of the Planning Phase the proposal has the product "Business, user, operational and systems Requirements List" in the business viewpoint.

P4. "Global Analysis" [8]. Hofmeister in [8] proposes the activity "Global Analysis" as an intermediary between the processes of the requirements analysis and the architecture design. Such a Global Analysis helps to guide the design process by capturing the design rationale and supporting traceability between requirements and architecture. This proposal has three products. The product Software Requirements Specification (SRS) belongs to the business viewpoint. The other two products, namely the Architecture Design Requirements (ADR) and the System Solutions Strategies (SSS) are part of the system viewpoint. Although ADR and SSS are constructed in the Global Analysis, ADR is part of the requirements specification whereas SSS is part of the architecture specification. ADR is located on the problem (requirements) side. Such an architecture must identify those requirements that affect the architecture design and also, must identify any additional quality attributes and constraints (e.g., pre-imposed design decisions and limitations, use of COTS software). It is also important to identify other factors that affect the architecture (e.g., culture of the organization, state of technology, existence of a related product line). On the other side, SSS is a set of strategies that guide the architecture design.

P5. "General Model of Software Architecture Design Process" [9]. This proposal defines the activity of Architectural Analysis (AA) that is an intermediary activity between the requirements analysis and the architectural design. Products of AA are: (i) Context Requirements and (ii) Architectural Significant Requirements. The requirements analysis presents the product "Context Requirements" for the business viewpoint and the product "Architectural Concerns" for the system viewpoint.

P6. "RUP-SOA" [21]. RUP-SOA [21] is an extended version of RUP for SOA (o RUPz). RUPz has two dimensions. Each dimension corresponds with an axis. The first dimension is the ***time*** represented over the horizontal axis. The second dimension corresponds with the ***disciplines*** or activities represented over the vertical axis. The horizontal axis represents also, the software development life cycle that is divided into four phases: Inception, Elaboration, Construction, and Transition. Each phase is divided in one or more iterations, this, according to the project needs. In addition, the vertical axis represents disciplines, such as requirements, analysis, design, or implementation, logically grouped according to its nature. The requirements products of RUPz are classified into the viewpoints of "business", "system" or "project": e.g., for the business viewpoint there is the product "Business Case", for the system viewpoint there is the product "Use-Case Model", and for the project viewpoint there is the product "Test Strategy".

P7. "The Missing Link -between Requirements and Design" [7]. In this proposal, Habe [7] describes Product Abstraction Levels (PAL) aligned with the V-Model. This shows the transformation Trail of the information and shows how the product is developed by keeping consistency and completeness, despite the high-fragmented activities of the engineering and the increasing parallelization of the sub-processes (like requirements-process) of the Product Creation Process (PCP). This proposal also tells us how the requirements are interconnected and that the requirements process is not integrated with the design process. The requirements product involves transforming the "stakeholder requirements" to the "functional and product requirements" which in turn

are transformed to the "system requirements". The latter are then transformed to the "technical requirements of implementation" which are then transformed to the "implementation".

P8. "General model defined by the SWEBOK V3.0" [4]. Bourque (2014) [4] proposes a product for three big viewpoints: e.g., the product "System Definition Document" for the business viewpoint, the product "System Requirements Document" for the business and system viewpoints, and the product "Software Requirements Specification Document" for the system viewpoint and "prototype" for the system and business viewpoints.

P9. "MDA-Based System Development Lifecycle" [3]. Asadi [3] proposes a software-system development life cycle (of the software development process) based on MDA. The proposal comprises five phases: Project Initiation, PIM Development, PSM & Code Development, Deployment, and Maintenance. Of such phases, the phase PIM Development and the phase PSM & Code Development are typically executed in an iterative and incremental way. This approach proposes products for two viewpoints: business and project. But the requirements process (project initiation) does not include products for the big system viewpoint, i.e., the approach advocates a separate vision to construct requirements and design.

P10. "SOCA-DSEM" [22]. The proposal P10 is an MDA approach that is focused on the development of SOSS and indicates what products are part of CIM, PIM, PSM, and PM. Although SOCA-DSEM proposes CIM products [29], it does not consider activities for requirements engineering. The proposed activities are more like sub-phases. It proposes that each MDA model should be constructed thoroughly in a phase or activity. It also proposes products, and artifacts to be constructed in a specific notation. The requirements phase has products that can be classified as the business and project viewpoints.

Regarding the component-analysis of the proposals of requirements-processes, Section 4 presents the normalisation of the products identified for each of one of the proposals under study.

4. Products of Requirements process classified into CIM, PIM or PSM

This section proposes the products that are part of the models CIM, PIM or PSM. We propose a comparative framework -that is shown in Table 2- to compare the products of the proposals. The framework enables the normalisation of the products. This normalisation involves identifying the products that can be part of a CIM, PIM or PSM.

We use the conceptual framework to identify the components, specifically regarding the products that compose the ten proposals under study. Then, given the identified products, we carried out a normalisation (or generalization) of the products on the requirements-process. Table 2 presents a comparative framework, in which the first column indicates the three viewpoints (business, system, and project) and the next n columns present the products of each proposal. The result of such a normalisation is presented in Table 4 and Table 5 involving the products of the business and system viewpoints, respectively. The general products of the requirement process, identified in

the previous section, are indicated in Table 4 and Table 5 with a their names and a label i.e., CIM-1..CIM-6, PIM-1..PIM-2, PSM-1.

Table 4. Normalisation of products oft he business viewpoint

Products maped from Requirements proposals reviewed regarding Business viewpoint										**SOCA-rap proposal CIM model**	
P1 RA	**P2 RA**	**P3 RE**	**P4 AD**	**P5 AD**	**P6 SOSE & RE**	**P7 RE**	**P8 RE**	**P9 MDA**	**P10 MDA-SOSE**	**products**	**docu-ments**
- Facts - Application Vocabulary	- Stakeholders list - Win-Win conditions	Vision/scope document	Software requirements specification (SRS)	X	Business case	X	System Requirements Document	Computational Independent Model (CIM)	X	CIM-1. Vision/scope (or businesss case document) CIM-2. Stakeholders list CIM-3. Win-win conditions CIM-4. Context (Facts & Application Vocabulary)	**System Definition Document**
Requirements document	X	Requirements List		Conceptual Business Analysis	X	- Operational Use level requirements - Functional level requirements			Business Process Model (BPM)	CIM-5. Requirements List CIM-6. Business Process Model (BPM)	**System Requirements Document**

As shown in Table 4 and Table 5:

(i) The products of the requirements processes are classified into the business viewpoint and are aligned with CIM. Also, the products of requirements processes, classified into the system viewpoint are aligned with PIM and PSM.

(ii) The products of the Requirements-process that were identified as part of the CIM, PIM, PSM models are: for CIM are CIM-1, .. , CIM-6, for PIM are PIM-1 and PIM-2 and for PSM is PSM-1.

The first version of the Requirements documents implies evolution during the development process. It is assumed, that the CIM documents and the document "Software Requirements Specification" enable requirements changes. Requirements evolve along all the development phases if there are changes on the requirements specification. In contrast, the "Initial Design Document" is assumed to be disposable in the early stages of the development process as it is about quick experiments with some non-detailed technological alternatives.

Table 5. Normalisation of products of the system viewpoint

Products maped from Requirements proposals reviewed regarding System viewpoint										**SOCA-rap proposal PIM and PSM models**	
P1 RA	**P2 RA**	**P3 RE**	**P4 AD**	**P5 AD**	**P6 SOSE & RE**	**P7 RE**	**P8 RE**	**P9 MDA**	**P10 MDA-SOSE**	**products**	**docu-ments**
Requirements analysis Model	- Operational Concept Description (OCD) - System and Software Requirements Definition (SSRD)	Functional Specification	X	X	- Vision - Use-Case Model - Supplementary Specifications - Glossary	Systems Requirements	Software Requirements Specification Document	Requirements Model	Enterprise architecture	PIM-1. Functional Specification of software (includes a minimal enterprise architecture for services)	**Software Requirements Specification (SRS)**
	System and Software Architecture Description (SSAD)	Design document (initial)	Architecture Design Requirements (ADR)	- System Context of architectural concerns - Architectural Concerns - Architecturally significant requirements (ASR)	Software Architecture Document	- Functional architecture - Systems architecture	X	X	System Solution Architecture: (- Service Provider , design, - Service Consumer design, - Database design)	PIM-2. System Architecture Solution (includes service provider design, service consumer design, database design)	
	Prototype	Prototypes, development and technological options, feasibility analysis	X	X	X	Models and design options	Prototype	X	Preliminary user interface design as a prototype	PSM-1. Prototypes, development and technological options, feasibility analysis and preliminary user interface design	**Initial Design Document**

Finally, Table 6 describes the identified products of the CIM, PIM and PSM as follows.

Table 6. CIM, PIM and PSM products of requirements-process for SOSS

CIM products	
CIM-1. Vision/scope (or business case document).	A high-level vision of the objectives and constraints of the project.
CIM-2. Stakeholders list	Indicates names commonly as roles of the stakeholders, and the description of the function and tasks of such roles.
CIM-3. Win-Win conditions	It Specifies the stakeholder win-win relationship throughout the system's life cycle. It involves conditions related to the stakeholders to reach success by getting the stakeholders to clarify, understand, and reconcile their success models.
CIM-4. Context (Facts & Application Vocabulary)	This regards the context of the system but mainly the Facts and the Application Vocabulary (or domain terms). A Fact refers to gathering information for better understanding an application. More specifically, it refers to the domain knowledge that can be gathered from documents, pre-printed forms, information between stakeholders, interviews with stakeholders, etc. The Application Vocabulary standardises the use of a common set of words in the model of the application and in the application itself.
CIM-5. Requirements List	This involves the Operation Use level requirements and Functional level requirements that are agreed with the user [7]: Operational Use level requirements captures the customer needs, the operational use terms and conditions, product roadmap, product scope, and the development-process conditions. Functional level requirements capture the properties, functions and constraints of the system and how to validate and ensure the behaviour and feasibility.
CIM-6. Business Process Model (BPM)	It includes the definition of the process and workflows at the business level and the identification of the services and their conceptual definition.
PIM products	
PIM-1. Functional Specification (includes a minimal Enterprise Architecture for Services)	This regards a set of requirements specification in machine-processable form that includes an early design [13] (i.e. the Functional Specification and the Enterprise architecture). The Functional Specification describes in detail how each feature (or requirement) looks like and behaves. It also describes the architecture and the design of all features [29]. It has Instructions for the developers on what to build, the Basis for estimating work, the Agreement with customer on exactly what will be built, and the Point of synchronization for the whole team. The Enterprise architecture is a logical structure for classifying and organising the descriptive representations of an Enterprise that are significant to the management of the Enterprise as well as to the development of the Enterprise's systems, manual systems, and automated systems [24].
PIM-2. System Architecture Solution (an initial architecture design of service provider, service consumer and database design)	It is the candidate architectural solution that may present alternative solutions, and/or may be partial solutions (i.e., fragments of an architecture). It reflects design decisions about the software structure. The architectural solutions include information about the design rationale, that is, commentary on why decisions are made, which decisions are accepted or rejected, and the traceability of decisions to requirements is defined. It refers to the initial or candidate solution, constructed in the first development phases [22] that includes the initial Service operations definition for the service provider design, and the semantic specification of the processes and workflows for the service consumer design; and the initial database design such as an Entity and Relation diagram.
PSM products	
PSM-1. Prototypes, development and technology options, feasibility analysis.	Prototypes, development and technology options, feasibility analysis [29] refers to selecting and proving the technology and an initial design of the system through prototypes for analysing the feasibility of constructing the system with the selected technology. And the initial interface design.

5. The SOCA-rap Process

SOCA-rap is a general MDA-based process that integrates MDA models for Requirements modelling and a process to guide the development. Such a process is represented in the Figure 1 and includes the four generic phases of a software development life-cycle (SDLC) i.e. Requirements, Design, Construction, and Operation. The process also includes the models of MDA -CIM, PIM, and PSM- and an extension the model E-PM (Executable-Platform Model) [14]. Similar to the RUP process, the area below the curves indicates the amount of development effort required in each phase to construct each model. The number of X on the general MDA-based process (see Figure 1) indicates the average amount of effort required for constructing each MDA model in each phase. The SOCA-rap then defines the MDA products (see Table 6), the model-to-model transformations (see Figure 3) and the generic activities of the requirements analysis process (see Table 7) to be specifically used in the SOCA domain. The SOCA-rap also defines a very generic SDLC for iterative and incremental development (see Figure 1).

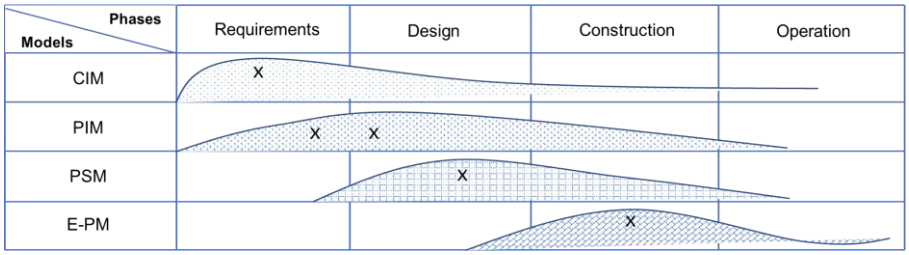


Fig. 1. General MDA approach of an SDLC

The distribution of activities and products is shown in Figure 2, which also presents the summary of our proposed MDA-based Requirements Analysis Process.

In Table 7 we define the general activities of the Requirements process presented in Figure 2, namely "Elicitation", "Analysis", "Modelling", and "Validation". We associate the products identified in the previous section (see Table 6) with such general activities. The three activities "Elicitation", "Analysis", and "Validation" corresponds to three of the four Knowledge Areas - "Requirements Elicitation", "Requirements Analysis", "Requirements Specification", "Requirements Validation"-, defined in [4]. Here we assume that the requirements specification is executed along all the requirement process. The products of the activities of Elicitation and Analysis are presented in terms of the business viewpoint. The names of such products have a prefix that indicates the MDA model to which each one belongs. Hence, the following products constitute the CIM model: CIM-1 Vision/scope (for the business case document), CIM-2 Stakeholders list, CIM-3 Win-Win conditions, CIM-4 Context (Facts & Application Vocabulary), CIM-5 Requirements List and CIM-6 Business Process Model (BPM). Besides, the PIM model includes the product PIM-1 Functional Specification (which includes a minimal Enterprise Architecture for Services) and PIM-2 System Architecture Solution. Finally, the PSM model for requirements includes only the product PSM-1 Prototypes, development and technology options, and feasibility analysis.

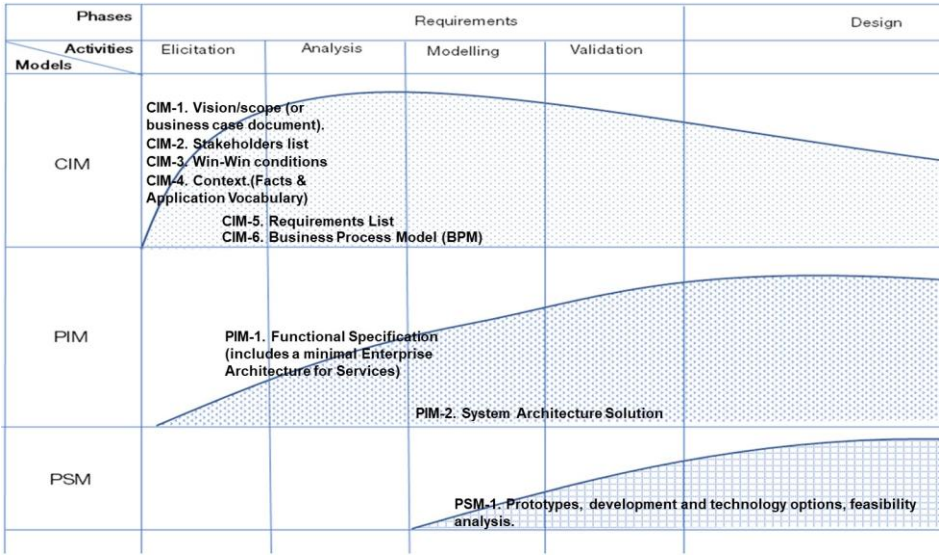


Fig. 2. MDA-based Requirements Analysis Process

The process is executed as follows. First, the requirements are taken during the activity of elicitation and analysis from the problem domain and then modelled (or transformed) as a CIM that is presented from a business viewpoint. Next, the modelling activity transforms the CIM into a PIM from the system viewpoint. Such a system viewpoint involves a model at a high-level of abstraction. At this time, we have the definition of requirements from the user and system perspectives [12]. Afterwards, the activity of modelling transforms the PIMs into PSMs. A PSM represents the requirements from the perspective of computing and involves a pre-design, e.g., it is expected that the PSM model includes some design of interfaces. Finally, the CIM, PIM, and PSM products are validated in the "Validation" activity.

We obtained the description of the requirements-process activities mainly from [4] and by matching with the description of the reviewed proposals -specially the descriptions of Elicitation, Analysis, and Validation were obtained from the methodology P10, and of the description of Modelling was obtained from the methodology P1-. Table 7 shows the description of the activities of the requirements process.

Finally, as shown in Fig. 3, the methodology uses three lines of model-to-model transformations [22], namely orchestration, choreography, and data lines. The first products constructed are CIM-1, CIM-2, CIM-3, and CIM-4, which are transformed into the CIM-5 Requirements List and the CIM-6 Business Process Diagram –involving both the Choreography and Orchestration lines-.

Also, in the Data transformation line, it is constructed a new version of CIM-4 Context by aggregating the Enterprise Data view or Application Vocabulary. Then the CIM products are transformed into the PIM products as follows. The products CIM-5 and CIM-6 in the Choreography line are transformed into the PIM-1 Functional specification (Use-case details). Also, CIM-5 and CIM-6 in the Orchestration line are transformed into the first version of the PIM-2 System Architecture Solution that

includes a System model diagram. Then PIM-1 in the Choreography line is transformed in a second version of PIM-2 that includes the Join Realization Table. Next, the second version of CIM-4 is transformed into a third version of PIM-2 that includes the E-R System data model. Finally, in the Choreography transformation line the third version of PIM-2 is transformed into PSM-1.

Table 7. Activities of the Requirements-Process

Activity	Description
Requirements Elicitation	It is "…concerned with the origins of software requirements and how the software engineer can collect them. It is the first stage in building and understanding of the problem the software is required to solve. It is fundamentally a human activity and is where the stakeholders are identified, and relationships established between the development team and the customer" [4].
Requirements Analysis	It is "… the process of analyzing requirements to detect and resolve conflicts between requirements; discover the bounds of the software and how it must interact with its organizational and operational environment; elaborate system requirements to derive software requirements. The traditional view of requirements analysis has been that it be reduced to conceptual modeling using one of a number of analysis methods, such as the structured analysis method" [4].
Modelling / Specification	Although the activities of Modelling and Specification are realised in parallel with each activity of requirements process, it is indicated here a modelling activity to highlight the modelling activity given that the specification part remains only as a parallel activity. Here the Modelling activity is considered a transition phase between the domain specification -conceptual model or domain model- and the systems requirements. In this activity, the system begins to be modelled/designed and the CIM-5 Requirements List, CIM-6 Business Process Model (BPM), PIM-1 Functional Specification and PIM-2 System Architecture Solution are completed. Also, in this activity is where it is initiated the elaboration of the products PSM-1 Prototypes, development and technology options, and feasibility analysis. Also, according to Leite (1988) [12] "*The resultant requirements analysis model, faces the problem that it has to serve different actors. First, it has to be readable by users and second, it should be the base for the designer's understanding of the application.*" Finally, the authors in [4] state that "The development of models of a real-world problem is key to software requirements analysis. Their purpose is to aid in understanding the situation in which the problem occurs, as well as depicting a solution. Hence, conceptual models comprise models to reflect their real-world relationships and dependencies. Several kinds of models can be developed. These include use diagrams, data flow models, state models, goal-based models, data models, and many others."
Requirements Validation	"The requirements documents may be subject to validation and verification procedures. The requirements may be validated to ensure that the software engineer has understood the requirements; it is also important to verify that a requirements document conforms to company standards and that it is understandable, consistent, and complete" [4].

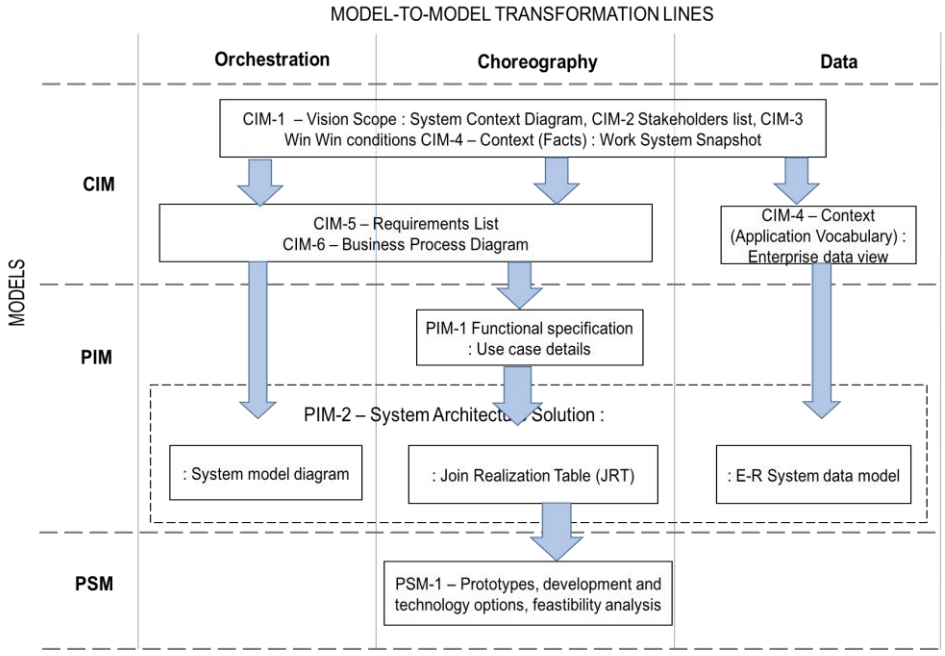


Fig. 3. Model-to-model transformation of SOCA-rap

SOCA-rap states how the model-to-model transformation process should be carried out. However, the details of how to carry out these transformations are out of the scope of this paper.

6. Illustrative Example

In this section, we illustrate the proposed requirements process by applying it to an example case. The illustrative example involves a system for managing job competences (JCM System- Job Competences Management System). The objective of this JCM System is to enable the users for managing such job competences, e.g. definition of competences, record of job competences of the personnel (or employees) of the organization, managing job descriptions by competencies, evaluation of job competences, and define career plans and training plans for the employees. The products of the CIM, PIM and PSM developed for SOSS, with the requirements-analysis process, are presented in the illustrative example. The notations used for the example, are optional, i.e. they are used only for illustration. Then, any other alternative notations can be used.

6.1 Computational Independent Model for the example

The product CIM-1 Vision Scope presents a high-level vision of the objectives and constraints of the project (see Figure 4). The notations for CIM-1 are from SOCA DSEM[2] [22] and UML[3] [34].

The product CIM-2 Stakeholders list indicates the names commonly used as roles of the stakeholders, and the description of the functions and tasks of each role; but, it is not included in this example for the sake of brevity. The product CIM-3 Win-Win Conditions is optional, hence, here it is not exemplified. In this case, we suggest using the approach of Boehm & Kitapci (2006) [28]. The product CIM-4 Context (Facts & Application Vocabulary) is exemplified in the Figure 5, which presents a work system snapshot[4] [32] and an excerpt of the application vocabulary as in [22].

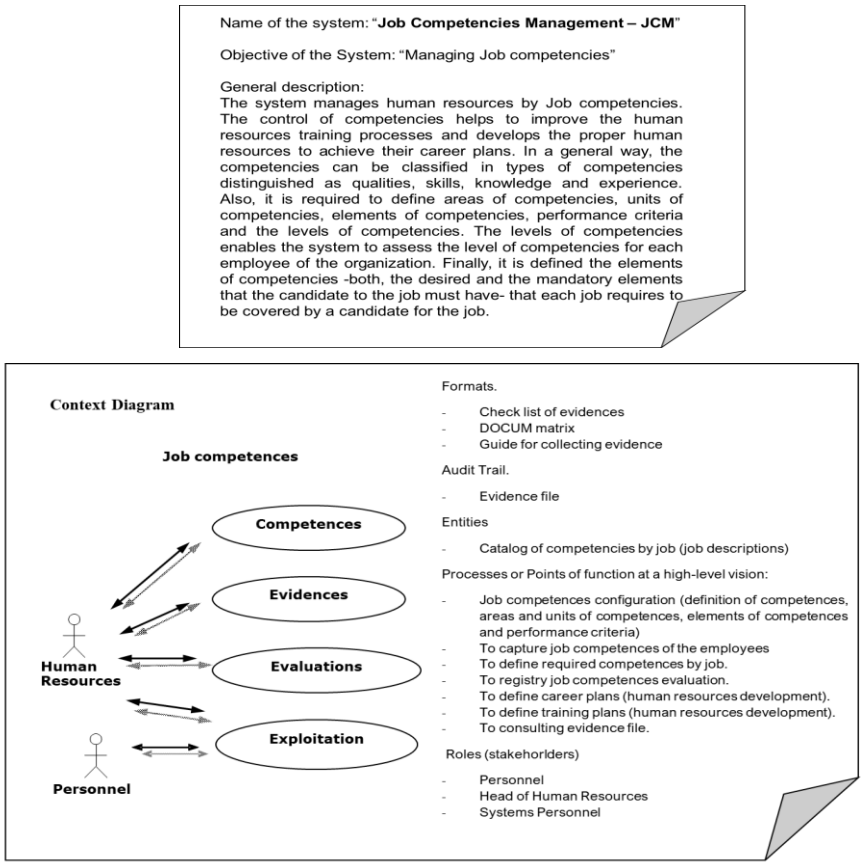


Fig. 4. CIM-1 Vision Scope

---

2 SOCA DSEM is a Software Engineering Methodology to develop Service-Oriented Computing Applications

3 UML is a Unified Modelling Language widely accepted by software engineering academicians and practitioners

Customer	Products and services
Personnel Head of Human Resources	Audit trail consulting and validation To configure job competencies, areas and units of competence, competency elements and levels of competence. To capture performance criteria for evaluating competences Scheduling training plans To define career plans To capture grades of competence evaluations To capture evidences of audit trail To consult career and training plans

Processes and activities

1. When the employee is hired, the Human Resources Head captures grades of competencies for a job of the employee and the respective evidences as an audit trail.
2. At the end of the first year of hiring, the Human Resources Head generates a Training and Career Plan of the employee .
3. From the second year, once in a year, the Human Resources Head captures for the employee new competencies, and the corresponding grades and evidences. The employee can generate changes on a training plan.
4. Each two years, the employee or the Human Resource Head can change the career or training plan.
5. At every moment, the Human Resources Head can consult candidates for a vacancy based on the grades of competencies of the employees for reaching career planes, independently of the current career plan.

Participants	Information	Technologies
Personnel Head of Human Resources Systems Personnel	- Audit trail information (grades, evidence, and real dates of evaluation) - Systems personnel database - Career Plans - Job descriptions - DOCUM matrix - Catalog of competencies - Evaluation criteria - Calendar of training	Database management Internet Worksheet for data exportation HTML templates

Definitions, acroyms and abbreviations

Standard of competence
The norm or standard of competence is an orderly description of the content of a work performance or work achievement. The standard states:
• Elements of competence.
• Performance criteria.
• Evidence of performance.
• Scope (field of application).
• Guide for collecting evidence of performance.

Elements of competence
It is the description of work achievements, grouped as units of competence. Each element of competence must be described in terms of: a verb (the action), an object (on which the action is performed), and the condition, which indicates the result of the action on the object. This includes required knowledge.

Performance criteria
It is a quality criterion with which the result of evaluation of the element of competence must be achieved. That is, it involves sources of verification on the quality of work achievement that the competence allows to obtain. They are classified as performance evidence (see types of performance evidence).

Evidence of performance.
It is the evidence that the achievement has been effectively achieved, this is under the established performance criteria.

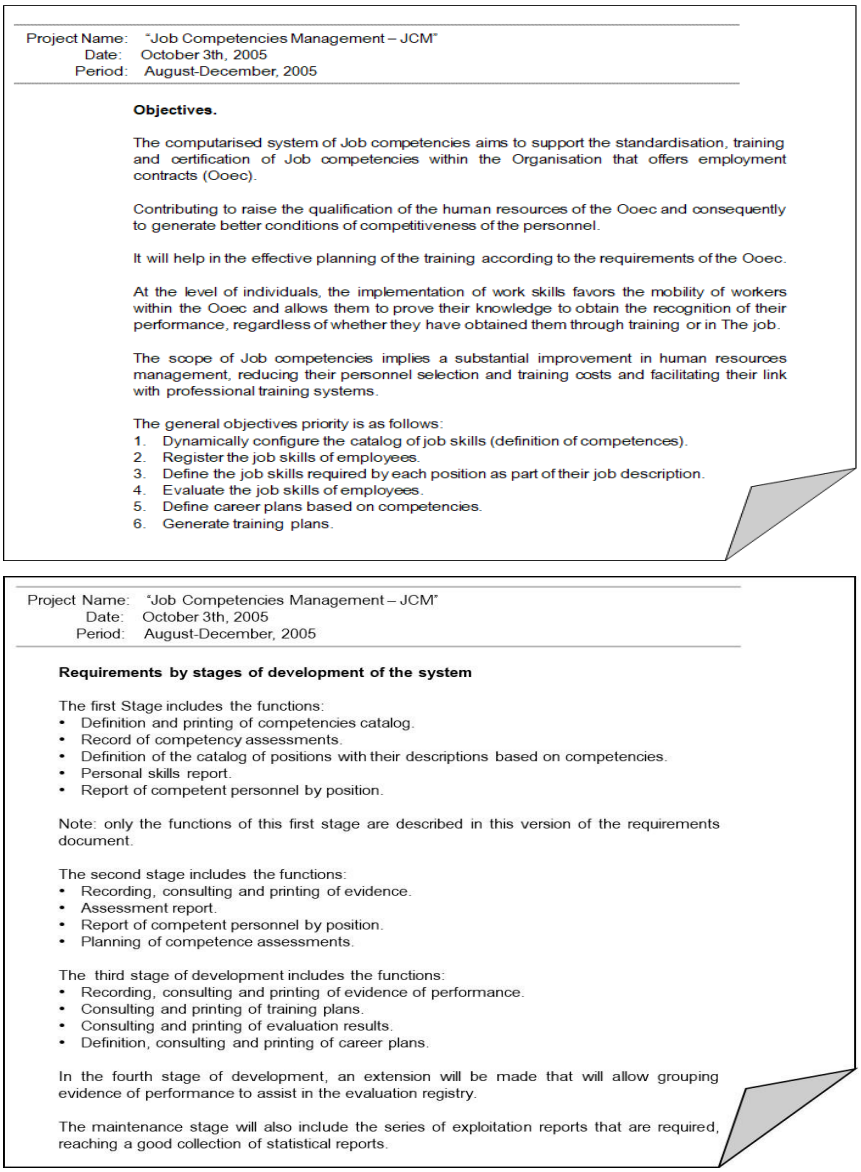
Scope (field of application).
Machines, tools, physical context, environment and interpersonal re-operation in which the scope performance is verified.

- - -

Fig. 5. CIM-4 Context (Facts & Application Vocabulary) represented as Work System Snapshot and Enterprise data view

4 Work System Snapshot is a format proposed by Alter [32] to show the relationship between business processes, their resulting products and services, their IT support and the stakeholders, from a business perspective.

Next, the product CIM-5 Requirements List presents, in a contextual way, the functions of the system to be constructed (see Figure 6). Figure 6 presents the objectives of the system and the requirements organised in system development stages. In addition, Figure 7 presents the contextual diagram of use cases of the system. Parts a and b of CIM-5 are presented in a non-specific or pre-established format. Part c is presented in the UML use case notation.



Project Name: "Job Competencies Management – JCM"
Date: October 3th, 2005
Period: August-December, 2005

Objectives.

The computarised system of Job competencies aims to support the standardisation, training and certification of Job competencies within the Organisation that offers employment contracts (Ooec).

Contributing to raise the qualification of the human resources of the Ooec and consequently to generate better conditions of competitiveness of the personnel.

It will help in the effective planning of the training according to the requirements of the Ooec.

At the level of individuals, the implementation of work skills favors the mobility of workers within the Ooec and allows them to prove their knowledge to obtain the recognition of their performance, regardless of whether they have obtained them through training or in The job.

The scope of Job competencies implies a substantial improvement in human resources management, reducing their personnel selection and training costs and facilitating their link with professional training systems.

The general objectives priority is as follows:
1. Dynamically configure the catalog of job skills (definition of competences).
2. Register the job skills of employees.
3. Define the job skills required by each position as part of their job description.
4. Evaluate the job skills of employees.
5. Define career plans based on competencies.
6. Generate training plans.

Project Name: "Job Competencies Management – JCM"
Date: October 3th, 2005
Period: August-December, 2005

Requirements by stages of development of the system

The first Stage includes the functions:
- Definition and printing of competencies catalog.
- Record of competency assessments.
- Definition of the catalog of positions with their descriptions based on competencies.
- Personal skills report.
- Report of competent personnel by position.

Note: only the functions of this first stage are described in this version of the requirements document.

The second stage includes the functions:
- Recording, consulting and printing of evidence.
- Assessment report.
- Report of competent personnel by position.
- Planning of competence assessments.

The third stage of development includes the functions:
- Recording, consulting and printing of evidence of performance.
- Consulting and printing of training plans.
- Consulting and printing of evaluation results.
- Definition, consulting and printing of career plans.

In the fourth stage of development, an extension will be made that will allow grouping evidence of performance to assist in the evaluation registry.

The maintenance stage will also include the series of exploitation reports that are required, reaching a good collection of statistical reports.

Fig. 6. CIM-5 Requirements list, part a: objectives and part b: requirements in system development stages

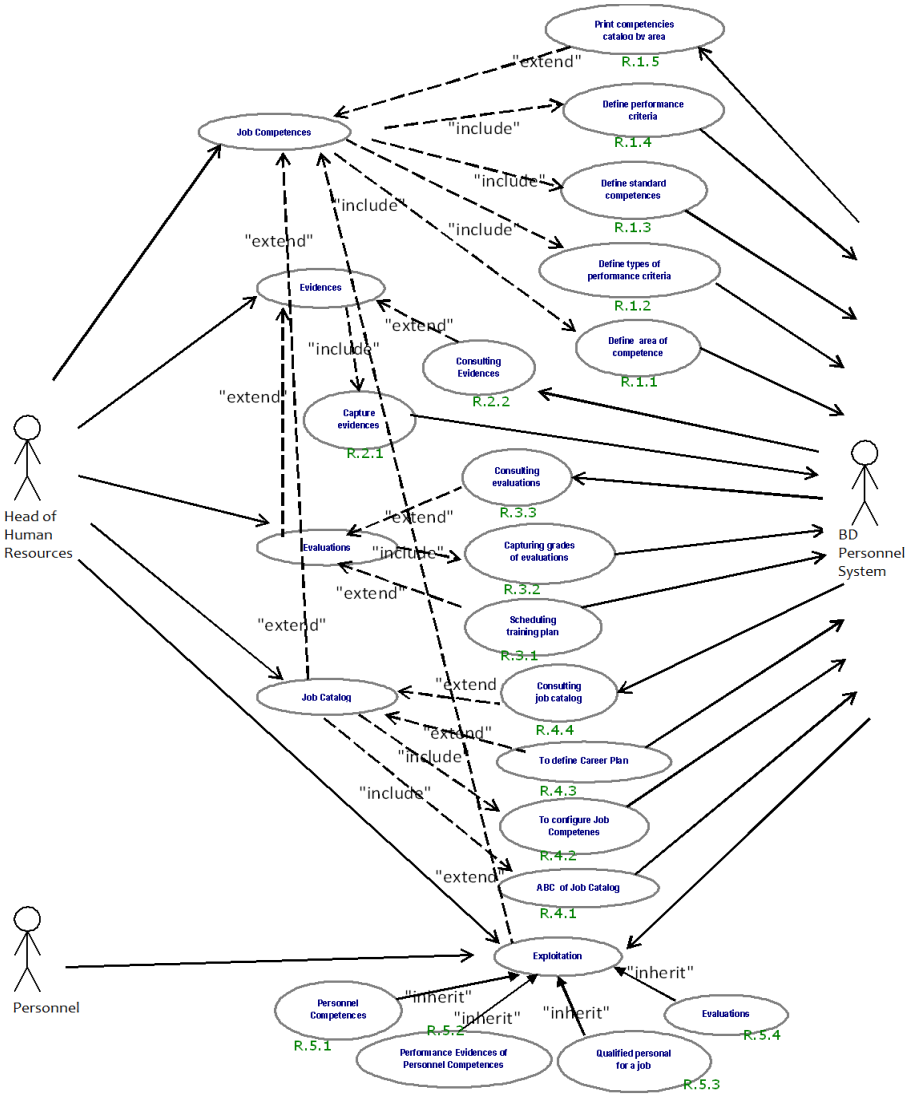


Fig. 7. CIM-5 Requirements list, part c: contextual diagram of use cases of the system

Finally, the last product of the model is CIM-6 Business Process Model. This model is presented in different perspectives - part a, b, and c (see Figures 8, 9 and 10)-. Figure 8 presents the private processes (part a), Figure 9 shows the process collaboration (part b), and Figure 10 depicts a detailed process diagram for the actor "Head of Human Resources". CIM-6 is presented in BPMN[5] as in [22].

5 BPMN is a Notation for Business Process Modelling

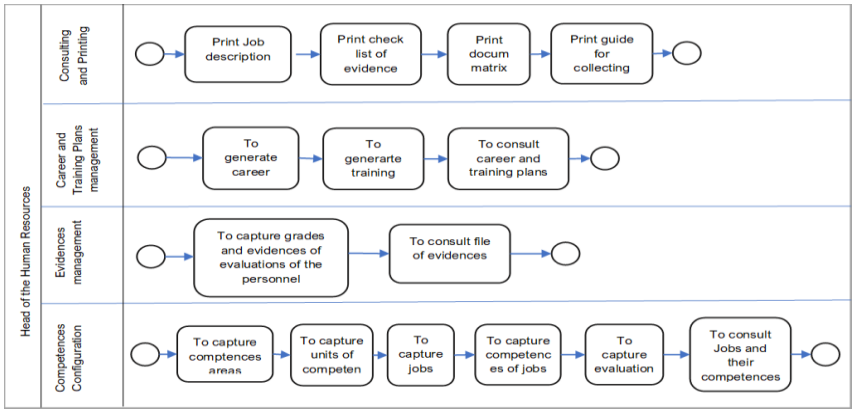


Fig. 8. CIM-6 Business Process Model (BPM), part a: Business process diagrams – private processes

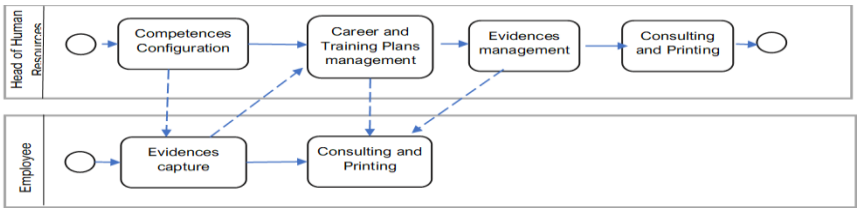


Fig. 9. CIM-6 Business Process Model (BPM), part b: process collaboration (actors' collaboration)

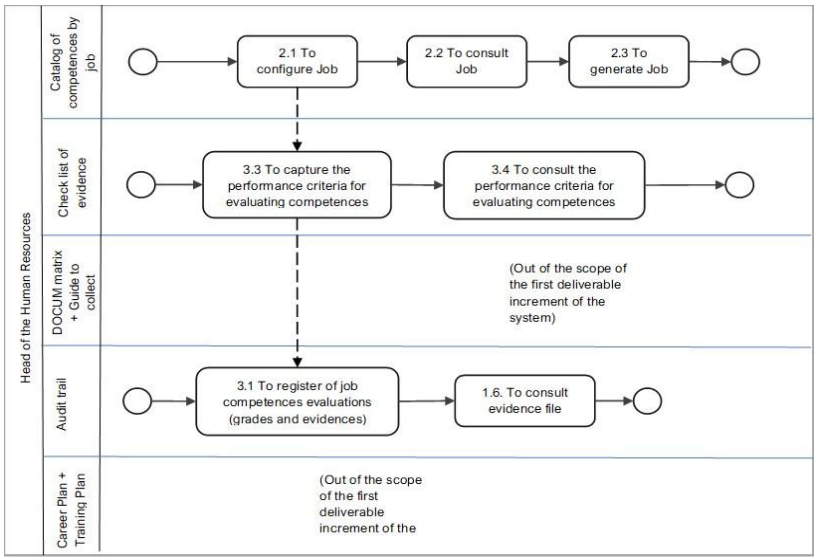


Fig. 10. CIM-6 Business Process Model (BPM) part c: A detailed business process diagram (workflows of the actor "Head of Human Resources")

6.2 Platform Independent Model for the example

For the product PIM-1 Functional Specification, one example, for a use case (see Table 8). It should be noted that each use case requires a Table as part of the product PIM-1. For PIM-1, this example uses the use case description format used in an example of RUP[6] in [33].

Table 8. Use case R.1.1 Define the competence area

Use case	R.1.1 Define competence area
Actor	Head of Human Resources
Purpose	Register and keep the competence areas updated in the competence catalogue.
Summary	The use case is initiated by the user who captures the data that identifies and describes the area of competence. The system records the data captured by the user. The user can modify or delete the data of an area of competence while maintaining the consistency and integrity of the system.
Exceptions	The system will indicate the user that due to the integrity and consistency of the data; an area of competence cannot be eliminated. In the same way, the system will validate the values of the area of competence according to the data rules defined below.

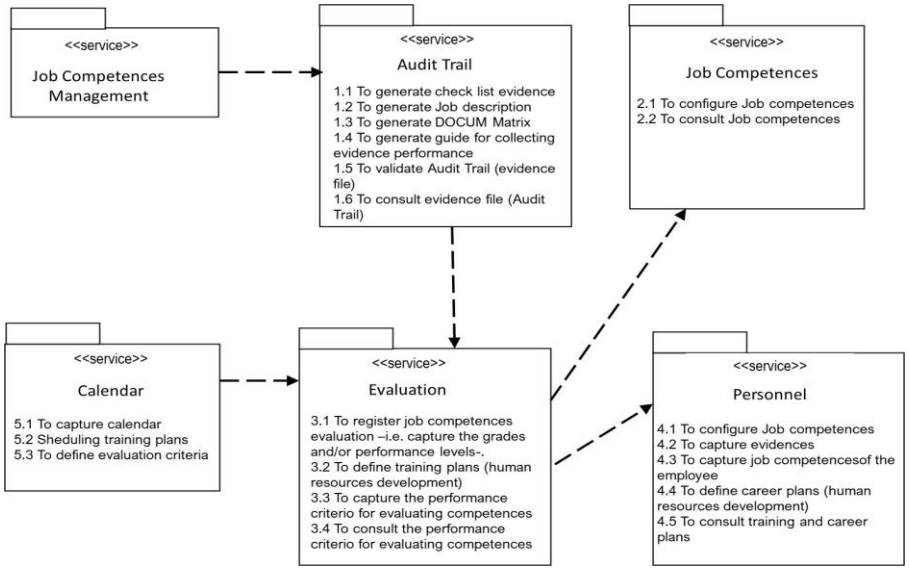


Fig. 11. PIM-2 System-Solution Architecture part a: Service Model diagram (regarding provider and composition of services)

6 RUP – Rational Unified Process

Table 9. PIM-2 System-Solution Architecture part b: JRT – Joint Realisation Table of use cases (regarding the choreograghy) for use case R.1.1 Define competence area

The use case R.1.1 Define competence area		
Action of the Actor	**White Box**	**Service and operation to use**
This use case starts when the Head of Human Resources selects the options of Competences / Define competence area	The system will display a text box and a search button. If the user knows the key competence area it is looking for: they can capture in the text box the code of the corresponding competence. Otherwise, you can search for a specific competence area by pressing the search button.	Job Competences service, operation 2.2 To consult Job competences.
When capturing the area key	The system will automatically search for the required data and display it. If the system does not find the password, the system will assume that it is desired to register a new password, at this point the user may cancel the registration operation or continue with the registration of the new area.	Job Competences service, operation 2.1 To configure Job competences.
The user presses the search button	The system will show a dialog that enables to do a contextual searching by the key competence, description or any other significative field for such a competence area. After locating a specific area of competence, the system will display the data of the area in question. At this point the user can indicate the system if he/she wants to modify the data of the area or if he/she wants to delete them.	Job Competences service, operation 2.2 To consult Job competences.

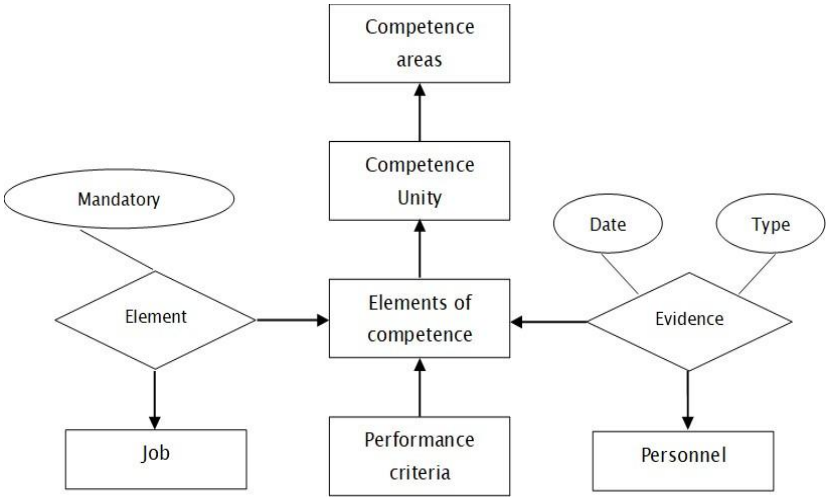


Fig. 12. PIM-2 System-Solution Architecture part c: Entity-Relationship diagram for the first stage or increment of development of the system

The product PIM-2 System-Solution Architecture includes an initial architectural design of the service provider and the service consumer (initial definition of composition, orchestration and choreography), and an initial design of the database (see the Figure 11 and the Table 9). The initial version of the architecture design has the need of refine, correct, and complete the design in order to achieve a better version on subsequent development phases. More specifically, Figure 11 presents the Service Model diagram (regarding service provider and choreography), Table 9 present the Join Realization Tables, each one for one of three use cases (regarding consumer and orchestration), and Figure 12 presents the Entity-Relation diagram (regarding the database). Regarding the notations used for the PIM-2 example, part a is presented as a stereotyped diagram for services as in [22]. Part b is a Joint Realisation Table taken from RUP in [33] and applied as in [22]. Part c is an Entity Relationship Diagram [35].

6.3 Platform Specific Model for the example

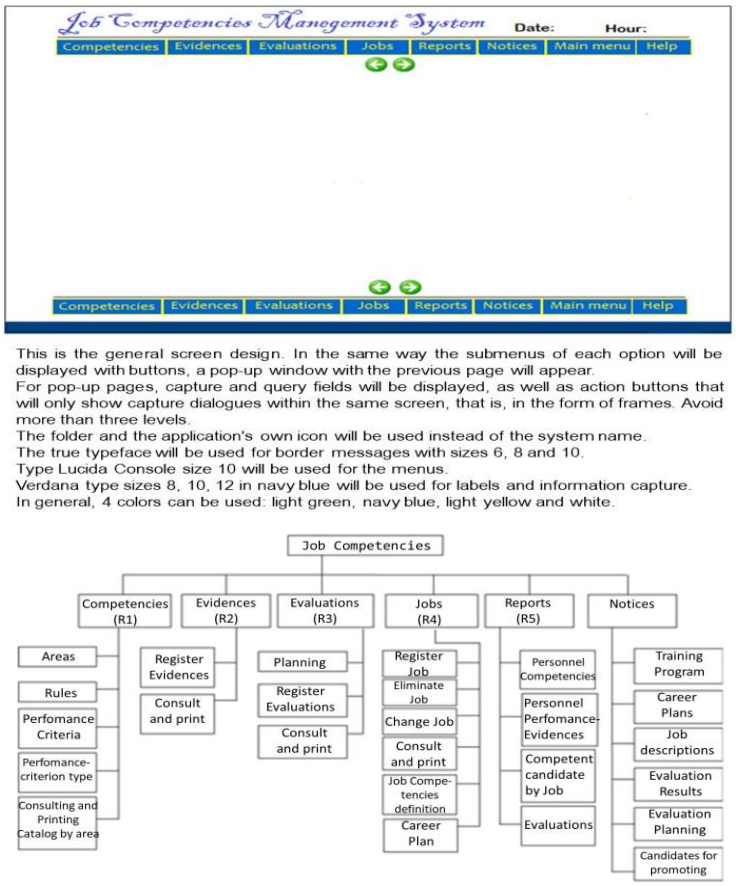


Fig. 13. PSM-1 Prototyping, part a: interface design of the main screen and part b: navigational design of the system

Finally, Figures 13 and 14 present the product PSM-1 Prototyping that involves the construction of prototypes for designing interfaces and prove the possibility of connection for the initial architecture design of provider and consumer. This PSM-1 also proposes technology options [29] for analysing the feasibility of constructing the system in any of them. The illustrative example presents only the preliminary design of the main screen of the system and a proposal of navigation between the options of the system. More specifically, Figure 13 presents the initial interface design of the main screen of the system. Here we use paper prototypes as in [22] to illustrate the PSM-1 Prototyping.

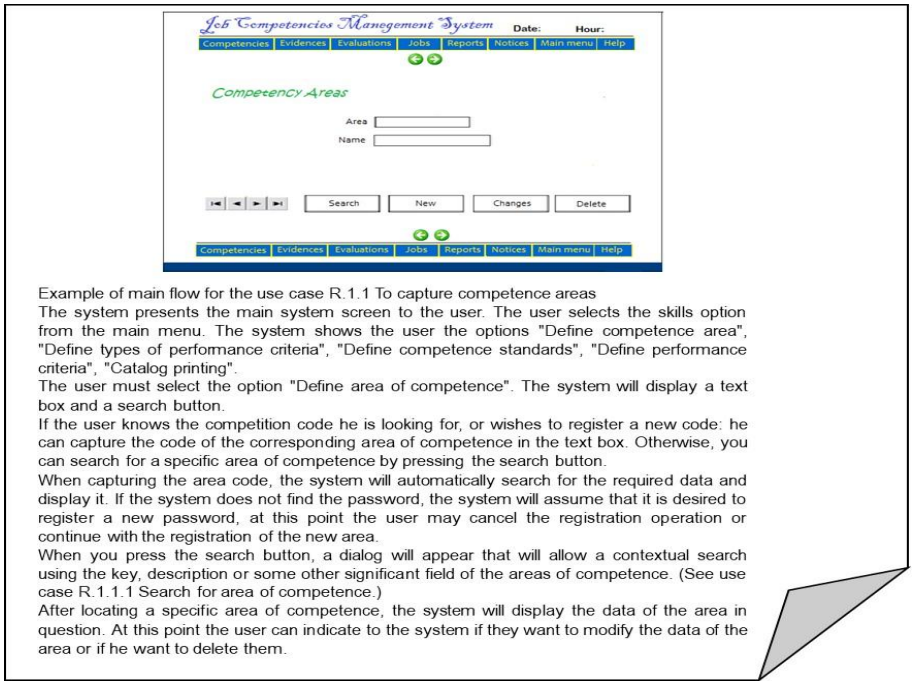


Fig. 14. PSM-1 Prototyping, part c: interface design for register, eliminate or change a competence area as a standard interface for the system

7. Evaluation of the proposal

The reviewed proposals present an evolution on the knowledge of the requirements. The first type of proposals, the RA proposals, regard the activity of “requirements analysis”. In contrast, in the RE proposals the requirements are managed as a process called “requirements engineering”. Then, the AD proposals add new characteristics whereby the relationship between “requirements analysis” and “system design” is described and eases the transformation of the requirements analysis into the system analysis. Consequently, the SOSE proposals adapt the requirements activities to construct

Service-Oriented Computing Applications. Finally, the MDA proposals change the process, i.e., it proposes a different way to construct a system focusing on the models to be constructed during the requirements process.

We claim that a solid requirements analysis processes for SOCA should provide support for RE, MDA, and SOSE. The reason of this is described as follows. In the early days, RA was incorporated and diluted into RE. Nowadays, AD is also being incorporated into RE. In this context, AD, in an analogous form, highlights the importance of the architecture design closely linked to the requirements specification. In addition, the research areas, namely RA, AD and MDA, highlight their respective link between (1) the requirements process and the system design, (2) the requirements and the architectural design; and (3) the products categorised in models with their approach to designing and transforming. In turn, RE provides the model outlining the activities to be performed during the requirements process, some products to be built during each activity, and the process management activities. However, RE, does not describe how to construct or design the products. MDA addresses this by defining a way to construct various evolving models throughout the development process, encompassing the requirements process. Finally, with the current trend toward web-based systems, constructed via services or microservices, it is required to focus on the construction of SOSE products.

We evaluate our approach by comparing it to the reviewed proposals. Table 10 shows the type of support provided by each proposal. The third, fourth, and fifth columns of Table 10 indicate whether each proposal supports RE, MDA, and SOSE, respectively, through a check mark for positive support or a cross for no support.

Table 10. Comparative analysis of the support provided by the proposals

Proposal & its name	Type of proposal	RE support	MDA support	SOSE support
P1 – Leite, 1988 [12] -"Viewpoint Resolution in Requirements Elicitation"	RA	√	×	×
P2 – Boehm, 2004 [27] - (MBASE)	RA	√	×	×
P3 – Getchell, 2002 [29] - "MSF Process Model v. 3.1"	RE	√	×	×
P4 – Hofmeister, 2005 [8] - "Global Analysis"	AD	√	×	×
P5 – Hofmeister, 2007 [9] - "General Model of Software Architecture Design Process"	AD	√	×	×
P6 – Péraire, 2007 [21] - (RUP-SOA)	SOSE & RE	√	×	√
P7 – Habe, 2013 [7] - "The Missing Link - between Requirements and Design"	AD	√	×	×
P8 – Bourque, 2014 [4] - (general model proposed in SWEBOK V3.0)	RE	√	×	×
P9 – Asadi, 2008 [3] - "An MDA-based System Development Lifecycle"	MDA	√	√	×
P10 – Rodríguez-Martínez, 2019 [22] - (SOCA-DSEM)	MDA & SOSE	×	√	√
SOCA-rap - integrated MDA-based requirements process - MDA & SOSE RE process	MDA - SOSE - RE	√	√	√

We analyse below the elements that are supported by the proposals for each type of support i.e. RE, MDA and SOSE.

7.1 Support for RE

Currently, RE refers to proposals that encompass both RA and AD. Initially, RA proposals focused primarily on the business point of view. Subsequently, RE proposals shifted focus towards process-oriented approaches. Currently, RE proposals tend to emphasise the system point of view, specifically regarding architectural design. Furthermore, ***requirements modelling*** and ***early system design*** remain relevant in the literature. P1, P5, and P6 approach modelling and early design from the ***software architecture research area*** [12], [8], [7], which has significantly expanded in the last decade. As a result, innovative proposals are needed to deal with the intersection of requirements and architecture. In essence, RA deals with the modelling of business requirements, while RE evolves describing the process derived from RA. The emerging AD then transforms business requirements into system design with modelling playing a vital role. And, RE emphasises the process but also considers modelling from both RA and AD.

All proposals, except P10, provide support for RE since they support the requirements analysis in either forms: (1) as requirements-process definition, or (2) as requirements-modeling strategy (e.g. CIM modeling with MDA). The first form mainly defines the process whereas the second one mainly establishes the products. Given that P1 and P2 are presented as definitions of a requirement-processes, they naturally support RE. Also, P3 to P8 inherently support RE as they are derived from RE and its corresponding AD. Regarding R9, it offers the two forms, as a result, it supports RE too. Lastly, P10 is presented as requirements-modeling strategy, therefore P10 does not support RE. Regarding *SOCA-rap*, it provides support for RE by defining the activities and their description, and by establishing how these activities iterate throughout the process. More specifically, SOCA-rap includes the activities of *Requirements Elicitation*, *Requirements Analysis*, *Modelling/Specification*, and *Requirements Validation* (see Table 7). The way to iterate in SOCA-rap consists of two dimensions related to the whole SDLC (see Figure 1). The SDLC is divided into four phases, namely Requirements, Design, Construction and Operation, which make up the horizontal dimension of the process. Each phase comprises specific activities. For instance, the SOCA-rap activities pertain to the Requirements phase. Meanwhile, the vertical dimension deals with products and entails building MDA models during specific phases. Table 6 shows each model that encompasses various SOCA-rap products.

7.2 Support for MDA

Support for MDA refers to proposals that focus on Model-Driven Development. Such proposals focus on defining products to be built that are grouped in the CIM, PIM, PSM and Executable Models, as well as the transformation between these models. Nevertheless, current approaches also strive to detail the development process for each

stage of the software development process. The CIM model is closely interconnected with the requirements stage. Also, the PIM, PSM and Executable model are closely related to the architectural design stage, the detailed design stage and the implementation stage, respectively.

Only P9, P10, and SOCA-rap provide support for MDA. P9 provides an SDLC description for developing MDA models. It is advised to construct each model in a singular phase, with corresponding activities and products allocated to each stage, respectively. Regarding RE, P9 propose to construct two products, namely "*CIM model*" and the "*Requirements model*," but does not provide details. P10 proposes stages, activities and products for a comprehensive SDLC process for developing service-based systems. The proposal does not encompass the requirements analysis process. Instead, it puts forward the CIM model for service-based systems. On the other hand, SOCA-rap proposes the requirements analysis process by outlining the activities and their products. These activities are grouped into the Requirements phase. Additionally, SOCA-rap categorises the products into the CIM, PIM and PSM. As the requirements process has an obvious connection to the CIM model, the majority of the products fall into this category. Although, there is a typical association of requirements with the CIM model, some products of the Requirements phase are not Computing Independent. As a consequence, these products should be included into the PIM or PSM model. SOCA-rap considers such grouping emphasizing the early design and the linkage of requirements and architecture design.

7.3 Support for SOSE

SOSE support pertains to processes that take into account service-oriented development. P10 proposes the products of the CIM model that are directly related to RE. Therefore, only P6, P10, and SOCA-rap provide support for SOSE. P6 introduces a RUP version for service-based systems, which describes each phase of the software development process. P10, on the other hand, focuses on the products of requirements and modelling questions but does not address the requirements phase process. In addition, SOCA-rap includes the necessary elements, and specially, the products mandated by SOSE requirements. Furthermore, P6, P10, and SOCA-rap address requirements, either through the definition of the requirements process definition and/or definition of requirement products.

SOCA-rap includes and describes most of the products that are present in RE and in SOSE proposals (see Table 6). The products are presented as part of the different MDA models; specifically, those that have products related to the requirements process and those models that are constructed during the requirements process. Such models are CIM, PIM, and PSM.

Overall, we can see that the proposals that provide better support are P6, P10, and SOCA-rap. However, the two former miss to provide support to MDA and RE, respectively, whereas SOCA-rap provides full support to all elements of Table 10.

8. Threats to Validity

In the case of threats to internal validity, the results of the accuracy evaluation could be affected by a bias in the selection of the ten works that were used to compare our work. We reduced this threat by following disciplined principles to search the reviewed works. These principles are: 1) the work explicitly or implicitly define artifacts to be developed during the requirements process, 2) the work define the activities of the requirements process, design process and/or CIM modeling, 3) the work defines its products in terms of viewpoints for the various stakeholders in the business or development process. These search principles privilege the manual analysis of the works. The manual analysis allowed us to identify and discriminate themes or topics, and refine searches. Furthermore, it allowed us to consider references whose documents are difficult to find even with the complete reference, especially, old references of doctoral theses or technical reports.

Threats to external validity are related to the extent the fulfillment of the evaluated aspects reflect the most important characteristics of a RA process for SOCA. We improved external validity by using an example case, which showed the feasibility of our approach. The example case shows how the requirements process builds the CIM, and also shows how to build the first stages of the PIM and the PSM. The artifacts constructed exemplify the requirements products of each MDA model according with the viewpoints proposed by SOCA-rap. The example case, also illustrates how, following the transformation lines in Figure 3, it is possible to transform some products into others. And how these transformations, progressively link the domain model (CIM) to the system models (PIM and PSM). Where the PIM is still independent of how it will be implemented, and the PSM is already a design to be implemented.

It is also shown that the definition of the artifacts is applicable to SOCA. The case example illustrates how to follow the SOCA-rap process, its transformation lines, and the construction of its products. This illustrates how each model in the transformation line is aligned first to business issues, then to system issues, and then to the implementation. Finally, the example case illustrates the execution of the first iteration of the development process for the requirements. Finally, the construction of the product CIM-5 provides a rough idea of the number of iterations/increments (i.e. stages of development) that are required to complete the system.

9. Conclusion

We methodologically studied ten approaches, which come from related complementary areas, namely RA, RE, AD, SOSE, and MDA. Our study found some characteristics that a good requirements analysis process for SOCA should possess. These characteristics are the following: (1) the process has an early design on the software development, (2) it considers the link between requirements analysis and architectural design, and (3) it describes the products and the models to be constructed. We then employed and applied a comparative framework to the reviewed methodologies in order to identify the elements that a SOCA requirements analysis process needs to meet the above-mentioned characteristics. Based on this, we developed SOCA-rap, which defines its elements in terms of phases, activities, products, and roles/viewpoints. SOCA-rap

covers all characteristics, except the means to carry out model transformation among the models. Hence, SOCA-rap defines the products of the requirements process. Each product is defined as part of an MDA model. In addition, each model is documented in specific documents of requirements. Concretely, the CIM is documented in the System Definition Document and the System Requirements Document, while the PIM and PSM are documented in the Software Requirements Specification (SRS) and the Initial Design Document.

In section 7 we evaluated the support that SOCA-rap and the reviewed methodologies provide to RE, SOSE, and MDA. We can see that SOCA-rap offers more support than the other approaches.

Finally, as future work, we will provide a model-to-model guide for carrying out model transformation of the models defined in the SOCA-rap.

References

1. Aguilar, J.A., Garrigós, I., Mazón, J.: Requirements Engineering in the Development Process of Web Systems: A Systematic Literature Review, Acta Polytechnica Hungarica, 13(3). (2016)
2. Amna, N., Anam, A., Farooque, A.: Model Driven Architecture – Issues, Challenges and Future Directions, Journal of Software, 11(9), 924-933. (2016)
3. Asadi, M., Ravakhah, M., Ramsin, R.: An MDA-based System Development Lifecycle, Proceedings of Second Asia International Conference on Modeling & Simulation, IEEE Computer Society, 836-842. (2008)
4. Bourque, P., Fairlley, R.E.: SWEBOK V3.0 Guide to the Software Engineering Body of Knowledge, IEEE Computer Society, 1-335. (2014)
5. Brown, A.W.: Model driven architecture: Principles and practice, Springer Verlag, Expert's voice. Software and Systems Modeling, 3(1) 314-327, Digital Object Identifier (DOI) 10.1007/s10270-004-0061-2. (2004)
6. Cantor, M.: Rational Unified Process for Systems Engineering Part 1, 2, 3, IBM Rational Software. (2003)
7. Habe, A., Michielsen, C.: The Missing Link -between Requirements and Design, Proceedings of the Posters Workshop at CSD&M. (2013)
8. Hofmeister, C., Nord, R.L., Soni, D.: Global Analysis: moving from software requirements specification to structural views of the software architecture, Iee Proc.-Softw., 152(4), 187-197. IEE Proceedings online no. 20045052, doi: 10.1049/ip-sen:20045052. (2005)
9. Hofmeister, C., Kruchten, P., Nord, R.L., Obbink, H., Ran, A., America, P.: A general model of software architecture design derived from five industrial approaches. Elsevier, The Journal of Systems and Software 80, 106-126. (2007)
10. Khan, S., Dulloo, A.B., Verma, M.: Systematic Review of Requirements Elicitation Techniques, International Journal of Information and Computation Technology, 4(2), 133-138. (2014)
11. Leite, J.: A Survey on Requirements Analysis, Technical Report, University of California, Department of Information and Computer Science. (1987)
12. Leite, J.: Viewpoint Resolution in Requirements Elicitation. PHD thesis, Department of Computer Science, University of California, Irvine. (1988)
13. Leite, J.: Viewpoint analysis: a case study, ACM SIGSOFT Software Engineering Notes, 14(3), 111-119. (1989)
14. Miller, J., Mukerji, J.: MDA Guide Version 1.0.1. OMG. (2003)
15. Neil, C.J., Laplante, P.A.: Requirements Engineering: The State of the Practice, IEEE Software, IEEE Computer Society, 20(1), 40-45. (2003)

18. Nuseibeh, B., Easterbrook, S.: Requirements Engineering: A Roadmap, ICSE. (2000)
19. Oktaba, H., Ibargüengoitia, G.: Software Process Modeled with Objects: Static View, Computación y Sistemas 1(4), 228-238. CIC-IPN ISSN 1405-5546. (1998)
20. Parviainen, P., Takalo, J., Teppola, S., Tihinen, M.: Model-Driven Development – Processes and practices, VTT Working papers 114. (2009)
21. Péraire, C., Edwards, M., Fernandes, A., Mancin, E., Carrol, K.: The IBM Rational Unified Process for System z, IBM Rational software, Redbooks. (2007)
22. Rodriguez-Martinez, L.C., Duran-Limon, H.A., Mora, M., Alvarez-Rodriguez, F.: SOCA-DSEM: a Well-Structured SOCA Development Systems Engineering Methodology, Computer Science and Information Systems (COMSIS), 16(1), 19-44. (2019)
23. Singh, Y., Sood, M.: Model Driven Architecture: A Perspective, IEEE International Advance Computing Conference (IACC 2009), 1644-1652. (2009)
24. Zachman, J.A.: The framework for Enterprise Architecture: Backgroud, decription and utility, https://www.zachman.com (2016)
25. Gu, Q., Lago, P.: A stakeholder-driven service life cycle model for SOA, In Proceedings of 2nd international workshop on Service oriented software engineering: in conjunction with the 6th ESEC/FSE joint meeting (IW-SOSWE'07) 1-7. Dubrovnik, Croatia: ACM. (2007)
26. Gu, Q., Lago, P.: Guiding the selection of service-oriented software engineering methodologies, SpringerLink, Service Oriented Computing and Applications (SOCA), 5(4), 203-223. (2011)
27. Boehm, B., Klappholz, D., Colbert, E., Puri, P., Jain, A., Bhuta, J., Kitapci, H.: Guidelines for Model-Based (System) Architecting and Software Engineering (MBASE), Center for Software Engineering, University of Southern California. (2004)
28. Boehm, B., Kitapci, H.: The WinWin Approach: Using a Requirements Negotiation Tool for Rationale Capture and Use. In: Dutoit A.H., McCall R., Mistrík I., Paech B. (eds) Rationale Management in Software Engineering. Springer, Berlin, Heidelberg, 173-190. (2006)
29. Getchell, S., Hargrave, L., Haynes, P., Lubrecht, M., Pervez, K., et al.; MSF Process Model v. 3.1. Microsoft Corporation, Microsoft Solutions Framework, White Paper. (2002)
30. Richards, M., Neal, F.: Fundamentals of Software Architecture: An Engineering Approach. Ed. O'Reilly. (2020)
31. Stahl, T., Völter, M., Bettin, J., Haase, A., Helsen, S.: Model-Driven Software Development. John Wiley & Sons. (2006)
32. Alter, S.: The Work System Method: Connecting People, Processes, and IT for Business Results, Work System Press. (2006)
33. Rational Software Corporation.: Rational Unified Process for Systems Engineering SE1.1, Rational Software Corporation, White paper. (2002)
34. Larman, C.: UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development, Third Edition, Addison Wesley Professional. (2004)
35. Silberschatz, A., Korth, H.F., Sudarshan, S.: Database System Concepts, Seventh Edition., Mc Graw-Hill. (2019)
36. Pressman, R.S.: Software Engineering: A Practitioner's Approach, Fifth Edition, McGraw-Hill series in computer science. (2001)
37. OMG.: Business Process Model and Notation (BPMN), Version 2.0.2, OMG. (2013)

Laura C. Rodriguez-Martinez is a full Professor at the Systems and Computing Department, Institute of Technology of Aguascalientes, Mexico. She holds a PhD in Computer Science at Universidad Autonomous University of Aguascalientes, Mexico in 2009. Her research interests include Software Systems Development Processes, Service-Oriented Software Engineering and Graphical-User Interfaces Development Processes.

Hector A. Duran-Limon Eng.D., has published more than 25 research papers in peer-reviewed journals listed in JCRs and more than 40 papers in international top conferences and research books. He obtained an IBM Faculty award in 2008. He holds an M.Sc. in Computer Science (1994) from the National Autonomous University of Mexico (UNAM), and a PhD in Computer Science (2001) from Lancaster University, England. Prof. Duran-Limon was a research assistant at Lancaster University (2002-2003), a Professor at the Tec Monterrey (2004-2005), and has been a full-time Professor, since 2006, at the University of Guadalajara, Mexico. Prof. Duran-Limon has directed the thesis of eight PhD students. His current research interests are software architectures, software product lines, and HPC in the cloud. Prof. Duran-Limon is also a Mexican National Researcher at Level I. Overall, Prof. Duran-Limon has more than 20 years of experience doing research, and teaching both undergraduate and graduate courses.

Francisco Alvarez-Rodríguez is a Professor of Software Engineering. He holds a BA. in Informatics (1994) and a MA. (1997) from the Autonomous University of Aguascalientes and a EdD degree from the Education Institute of Tamaulipas, México and he is PhD from the National Autonomous University of Mexico. He has published research papers in several international conferences in the topics of software engineering and e-learning process. His research interests are software engineering lifecycles for small and medium sized enterprises and software engineering process for e-learning. He is currently president of the National Council for Accreditation of programs and Computing , A.C. (CONAIC).

Ricardo Mendoza-González is a full-time professor at the Tecnológico Nacional de México/ IT Aguascalientes. Member of the National Researchers System (Level 1) from the Secretariat of Science, Humanities, Technology and Innovation, SECIHTI (in Spanish), Mexico. His current research interests include several topics on (but not limited to): Human-Computer Interaction, User Research, User-Centered Design, Usability, Accessibility and Equity in technology, Design Thinking, User Interfaces Design, Innovation processes, Educational Technology, Open Educational Resources, Software Engineering and Artificial Intelligence.

Received: July 01, 2024; Accepted: February 16, 2025.

Computer Science and Information Systems 22(3):815–837 https://doi.org/10.2298/CSIS240729017M

PI2M-ITGov – Panel of Indicators for Monitoring and Maintaining the Information Technology Governance: Method and Artefacts

Altino J. Mentzingen Moraes* and Álvaro Rocha

Advance, ISEG, University of Lisbon
Rua Miguel Lupi, nº 20 (Gab. 101)
1249-078 Lisbon, Portugal
altino.moraes{@gmail.com, @gestao.gov.br}
amrrocha@gmail.com

Abstract. In today's highly competitive corporate world, effective management of Information Technology (IT) resources and the consequent need for structured management and governance are essential. Therefore, prior planning of goals, definition of effective actions, and monitoring of results (through Indicators) is the way to apply effective control. This article presents the method named as PI2M-ITGov - Panel of Indicators for Monitoring and Maintaining Information Technology Governance, which covers 12 identified IT Areas and consists of 12 key monitoring indicators (KMIs) and their 36 sub-KMIs (3 sub-KMIs for each 12 identified IT Areas). This method is the result of many years of IT Governance Models implementation expertise of the Authors besides theories presented by both in several Technical Congresses participations. The artefacts created are available at the provided links. The simulation (through a case study) demonstrated a high level of acceptance of the tools as a practical IT Governance alternative.

Keywords: IT Management; IT Governance; Areas of Management of IT; IT Strategic Alignment; MM - Maturity Models; KMI - Key Monitoring Indicators; Case Study; DSR - Design Science Research.

1. Introduction

Organizations can be classified into two distinct domains: 1) Those that utilize information technology (IT) resources for their operations; and 2) Those that serve as providers of IT resources. Figure 1 visually illustrates this division, emphasizing the differentiation between these two domains.

In both cases, even the organizations belonging to the second category (IT resources providers) have an internal administrative area that functions similarly to companies in the first category (IT resources users). This internal administrative area also needs to effectively manage its performance and actions to ensure the success of the overall business.

Consequently, both domains – without exception – make significant investments in IT, which are expected to yield tangible results for the business. Whether operating in the industrial, commercial, or financial sectors, organizations aim to maintain a competitive position in the globalized business environment.

* Corresponding author

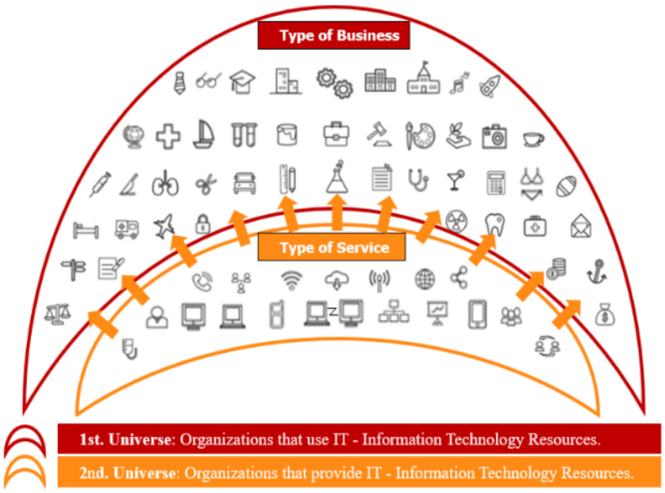


Fig. 1. Enterprise domains

To ensure effective monitoring of IT initiatives and their alignment with the organization's strategic planning, a set of guidelines known as IT governance is implemented.

Peter Weill [41] is one of the prominent authors in the field of IT governance, emphasizing the importance of IT being perceived as a sustainable area that generates results (revenues), rather than merely incurring costs (expenses). According to Weill, IT should strongly focus on 'aligning its initiatives with the business'.

In the words of Rockart [32]:

> *In sum, the load of IT on organizations is heavier than ever before and the management of it is more complex.*

In light of this reality, as highlighted by one of the authors [24] in this research, significant challenges need to be addressed to effectively implement IT governance and strive for high-quality indices. The doctoral thesis presents Figure 2, referred to as the 'puzzle' of the governance process, illustrating these challenges. Overcoming these challenges is essential to successfully implement IT governance and achieve desirable quality metrics.

Consequently, it becomes evident that monitoring IT governance activities requires an initial and meticulous planning phase (structuring), followed by effective management (administration), and rigorous governance (control) processes.

To accomplish this objective, a practical approach is to apply the plan, do, check, act (PDCA) cycle, initially formulated by the American statistician William Edwards Deming. The PDCA cycle is advocated by the author Ruks Rundle [33] and is depicted in Figure 3, adapted here to provide a more operational interpretation.

The problem at hand revolves around how management can effectively apply governance processes, particularly those recommended by IT governance techniques. It is crucial to adopt a proactive approach to administration control, rather than a reactive one, in order to anticipate and prevent misalignments between the actions of the IT de-

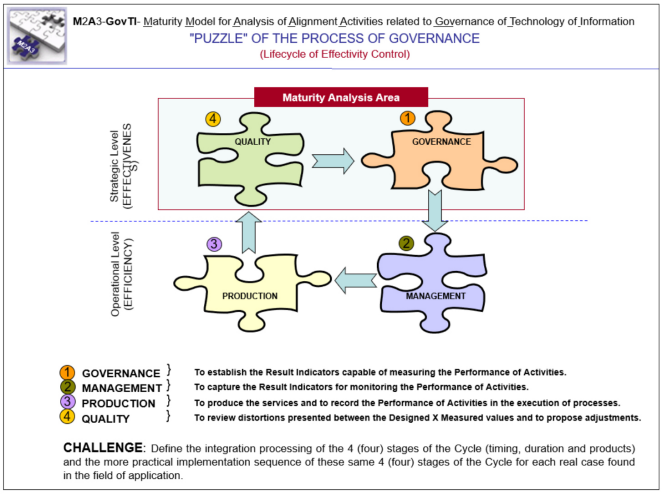


Fig. 2. Process of governance 'puzzle' [24]

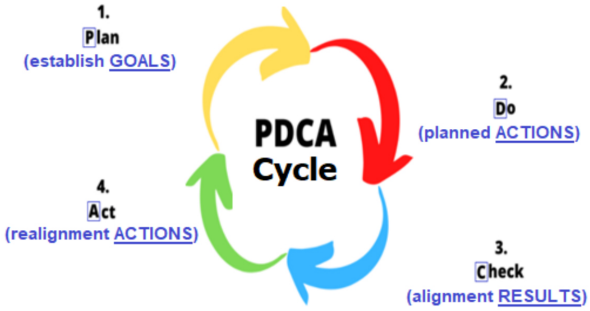


Fig. 3. Plan, do, check, act (PDCA) cycle: adapted by authors from Rundle [33]

partment and the initial planning. Failure to address these misalignments can result in non-compliance with the organization's expected outcomes.

To address this challenge, the panel of indicators for monitoring and maintaining information technology governance (PI2M-ITGov) method, developed through research and presented in this article, aims to fulfill the management's need in this field. It provides a panel of indicators that can measure the degree of alignment between the actions implemented by the IT department and the organization's strategic planning. This enables decision-makers to identify and rectify any distortions, thereby guiding the organization back to the correct course of action.

It is important to inform that this article is the result of a deep revision of one first and initial proposal presented by Moraes & Rocha [25] in a European IT Congress held in the City of Aveiro/Portugal (CISTI'23) – which revision besides other aspects – involved, as the main purpose, the build of artefacts.

Those artefacts were created (such as a spreadsheet, a guideline, forms and other tools), in order to make, this first and initial proposal became from theory to practice in a real operational environment. Their links are provided and mentioned in the context of this paper.

Additionally, it is also important to note that the PI2M-ITGov method ensures the maintenance of confidentiality and privacy of personal data. The method does not handle information that possesses sensitive content or characteristics, thereby safeguarding the privacy of individuals involved.

2. Applied Methodology

This research project is guided by the design science research (DSR) methodology and its principles.

In a broader sense, scientific research involves the practical application of objective procedures by researchers to develop experiments and generate new knowledge that integrates with existing knowledge, as explained by Fontelles [11].

Within the framework of DSR, the development of artefacts plays a crucial role in producing scientific knowledge from an epistemological perspective, as described in Simon's work 'The Sciences of the Artificial' [35].

According to Peffers et al. [27], any designed entity intended to achieve a goal can be considered an artifact. However, the creation of a well-designed artifact and its investigation in a specific context are key elements in knowledge production within scientific research, as emphasized by Dresch, Lacerda & Antunes Jr. [8].

This paper adheres to the process flow of DSR as recommended by Wieringa [42], with graphical adaptations made by the authors, as depicted in Figure 4.

The DSR methodology consists of several cycles, including the design cycle (steps 1, 2, and 3) for conceiving and constructing the solution, and the engineering cycle (steps 3 and 4) for adapting the solution to real-world applicability.

In DSR, one of the objectives of research is to contribute to the knowledge base of the research area. Therefore, it is crucial for researchers to engage with existing knowledge bases to ensure their work provides an original contribution to scientific knowledge, rather than solely focusing on technological advancements [15].

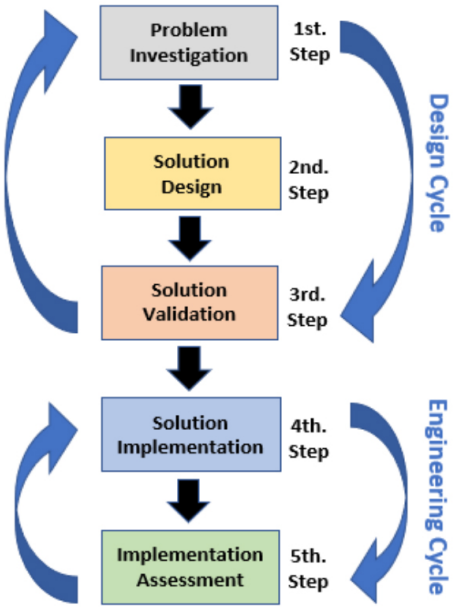


Fig. 4. The design science research flow: adapted by authors from Wieringa [42]

Following the guidelines of step 1 (problem investigation) in DSR, this research aimed to determine if there was a sufficient degree of relevance to justify its development. The investigation sought to identify whether other studies had addressed the problem that needed to be resolved and considered important. The results of this investigation are presented in the subsequent section titled 'problem investigation'.

Within the context of DSR, artefacts can be classified into four types: models (abstractions and representations), methods (algorithms and practices), instantiations (implementations and prototypes), and constructs (vocabularies and symbols).

This research focused on creating artefacts in alignment with the principles of DSR. To ensure their effectiveness in practical application, the research underwent three iterations of the engineering cycle (steps 3 and 4). These iterations aimed to refine the artefacts and adapt them to the realities of corporations.

This clarifies the rationale behind the versioning of the artefacts, which are made available in the provided link in Section 3 ('Solution Validation'). The version indicated as V.01c signifies the third revision of the artifact.

3. Problem Investigation

As mentioned in Section 1, this research aims to address the need for a tool within IT governance that can manage a dashboard to guide the alignment of IT initiatives with corporate strategic planning.

Following the guidelines of the DSR methodology, the research commenced with step 1, 'problem investigation'. This step aimed to determine whether conducting this research and allocating effort to this task would contribute to advancing scientific knowledge.

To assess the existing literature related to the research objectives, a bibliographic search was conducted. Nine databases were searched, namely ACM, IEEE, MIT, Research Gate, Scholar Google, SciELO, Science Direct, Scopus, and Web of Science (WoS). The search keywords used were 'IT Governance NOT Social NOT Health NOT Education AND Performance Indicator AND Spreadsheet AND Guideline'. The exclusion criteria of 'NOT Social NOT Health NOT Education' were applied to filter out articles not directly linked to the business domain, such as banking/financial, industry, and services.

Of course, the papers those were focused on these 3 areas (Social, Health and Education) and in their business tools, what means no in its final targets as to support Social, Health or Education problems in its main themes, were also considered because, in this interpretation, these can be sort as integrating the areas of banking/financial, industry, and services as well.

Additionally, the search was limited to the last decade (2012–2022) to ensure relevance and up-to-date information.

The initial search in the ACM database yielded only seven occurrences, which did not align with the research focus. Subsequent searches in IEEE, MIT, Research Gate, Scholar Google, and SciELO databases resulted in zero occurrences, while Science Direct yielded only two occurrences. In response to these limited results, two additional databases, Scopus and Web of Science, were searched, but no relevant articles were found.

Detailed information regarding the search results can be found in the '1. Problem Investigation/Investigação do Problema' folder accessible via the following link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

Based on the analysis of these search results, which demonstrated the lack of literature on this specific topic, it became evident that the research should proceed. It was determined that there was a significant gap that could be filled by concluding this research in the planned format. The intention is to generate practical material that can effectively assist IT managers in their work and activities.

4. Solution Design

Following the principles of the DSR methodology, this research project progressed to step 2, 'solution design'. The objective of this step was to conceptualize and construct the architecture required to address the identified research question.

4.1. The 12 IT Monitoring Areas

As a way of delimiting the scope of this study – among the various existing possibilities – the Theory of Number 12 as described by Tesla [1] [39] which highlights the significance of this 'magic number' in various elements of the universe, served as the inspiration for the establishment of a quantitative framework for IT work processes.

[1] Nikola Tesla was a Serbian-Croatian scientist who was known for his important discoveries in the field of electricity. His work was fundamental in improving the transmission of electrical energy.

Since Tesla was one of the most recognized scientists by the Scientific Community in studies related to electricity, and was even a direct competitor of Thomas Edison in his discoveries, the line of work applied by this research – in the composition of the quantity of IT work processes to be studied – followed his "magic number" equal to 12 (instead of being applied other value within another focus of reasoning), also because, Tesla related his "magic number" to many elements of the universe and not only to elements involved with the theme of the electricity.

Following this line of reasoning, IT work processes were organized into 12 IT management areas, which became the focal point for the creation of indicators.

To determine the 12 IT management areas, in addition to delimiting the scope of this study within 12 areas according to Tesla's Theory (as the initial idea already mentioned above), the empirical experience of the Authors (who have each worked in the IT field for almost 5 decades) was considered, as well as a criterion based on the identification of IT work processes that had published and recognized maturity level assessment models was applied.

These 12 areas, which had corresponding maturity level assessment models, were subsequently renamed as IT monitoring areas. For each of these areas, 12 key monitoring indicators (KMIs) were created, with 3 sub-KMIs for each one, resulting in a total of 36 sub-KMIs for the proposed solution (which will be the focus of the following "4.2. The Scope of the 36 Sub-KMIs - Key Monitoring Sub Indicators").

To ensure precision in the analysis, three key monitoring sub-indicators (sub-KMIs) were developed for each of the 12 IT monitoring areas, leading to a total of 36 sub-KMIs within the proposed solution. The data collected to determine the percentages for these 36 sub-KMIs were derived from aligning the results with the highest maturity levels (ranging from 3 to 5) identified in the considered maturity level assessment models.

The research process involved the application of technical and professional knowledge by one of the authors, combined with the principle of empiricism, as described by Locke [22]. This approach allowed for an analysis of the data to determine whether the maturity level assessment models under investigation could be aligned with an IT management area, which in this case is referred to as IT monitoring areas.

Even when the framework which the maturity level assessment models was based could have more than just one IT area focus it was considered, to make possible the application of the idea behind this proposed study and its resulted model, its main focus (e.g.: CobiT® mentioned in the sequence of this paper).

Despite the availability of 26 maturity level assessment models, the aim was to maintain the framework within the limit of 12 IT monitoring areas. Detailed descriptions of these models can be found in the '2. Solution Design/Projeto da Solução' folder in the provided link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

These findings led to the determination of the 12 IT monitoring areas considered by the PI2M-ITGov. It is important to note that the references listed at the end of this work provide the foundational understanding for readers. However, additional literature was also researched to support the continuation of this study.

Two significant maturity models identified were ITIL® (Information Technology Infrastructure Library) by Axelos [3] and ITSCMM© (Information Technology Service Capability Maturity Model) by Clerc & Niessink [6]. Both of these models pointed towards

a set of assessments that could be categorized under a services management area, which was subsequently renamed as the services monitoring area.

Furthermore, the CMMI© (Capability Maturity Model Integration) of the SEI® [34], the Software Engineering Institute, as referenced by [1], and with strong alignment with the Technical Standard 15.504 ISO/IEC [18], along with the MR MPS br© (Reference Model for Improvement of the Brazilian Software Process [*Modelo de Referência para Melhoria do Processo de Software Brasileiro*]) by SOFTEX [36], were also investigated. After conducting the research, these models were classified as relevant to the development management area, which was then renamed as the development monitoring area.

The research also investigated the CobiT® (Control Objectives for Information and Related Technology Maturity Model) from ISACA [17], referenced by Gartner [12], and found that it specifically addresses the management of IT work processes. Similarly, the TOGAF©(The Open Group Architecture Framework) from the OPEN GROUP was identified as relevant to the same area. Both models were categorized under a business management area, which will be renamed as the business monitoring area.

The exploration of the B-ITa©(Business–IT Alignment Maturity Model) by Tapia, Daneva, Eck & Wieringa [38], EAG©(Enterprise Architecture Governance) by CIOIndex [5], and ITGAP©(IT Governance Assessment Process) by Peterson [28] highlighted the need for a governance management area specifically linked to IT. This area will be renamed as the governance monitoring area in the context of this research.

The research also encompassed maturity models such as MMGP©(Project Management Maturity Model) by Prado [30], OPM3©(Organizational Project Management Maturity Model) from PMI® [29], PMMM©(Project Management Maturity Model) by Kerzner [19], and P3M3©(Portfolio, Program, and Project Management Maturity Model) by Axelos [4]. These models were associated with project management, leading to the identification of a projects management area, which will be renamed as the projects monitoring area.

The investigation revealed a maturity model called OKA©(Organizational Knowledge Assessment) presented by Fonseca [10], which mentioned the use of a software application called SysOKA©(OKA – Organizational Knowledge Assessment System). This model was classified as relevant to the knowledge management area, which will be renamed as the knowledge monitoring area for the purpose of this work.

The maturity models SIMM©(Service Integration Maturity Model) by [2] and SOAMM©(SOA©Maturity Model) by Sonic [37] were evaluated and identified as relevant to the integration management area, which will be renamed as the integration monitoring area.

During the assessment of maturity models, the research identified the requirements management area, which will be renamed as the requirements monitoring area, through the exploration of the RMM©(Requirements Management Maturity Model) by Heumann [14].

Further analysis of maturity models included the BPMM© (Business Process Maturity Model) from OMG [26] and the BPMMM©(Business Process Management Maturity Model) referenced by Tapio Hüffner [16]. These models were associated with the processes management area, which will be renamed as the processes monitoring area.

Additionally, the EFQM©[9], the European Foundation for Quality Management and the Six-Sigma©Maturity Model described by Prasad[31], were evaluated and found to be relevant to the quality management area. This area will be renamed as the quality monitoring area to align with the research terminology.

The research continued with the exploration of two more maturity models: the BSIMM© (Building Security in Maturity Model) from Synopsys [**?**] and the ISM3©(Information Security Management Maturity Model) from the ISM3 Consortium [7]. Both models were classified under the security management area, which will be renamed as the security monitoring area.

Furthermore, three additional maturity models were researched. The first model, MMAST©(Automated Software Testing Maturity Model) by Mitchel Krause [20], the second model, TMMi©(Test Maturity Model Integration) referenced in the TMMi – Test Maturity Model Integration Foundation [40] link, and the third model, TOM©(Test Organization Maturity Model) by Systeme Evolutif Ltd. [23], were found to be relevant to the tests management area. This area will be renamed as the tests monitoring area.

As a result of these research activities, Figure 5 was created, presenting the identified maturity models that guided the definition of the 12 IT monitoring areas considered by the PI2M-ITGov. Additionally, these 12 IT monitoring areas were categorized into 3 IT monitoring groups for ease of interpretation and application of their activities: the planning monitoring group, the execution monitoring group, and the control monitoring group.

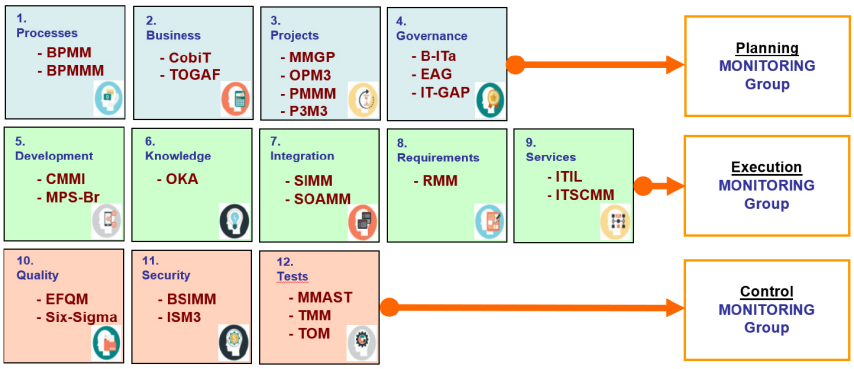


Fig. 5. The 12 monitoring areas considered by the PI2M-ITGov

After conducting the research and evaluating the requirements, it was determined that the KMIs applicable to each of the 12 IT monitoring areas would be established in order to achieve the highest maturity levels (ranging from 3 to 5) of the maturity level assessment models considered.

4.2. The Scope of the 36 Sub-KMIs - Key Monitoring Sub Indicators

The GQ(I)M framework, developed by Goethert & Hayes [13], is widely recognized in the literature as a valuable approach for constructing indicators. This framework follows a

structured process consisting of four components: goal, question, indicator, and measurement. Figure 6 exemplifies the workflow of this framework, illustrating its step-by-step approach to indicator construction.

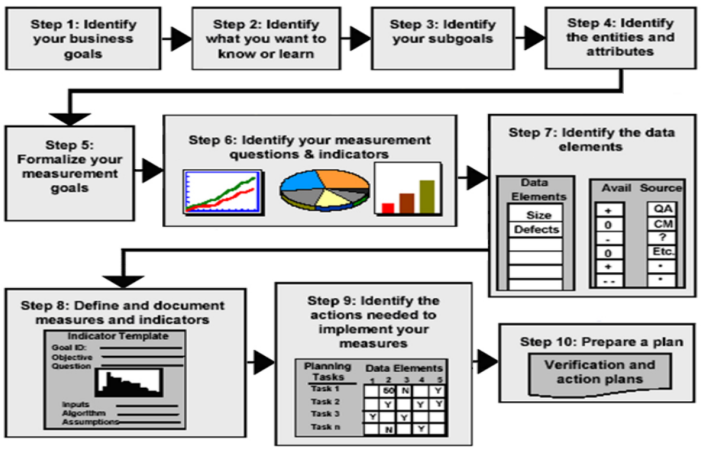


Fig. 6. GQ(I)M – Goal, question, indicator, and measurement [13]

During the synthesis process following the research, it was observed that three requirements were common and essential for attributing the highest maturity levels (ranging from 3 to 5) of the maturity level assessment models considered. These requirements were related to achieving satisfactory performance of activities, complying with schedule and budget, and meeting quality requirements.

These three evaluation criteria also aligned with the proposal put forward by one of the authors [3, p. 180–189] of this research in their doctoral thesis. They suggested reducing the Likert [21] scale, which typically has five levels of evaluation, to just three levels (referred to as the AJMM table) for greater precision in assessment. The three levels of evaluation in the AJMM table were defined as: 1st = Ok, 2nd = Ok with Restriction, and 3rd = Not Ok.

In accordance with the previously mentioned indicator construction technique and the planned approach, the assembly of the three sub-KMIs for each of the 12 IT monitoring areas was carried out as follows: one sub-KMI for the assessment of work results, one sub-KMI for the assessment of delivery commitment, and one sub-KMI for the assessment of customer satisfaction.

The following section presents the content of the three sub-KMIs, which are replicated for each of the 12 identified IT monitoring areas. The placeholder '???' will be replaced by the monitoring area code ranging from 1 to 12:

It is apparent that sub-indicator 1, WR, has the least demanding requirements among the three sub-indicators. Sub-indicator 2, DC, has moderate requirements, as it establishes that the volume of sub-indicator 1, WR, has fulfilled its restriction. On the other hand, sub-

KMI ?? – 1. WR (work result)	= Percentage of work concluded.
KMI ?? – 2. DC (delivery commitment)	= Percentage of work concluded and delivered on schedule planned and within the estimated cost.
KMI ?? – 3. CS (customer satisfaction)	= Percentage of work concluded and delivered on schedule planned and within the estimated cost, which in addition, received the definitive acceptance (and not provisional).

indicator 3, CS, is the most demanding, as it requires the cumulative fulfillment of both the volumes of sub-indicator 1, WR, and sub-indicator 2, DC.

5. Solution Validation

Following the principles of DSR, the next step, step 3, involves 'solution validation'. This step focuses on planning the implementation of the solution designed in step 2.

To calculate the sub-KMI, which represents the resulting data as a percentage for the KMI %, the following equation can be utilized.

$$\left\{ \sum \text{inspected numbers} - m \left(\frac{\text{maximum goal} - \text{minimum goal}}{2} \right) \right\} = \times \frac{100\%}{m}$$

where: $\sum = Summation m =$ Average

The '$\sum$ inspected numbers' data implies that it may be necessary to aggregate individual values for the same area/sector or across multiple areas/sectors if the time unit of the maximum goal and minimum goal differs from the volume being inspected for these data.

The maximum goal can be a smaller value than the total of an existing value in practical reality, which will be considered as the expected achievement for the KMI, or it can even be a larger value than the existing total in the real world.

Conversely, the minimum goal can be a greater value than the total of an existing value in practical reality, which will be considered as the accepted achievement for the KMI, or it can even be a smaller value than the existing total in the real world.

It is important to note that both the maximum goal and the minimum goal should be provided as quantitative numerical values rather than percentages. This approach is more feasible and simpler for obtaining this data.

If there are changes in the maximum goal and/or the minimum goal for certain KMIs, these changes should be considered simultaneously within the same time period. This ensures an equal assessment and allows for comparisons across different scenarios.

The balance for all KMIs is determined by the arithmetic average of the maximum goal and the minimum goal, also known as the maximum–minimum average. This average is compared with the 'inspected numbers' data, which enables evaluation of the KMI

against this criterion for the entire PI2M-ITGov framework. Based on this comparison, there are two interpretations: A KMI greater than 0 indicates a positive percentage calculated above the maximum–minimum Average, while a KMI smaller than 0 indicates a negative percentage calculated below the maximum–minimum average.

The interpretation of the KMI calculated as 0% would be that the data 'inspected numbers' remained exactly at the maximum–minimum average, and therefore, there was no positive or negative result. As a sample of a hypothetical example, we can say that for a maximum goal established as 8 and a minimum goal established as 4, which would result in a maximum–minimum average equal to 6: 1) if the data 'inspected numbers' were 6 the KMI would be 0%; 2) if the data 'inspected numbers' were 3 the KMI would be –50% which demonstrates a negative result, falling halfway below the maximum–minimum average; 3) if the data 'inspected numbers' were 9 the KMI would be 50% which demonstrates a positive result, exceeding the maximum–minimum average by half.

According to this logic, in the case described in the previous paragraph, the percentage of 100% of the KMI would only be calculated when the value of the data 'inspected numbers' was equal to 12, which is the sum of the maximum goal and the minimum goal. Under this same logic, positive percentages greater than 100% can be calculated when the value of the data 'Inspected Numbers' is greater than 12.

When the value of the data 'inspected numbers' is equal to 0, the KMI percentage will be negative and equal to –100%. It is not common or normal for the data 'inspected numbers' to have a negative value. However, in certain cases where it is necessary to account for previous poor results or 'debit' them in the current evaluation, this scenario could be plausible. In such cases, the KMI will be negative and lower than just 100%.

In some companies that operate in different industries or sectors, the evaluation and score of a specific IT Monitoring Area among the 12 treated areas may be considered more important than the values calculated for the other areas. This prioritization can vary depending on the nature of the business and the specific needs and goals of the company.

For instance, a company in the communication sector (such as a newspaper or magazine) may place greater emphasis on the sub-KMIs focused on knowledge management (IT monitoring area code = KMI 6). Conversely, a company in the financial sector (such as banks or loan institutions) may prioritize the sub-KMIs focused on security management (IT monitoring area code = KMI 11). Similarly, a company in the manufacturing sector (such as a factory or assembly plant) may attach more significance to the sub-KMIs focused on project management (IT monitoring area code = KMI 3).

To account for the varying significance of different KMIs in distinct scenarios, significance weights can be assigned to highlight the importance of one KMI over another during the comparison process. Therefore, the final formula used by PI2M-ITGov for calculating the KMI percentage will be modified to include the application of significance weights (represented as the letter 'W' in the equation):

$$\left\{\sum \text{inspected numbers} - m\left(\frac{\text{maximum goal} - \text{minimum goal}}{2}\right)\right\} = \times\frac{100\%}{m} \times W^2$$

or: W^3]

It is important to note that the KMI (final) and its sub-indicators are calculated with one decimal place. This precision is necessary because in decision-making processes, the decimal place can hold significant information that differentiates similar values.

A small spreadsheet has been created to simulate this formula and can be accessed in the folder '3. Solution Validation/Validação da Solução' in the link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

This spreadsheet allows for the exercise of the hypothetical example mentioned earlier, with a maximum goal of 8, a minimum goal of 4, and data 'inspected numbers' of 6, 3, and 9, resulting in KMI percentages of 0%, –50%, and 50%, respectively.

6. Solution Implementation

Step 4 of the DSR, referred to as 'solution implementation', requires exercising the solution theory and implementing the constructed models (from the prior step 3). This involves applying and evaluating the created artefacts (spreadsheet and guidelines) in a real operational environment.

To facilitate the assessment of IT governance in a given situation using the PI2M-ITGov assessment, two artefacts have been developed and are available for use in the folder '4. Solution Implementation/Implementação da Solução' in the link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

These artefacts include the MS® Excel spreadsheet entitled 'PI2M-ITGov - Spreadsheet {V.01c}' Edition=_____+Scenario=_____' and the accompanying guidelines 'PI2M-ITGov - Guidelines {V.01c}'. The guidelines provide instructions on how to complete the spreadsheet and interpret its results.

In the same link, you will also find the form for each KMI, which consists of two frames. The first frame is used to record the goals established during the evaluation planning phase in collaboration with the strategic management department. The second frame is used to record the actual data captured in the field during the execution phase of the PI2M-ITGov evaluation.

To facilitate identification, it is suggested to copy the fields 'Edition' and 'Scenario' from the spreadsheet's header to the file name, which will help in organizing and identifying the content of the spreadsheet in the directory or archive folder.

The layout of the cells in the MS®-Excel spreadsheet follows the predefined order of the 12 IT monitoring areas, divided into the three IT monitoring groups: planning monitoring group, execution monitoring group, and control monitoring group.

The current version of these artefacts, at the time of publication of this article, is V.01c, representing the third revision (letter 'c') of the first version (number '01'). Further updates and improvements may be made available in the same link following subsequent revisions.

The previous revisions (letters 'a' and 'b') were identified during the execution of the fifth and final stage of the DSR, which will be presented in the next topic. This stage involved a case study conducted over three rounds until reaching the current version in the third and final round.

Appendix A of this article presents the four main tabs of the MS® Excel spreadsheet. The first tab displays the data entry cells, while the other three tabs contain graphs for interpreting the results. The 'INSTRUCTIONS' tabs, which provide guidance on data entry and analysis of status icons, are not displayed.

It is important to note that the instructions provided in the 'PI2M-ITGov - Guidelines V.01c' document are based on the results obtained from simulating the implementation of this method in a real corporate environment, as described in Section 7, 'Implementation Assessment'.

Appendix B of this paper showcases the Chronogram (created in MS® Project) used to coordinate the execution of step 4 of the DSR. This chronogram is available in the same link under the folder '4. Solution Implementation/Implementação da Solução'. Although it is already filled, it can be copied and reused, as it provides average duration data for the tasks that can be replicated.

7. Implementation Assessment

In the final step of the DSR, step 5, known as 'implementation assessment', the results obtained from the previous step 4 were interpreted and examined to determine the utility and applicability of the proposed solution from step 3.

To assess the effectiveness of the solution, a simulation was conducted in a Brazilian Government Agency, consisting of three rounds. The outcomes of this simulation were utilized in the development of the guidelines, as mentioned earlier, which can be accessed through the provided link.

The results obtained from the simulation are presented in the spreadsheet titled 'PIM2-GovTI - Spreadsheet V.01c Edition=2021 2nd.Quarter+Scenario=Organizational Restructuring' available in the folder '5. Implementation Assessment/Avaliação da Implementação' in the link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

Additionally, in the same link, a form for each KMI (KMI Form) with two frames is available where the first frame captures the goals defined during the evaluation planning phase in collaboration with the Strategic Management Department, as part of the application of the PI2M-ITGov method. The second frame records the actual data collected in the field from the business areas during the execution phase of the PI2M-ITGov evaluation.

The 3rd. and final round of this simulation, was to evaluate the 2nd. Quarter of the assessment year and, in this quarter, this assessment was performed at a time when the Company was undergoing Organizational Restructuring.

There are comments, in the above Spreadsheet, where this scenario was described and where is possible to be understood that even with this issue the Method could be implemented and retrieve extra-results those were applied to support decision about this Organizational Restructuring in progress.

Following the Chronogram (created in MS® Project) used to coordinate the execution of step 4 of the DSR (which is available in the same link under the folder '4. Solution Implementation/Implementação da Solução' and presented in the Appendix B) this simulation handles the numbers exposed in the Figure 7 when was applied the KMI Form.

Many interpretations, which led the need of immediate corrective actions, could be found analyzing the results in the spreadsheet titled 'PIM2-GovTI - Spreadsheet V.01c

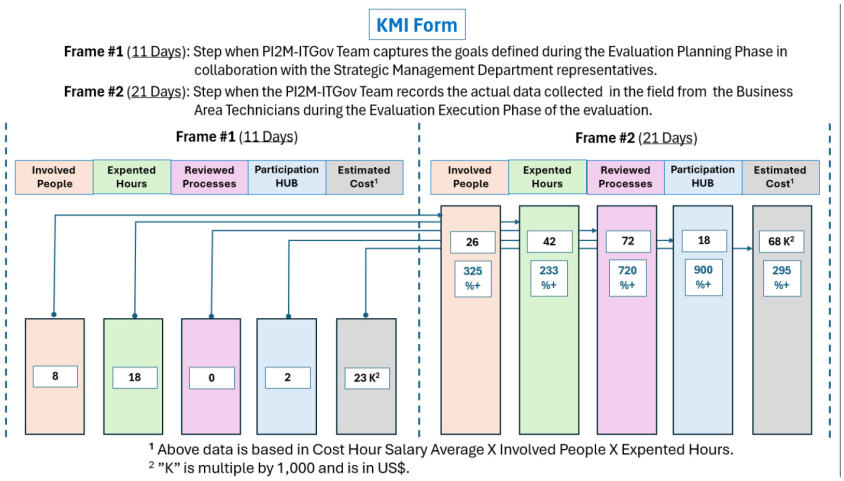


Fig. 7. Numbers retrieved from Final Simulation

Edition=2021 2nd.Quarter+Scenario=Organizational Restructuring' available in the folder '5. Implementation Assessment/Avaliação da Implementação' in the link `https://drive.google.com/drive/folders/1sFa5845Wcxr7KXQlH9hePXxN0WdRKchV?usp=sharing`.

Those corrective actions were discussed with the Board of Directors of the Brazilian Government Agency, where this simulation and assessment were applied, during the 4th. Phase = Evaluation of the Chronogram (created in MS® Project which is available in the same link under the folder '4. Solution Implementation/Implementação da Solução' and presented in the Appendix B)

One of the interpretations about the KMI 1 - Processes is that this Processes had a high level of Production with 100% in the Sub-KMI 1. WR (work result) but with many difficulties to get the values for the Sub-KMI 2. DC (delivery commitment) and 3. CS (customer satisfaction), what of course, impacted indeed the analysis that could be made by the Strategic Area.

Other of them is according to KMI 6 - Knowledge which the evaluation shown that this Processes had no procedures implemented in this theme, what can be interpreted that the Organization in assessment is losing the monitoring of its successful actions to register these and reuse as practical and useful feedback.

Also, could be found that the KMI 11 - Security, even this Process having a very high importance for a effective IT Governance, was impossible to be defined and the explanation for that was: "All 3 Subindicators of this KMI will be evaluated only in the next Time Period as this Security Area is in the process of reviewing the Internal Processes." (as is in the comments in the Spreadsheet).

Regarding the KMI 3 - Projects must be highlighted that its 3 Sub-KMIs, what means 1. WR (work result), the Sub-KMI 2. DC (delivery commitment) and 3. CS (customer satisfaction), had these 3 ones done under the expectations and can be considered as a successful Process all concluded.

These above interpretations were consistent with the evaluation obtained in the KMI 4 - Governance where the Sub-KMI 3. CS (customer satisfaction) had a negative value even with the follow explanation which can justify this grade but drive to reinterpret this result: “The Division Manager justified the very low value of the 3. SC above due to the large number of his Customers on vacation or on leave that still could not assess the Degree of Satisfaction of the reports delivered.” (as is in the comments in the Spreadsheet).

The guidelines provided in the link serve as a practical resource for understanding how the simulation was conducted, in addition to being a tool for applying the PI2M-ITGov method.

As mentioned in the initial summary of this research work, the PI2M-ITGov method can be utilized to (1) assess the performance status of an IT area. Furthermore, it can also be used to (2) compare the results of suppliers as a means of supporting decision-making, particularly in the context of proof of concept (POC) evaluations for contracting or renewing commercial agreements.

The MS® Excel spreadsheet discussed earlier primarily focuses on objective (1) mentioned above. However, to address objective (2) and provide an immediate assessment of a POC, the PI2M-ITGov KMIs can be applied in the manner depicted in Figure 8 (for evaluating service provider enterprises in the field of systems development, specifically as a software factory [FSW]) and Figure 9 (for evaluating service provider enterprises in the field of project management, particularly as a project management office [PMO]).

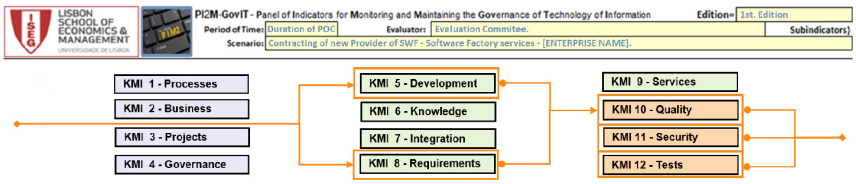


Fig. 8. Example of proof of concept (POC) to build a comparative framework in system development

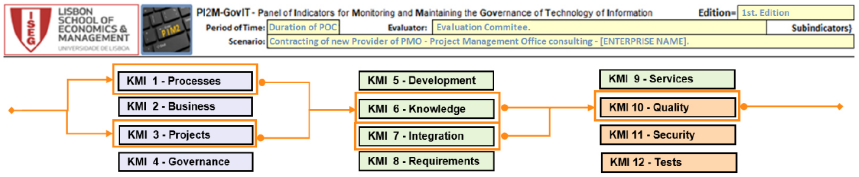


Fig. 9. Example of POC - Proof of Concept to build a Comparative Framework in Project Management

In both examples, the assessment of sub-indicator 2. DC should focus solely on the deadline criterion, excluding the cost criterion. In a POC evaluation, cost evaluation is not typically a common factor.

Furthermore, in both examples, regarding sub-indicator 3. CS, the term 'customer' represents the evaluation committee responsible for applying the final acceptance criteria for the POC presented by the assessed supplier.

For these examples, the maximum goal and minimum goal for the KMIs should be the same. Since the POC has a short duration and aims to achieve a specific outcome, the maximum goal and minimum goal reflect the singular target value that the evaluated supplier must meet.

8. Future Work

This article limited its approach in retrieve data and analyzes results from how the IT Governance, in a specific Organization, are being conducted – with the target – to gain able time to realign deliveries and expectation, based in the status obtained with a precise monitoring, before a massive control and auditing are not yet necessary in fact.

Nevertheless, we can expand this approach to compare the result obtained by a specific Organization with another to sort its ranking, what means, transforming PI2M-ITGov Method into a real MM - Maturity Model itself acting in distinct Organizations.

As a first step, we can think (as an initial idea) of creating a ranking with Gold, Silver and Bronze levels, what will drive us to need to define the maximum and minimum range grade for each Level according to each 36 sub-KMIs results.

In the sequence of this new approach implementation, it is necessary to establish criteria (such as: Capital Amount, Employes Number, Type of Business Finance; Industry; Services, Capillarization (Local; National; Intl), and other points) for sorting the Organizations in Small, Medium and Big sizes, so it will be possible compare different scenarios and apply an "Adjustment Weight" in order to equalize the grades obtained and attributes the correct Gold, Silver and Bronze level.

If this idea will be put in practice, of course, a new article will be built and explaining how it was done in details and how to use the new approach of PI2M-ITGov Method as a MM - Maturity Model between different Organizations to obtain a unique Pannel of Statistics."

9. Conclusion

This research adhered to the principles of a scientific method by following the steps of DSR, including the application of a case study consisting of three simulation rounds to validate the constructs and artefacts created (guidelines and spreadsheet).

Based on the evaluation, it can be concluded that the PI2M-ITGov, which is a panel of indicators for monitoring and maintaining IT governance, effectively serves as a practical tool for executing the necessary procedures in IT management.

While the PI2M-ITGov was constructed in a structured manner, its continued use and feedback from users will likely lead to new versions with added functionalities and improvements.

The author welcomes and appreciates users who provide feedback and share their experiences in using the PI2M-ITGov. This feedback will contribute to enhancing the quality of the method.

It is recommended to periodically check the provided link for any new updates or revisions that can be downloaded.

Furthermore, the author encourages future contacts and assures that any inquiries will be handled with care and attention. The author is dedicated to guiding interested individuals in implementing the PI2M-ITGov in their organizations and providing necessary support to ensure the success of this initiative.

Acknowledgments. Partial financial support was received from ITMA - Information and Technology Management Association for sponsoring the proofreading revision costs.

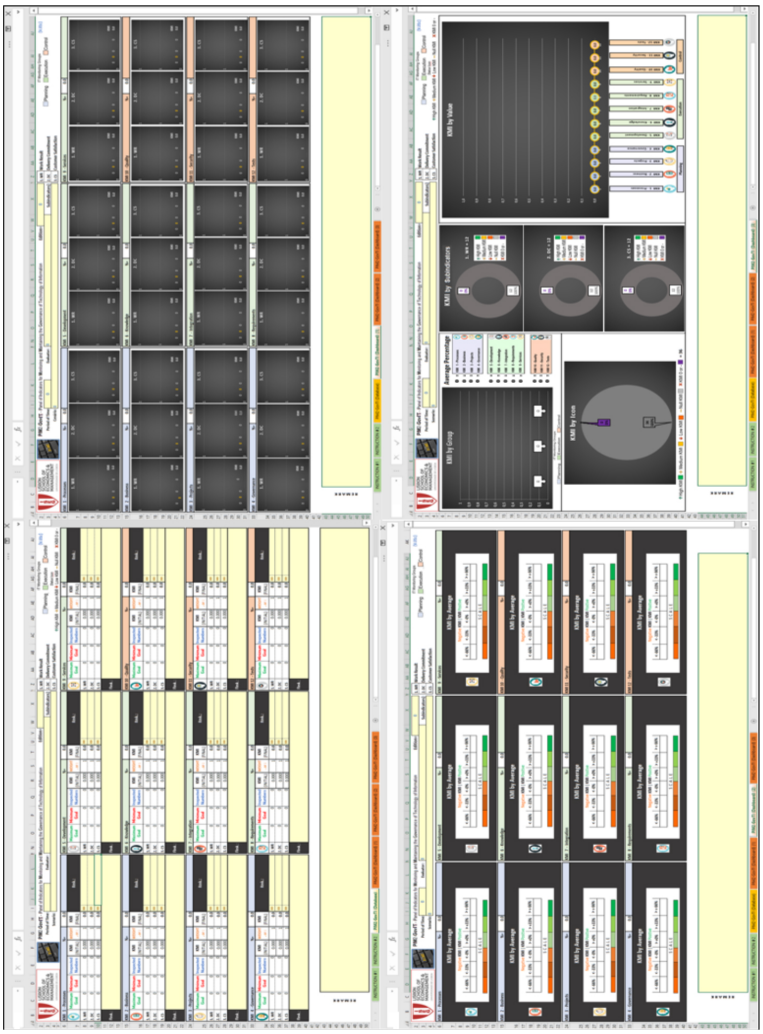
References

1. Ahern, D.M., Clouse, A., Turner, R.: CMMI Distilled – A Practical Introduction to Integrated Process Improvement. 2nd ed. Pearson Education Inc., Boston, Massachusetts, USA (2002)
2. Arsanjani, A., Holley, K.: The Service Integration Maturity Model: Achieving Flexibility in the Transformation to SOA©, Services Computing, IEEE International Conference on, p.515, IEEE International Conference on Services Computing (SCC'06), ISBN: 0-7695-2670-5. Chicago, Illinois, USA, Sep 18–22 (2006)
3. Axelos: ITIL® Foundation: ITIL 4 Edition. Axelos Ltd. (2022), https://www.axelos.com/best-practice-solutions/itil
4. Axelos: P3M3 – Portfolio, Programme and Project Management Maturity Model. Axelos Ltd. (2022), https://www.axelos.com/best-practice-solutions/p3m3
5. CIOIndex: EAG© – Enterprise Architecture Governance. CIOIndex – Global Community For Chief Information Officers (2022), https://cio-wiki.org/wiki/Enterprise_Architecture_Governance
6. Clerc, V., Niessink, F.: IT Service CMM, A Pocket Guide. Van Haren Publishing, Holland (2005)
7. Consortium, I.: Information Security Management Maturity Model., chap. ISM3 Consortium - Creative Commons Attribution. ISM3 Consortium - Creative Commons Attribution, ISBN: 978-54-611-9825-2. San Francisco, California, USA (2007)
8. Dresch, A., Lacerda, D.P., Antunes Jr., J.A.V.: Design Science Research: Research Method for Advancing Science and Technology (Método de Pesquisa para Avanço da Ciência e Tecnologia). Bookman, Porto Alegre, RS, Brazil (2015)
9. EFQM©: The Fundamental Concepts of Excellence. European Foundation for Quality Management (2022), https://www.onecaribbean.org/wp-content/uploads/Fundamental-Concepts-of-EFQM.pdf
10. Fonseca, A.F.: Organizational Knowledge Assessment Methodology. World Bank Institute, Washington, DC, USA (2006)
11. Fontelles, M.J., Simões, M.G., Farias, S.H., Simões, R.G.: Scientific Research Methodology – Guidelines for Elaboration of a Research Protocol. Cercomp, Brazil (2009)
12. Gartner, G.: CobiT® – Control Objectives for Information and Related Technology. Gartner Group (2022), https://www.gartner.com/en/information-technology/glossary/cobit-control-objectives-for-information-and-related-technology
13. Goethert, W., Hayes, W.: Experiences in implementing measurement programs. Technical note, Software Engineering Measurement and Analysis Initiative, SEI® – Software Engineering Institute (2001), CMU/SEI-2001-TN-026, NOV

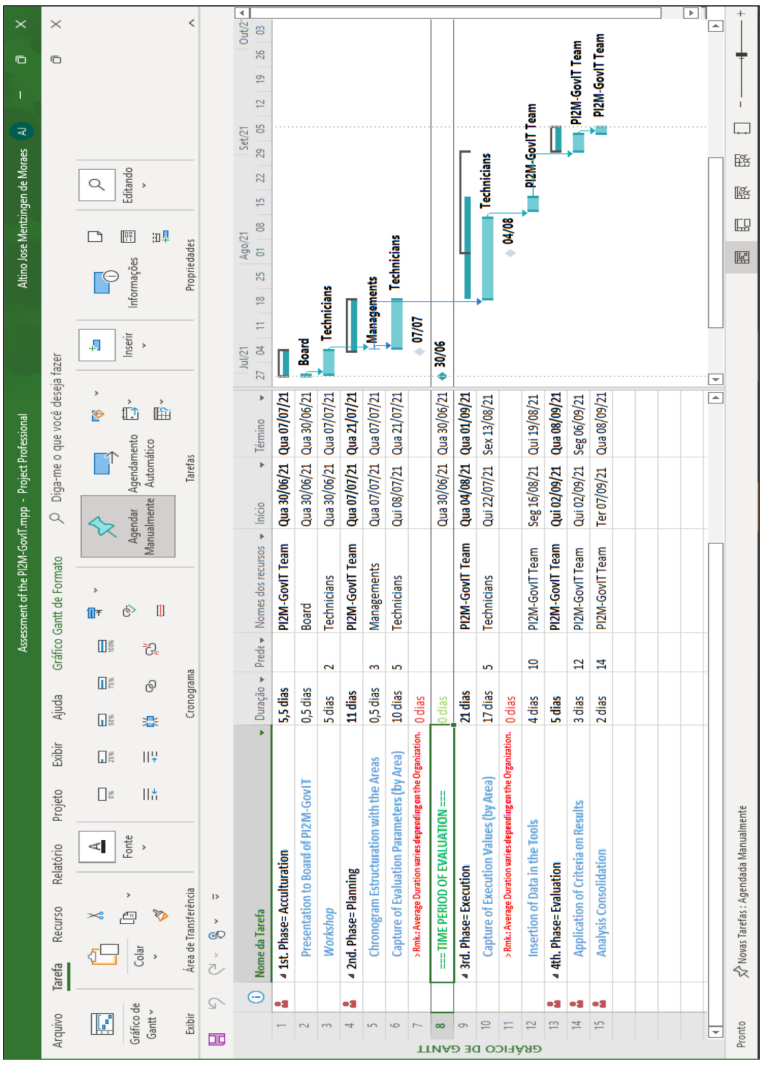
14. Heumann, J.: The Five Levels of Requirements Management Maturity – Rational Software (2022), https://pt.slideshare.net/ibmrational/the-five-levels-of-requirements-management-maturity
15. Hevner, A.: A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems 19(2), 4 (2007)
16. Hüffner, T.: The BPM® Maturity Model – Towards A Framework for Assessing the Business Process Management Maturity of Organizations. GRIN Verlag, Munich, Germany (2007)
17. ISACA: CobiT® – Control Objectives for Information and Related Technology. Information Systems Audit and Control Association (2022), https://www.isaca.org/resources/cobit
18. ISO/IEC: 15.504 – Process assessment – Part 5: An Exemplar Software Life Cycle Process Assessment Model. International Organization for Standardization (2022), https://www.iso.org/standard/60555.html
19. Kerzner, H.: Using the Project Management Maturity Model: Strategic Planning for Project Management, 2nd ed. John Wiley & Sons Inc, New York, NY, USA (2004)
20. Krause, H.M.: A maturity model for automated software testing. Medical Device & Diagnostic Industry Magazine (1994), mDDI Article Index
21. Likert, R.: A technique for the measurement of attitudes. Archives of Psychology 140, 1–55 (1932)
22. Locke, J.: The Works of John Locke: An Essay Concerning Human Understanding. Halcyon Classics – Kindle Edition, Jun-23 (2009)
23. Ltd., S.E.: Test Organization Maturity Questionnaire V2.0. Systeme Evolutif Ltd, Gloucester House, 57/59 – Gloucester Place, London, England (2000)
24. Moraes, A.J.M.: M2A3-GovTI Proposal – Maturity Model for Analysis of the Alignment of Activities Related to Governance of Technology of Information in Accordance with the Outcome Expectations Planned by the Organization (Proposta do M2A3-GovTI – Modelo de Maturidade para Análise do Alinhamento das Atividades relacionadas à Governança da Tecnologia da Informação em conformidade com as expectativas de resultado planejadas pela Organização). Doctoral thesis, Stricto-Sensu Postgraduate Program in Production Engineering Doctorate. UNIP – Paulist University, São Paulo, SP, Brazil (2010)
25. Moraes, A.J.M., Rocha, M.R.: IT Management: Proposal of the PI2M-ITGov - Panel of Indicators for Monitoring and Maintaining IT Governance: Research and Model (2023), https://ieeexplore.ieee.org/document/10211818
26. OMG: BPMM® – Business Process Maturity Model. Object Management Group (2022), http://www.omg.org/spec/BPMM/1.0/PDF
27. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of Management Information Systems 24(3), 45–77 (2007)
28. Peterson, R.: Crafting Information Technology Governance. Taylor & Francis Online (2022), https://www.tandfonline.com/doi/abs/10.1201/1078/44705.21.4.20040901/84183.2
29. PMI®: OPM3® – Organizational Project Management Maturity Model. Project Management Institute (2022), https://www.gartner.com/en/information-technology/glossary/cobit-control-objectives-for-information-and-related-technology
30. Prado, D.: MMGP – A Brazilian Model of Project Management Maturity (Um Modelo Brasileiro de Maturidade em Gerenciamento de Projetos). Ponto GP (2005), http://pontogp.wordpress.com/2006/05/06/mmgp
31. Prasad, R.: Maturity Model Describes Stages of Six-Sigma® Evolution. iSixSigma (2010), https://www.isixsigma.com/basics/maturity-model-describes-stages-six-sigma-evolution/

32. Rockart, J.F., Earl, M.J., Ross, J.W.: Eight imperatives for the new it organization. Sloan Management Review 38(1), 43–55 (Fall 1996)
33. Rundle, R.: Deming Cycle PDCA – Plan Do Check Act Journal in Daily Life Toyota Way. Independently Published (2019)
34. SEI®: CMMI® – Capability Maturity Model Integration. Software Engineering Institute (2022), https://resources.sei.cmu.edu/asset_files/TechnicalReport/2010_005_001_15287.pdf
35. Simon, H.A.: The Sciences of the Artificial. 3rd ed. MIT Press, Cambridge, Massachusetts, MA, USA (1996)
36. SOFTEX: MR MPS br®: Reference Model for Brazilian Software Process Improvement (Modelo de Referência para Melhoria do Processo de Software Brasileiro). Softex (2022), http://www.softex.br/mpsbr/_guias/default.asp/
37. Sonic: The New SOA® Maturity Model. Sonic Software Corporation, AmberPoint Inc., BearingPoint Inc., Systinet Corporation. All rights reserved (2005)
38. Tapi, R.S., Daneva, M., Eck, P., Wieringa, R.J.: B-ITa – Business-IT Alignment Maturity Model. University of Twente (2022), https://research.utwente.nl/en/publications/towards-a-business-it-alignment-maturity-model-for-collaborative
39. Tesla, N.: Map of Multiplication. Unicentro, Paraná, PR, Brazil (2021), https://www3.unicentro.br/petfisica/2017/05/18/o-mapa-da-multiplicacao-de-tesla/
40. TMMi: Test Maturity Model Integration. TMMi Foundation (2023), https://www.tmmi.org/tmmi-documents/
41. Weill, P.: The relationship between investment in information technology and firm performance: A study of the valve manufacturing sector. Information Systems Research 3(4), 307–333 (Dec 1992)
42. Wieringa, R.J.: Design Science Methodology for Information Systems and Software Engineering. Springer-Verlag Berlin Heidelberg, Berlin, Germany (2014)

Appendix A: MS® Excel Spreadsheet of PI2M-ITGov



Appendix B: MS® Project 'Assessment of the PI2M-ITGov'



Altino Moraes holds the title of Ph.D. in Production Engineering (with emphasis in Information Systems), M.Sc. in Information Management, and BCs in Information Technology. He also developed Post Doctoral Research with an IT Governance focus. Additionally, he has two lato-sensu courses: one MBA in Project Financial Management and one Specialization in System Analysis. He is certified in five Technical Certifications applied to the IT Area (CobiT, ITIL, PMI, ISTQB, and ABPMP). He is a Researcher at the ADVANCE (the ISEG Centre for Advanced Research in Management at the University of

Lisbon). He is a member of Scientific Committees in congresses sponsored by IIIS - International Institute of Informatics and Systemics, ITMA - Information and Technology Management Association and AISTI - Iberian Association of Information Systems and Technologies. He often participates in Conferences related to the IT Area and publishes in its Proceedings. Besides, he has been participating in books (as coeditor/coauthors) distributed by IGI Global. He is also a Brazilian Government Employee currently allocated to MGI - Ministério da Gestão e Inovação (Management and Innovation Ministry).

Álvaro Rocha holds the title of Honorary Professor, and holds a D.Sc. in Information Science, Ph.D. in Information Systems and Technologies, M.Sc. in Information Management, and BCs in Computer Science. He is a Professor of Information Systems at the University of Lisbon - ISEG, researcher at the ADVANCE (the ISEG Centre for Advanced Research in Management), and a collaborator researcher at both LIACC (Laboratory of Artificial Intelligence and Computer Science) and CINTESIS (Center for Research in Health Technologies and Information Systems). His main research interests are maturity models, information systems quality, online service quality, requirements engineering, intelligent information systems, e-Government, e-Health, and information technology in education. He is also Vice-Chair of the IEEE Portugal Section Systems, Man, and Cybernetics Society Chapter, and Founder and Editor-in-Chief of two Scopus and SCIMago journals: JISEM - Journal of Information Systems Engineering & Management; and RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação / Iberian Journal of Information Systems and Technologies. Moreover, he has served as Vice-Chair of Experts for the European Commission's Horizon 2020 Program, and as an Expert at the COST - intergovernmental framework for European Cooperation in Science and Technology, at the European Commission's Horizon Europe Program, at the Government of Italy's Ministry of Universities and Research, at the Government of Latvia's Ministry of Finance, at the Government of Mexico's National Council of Science and Technology, at the Government of Polish's National Science Centre, and at the Government of Cyprus's Research and Innovation Foundation.World's Top 0.05% Scientist, according to ScholarGPS. World's Top 1% Scientist, according to Stanford University and Elsevier. World's Top 1% Scientist, according to ResearchGate. ISEM's Book Series Scientific Manager at Springer Nature: https://www.springer.com/series/17396. Chair of ITMA - Information and Technology Management Association: http://itmas.org. Founder and Vice-Chair of IEEE SMC Portugal Chapter. Invited Full Professor at University of Calabria, Italy. Honorary Professor at Amity University, India. Professor of Information Systems, ISEG, University of Lisbon, Portugal.

Received: July 29, 2024; Accepted: January 12, 2025.

Computer Science and Information Systems 22(3):839–858 https://doi.org/10.2298/CSIS240831018Z

Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy in Spark

Yanhao Zhang, Congyang Wang, Xin He* , Junyang Yu , Rui Zhai and Yalin Song

School of Software, Henan University
Kaifeng,Henan,China,475000
zhangyanhao@henu.edu.cn
wangcongyang@henu.edu.cn
hexin@henu.edu.cn
jyyu@henu.edu.cn
zr@henu.edu.cn
122648935@qq.com

Abstract. For solving the low CPU and network resource utilization in the task scheduler process of the Spark and Flink computing frameworks, this paper proposes a Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy (DRTS). This algorithm can schedule parallel tasks in a pipelined fashion, effectively improving the system resource utilization and shortening the job completion times. Firstly, based on historical data of task completion times, we stagger the execution of tasks within the stage with the longest completion time. This helps optimize the utilization of system resources and ensures the smooth completion of the entire pipeline job. Secondly, the execution tasks are categorized into CPU-intensive and non-CPU-intensive phases, which include network I/O and disk I/O operations. During the non-CPU-intensive phase where tasks involve data fetch, parallel tasks are scheduled at suitable intervals to mitigate resource contention and minimize job completion time. Finally, we implemented DRTS on Spark 2.4.0 and conducted experiments to evaluate its performance. The results show that compared to DelayStage, DRTS reduces job execution time by 3.18% to 6.48% and improves CPU and network utilization of the cluster by 6.33% and 7.02%, respectively.

Keywords: Job execution time, delay-aware, Spark, task scheduler.

1. Introduction

With the rapid development of the Internet, the size of data has increased dramatically, creating a greater demand for real-time data processing. Big data analytics has aroused the interest of scholars because of its ability to deal with large data sets. Many open-source parallel processing frameworks, such as MapReduce[26] , Hadoop[8], Storm[3] to the later Spark[2] and Flink[1], have been developed to handle large data volumes. These frameworks have evolved through various stages, including the Map-Reduce model, the DAG model, the streaming model, and the real-time model.

These computing frameworks break a job into multiple tasks and assign them to work nodes for large-scale data processing. Task scheduling efficiency is a major bottleneck that affects the framework's performance[10].

* Corresponding author

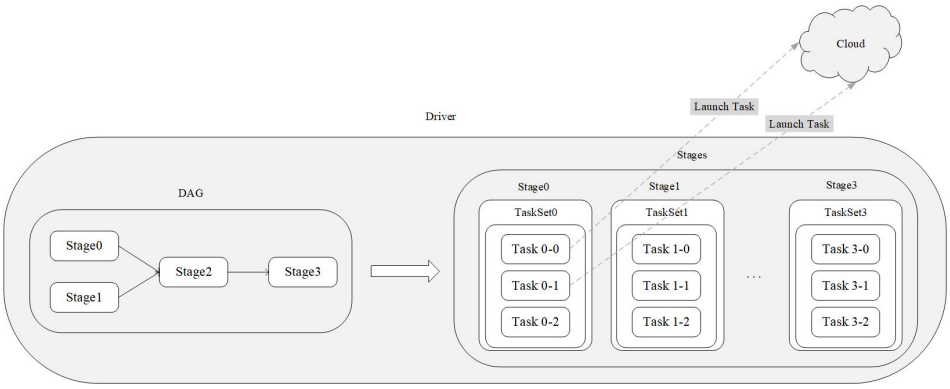


Fig. 1. From DAG scheduling to task scheduling

Moreover, as data centers continue to expand in scale, their energy consumption issues have become increasingly prominent. Research indicates that the energy consumption of cloud data centers accounts for a significant portion of overall operational costs[19]. Adopting efficient task scheduling algorithms can not only enhance computational performance but also significantly reduce energy consumption[21][20]. In this context, optimizing Spark's task scheduling not only improves computational efficiency to meet the demands of real-time data processing but also reduces data center energy consumption through more rational resource management.

In Spark, tasks are executed in parallel, but the scheduler on the driver side sends batches of tasks serially to the target machine based on a priority selection algorithm. The target machine then performs network I/O to fetch the data and carry out the computation, as shown in Fig. 1. This process results in severe CPU and network contention due to the simultaneous submission of parallel tasks. Until these tasks are completed, subsequent tasks that depend on their results cannot be scheduled. This contention leads to unbalanced resource utilization, reduces efficiency, and prolongs task completion time.

Although tasks in Spark are executed in parallel, the scheduler on the Driver side calculates task priorities based on a selection algorithm and sends tasks serially in batches to target machines. The target machines then retrieve data through network I/O and perform computations, as illustrated in Fig. 1. However, existing studies[22] have pointed out that this scheduling method may lead to severe CPU and network resource contention. In this process, concurrently submitted parallel tasks compete for limited computing and network resources, resulting in decreased resource utilization. For example, experimental results in[13][7] show that under conditions of high concurrent task submissions, task completion times are significantly extended. Furthermore, until all tasks in these parallel stages are completed, subsequent stage tasks that depend on their results will not be scheduled. This not only leads to unbalanced resource utilization but may also cause system bottlenecks[4]. Therefore, considering resource contention factors in scheduling strategies is crucial for improving system performance and resource utilization efficiency.

In Spark's default scheduling strategy, tasks from different stages are almost all executed in parallel. These tasks compete for network resources and process data while keeping CPU resources idle, or they keep bandwidth or disk idle while contending for CPU

resources. Resource utilization fluctuates drastically between extremes of overutilization and underutilization, resulting in inefficient use.

We conducted trace sampling of the average CPU utilization and network throughput across multiple machines, as shown in Fig.2. The results indicate that the CPU and network resources of most machines are not fully utilized; the average CPU utilization and network bandwidth utilization consistently remain at low levels. Therefore, we infer that when resource contention occurs, using Spark's default scheduling strategy leads to low resource utilization during job execution in the cluster.

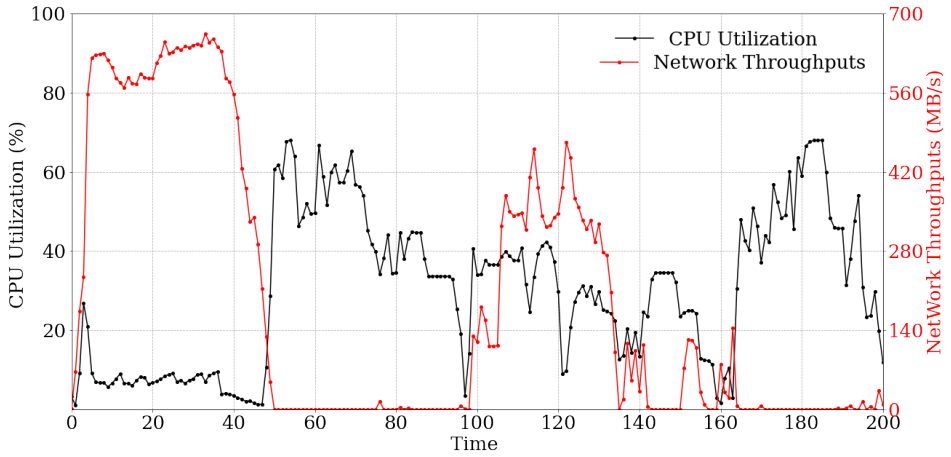


Fig. 2. The utilization of CPU and Network Throughputs

Nowadays, many scholars are focused on job scheduling [27] [11], stage scheduling [5][16], and task scheduling [17][18] to minimize job completion time and improve cluster performance. However, most of these scheduling strategies are coarse-grained and only optimize the overall execution of tasks. They do not consider that all tasks in the task set will be submitted serially to different target machines and start executing simultaneously, leading to peak resource contention.

To address this gap and further enhance cluster resource utilization while reducing job completion time, this paper approaches the problem from the perspective of task scheduling. By incorporating the scheduling priorities among parallel tasks, we analyze the various factors that affect resource utilization. By greedily selecting the tasks with the longest execution times within nodes, we determine when tasks should be scheduled.

We can illustrate this situation with a typical parallel computing task. When a task includes multiple sets of parallel parent tasks and one set of child tasks, Spark's default scheduling will relay and send parallel tasks serially. This can lead to severe network and CPU contention within the cluster during certain periods. However, the application's completion time is only related to the completion time of the longest stage. The simultaneous execution of other parallel parent task sets actually competes for resources with the longest stage, thereby affecting the final completion time.

To address this, this paper designs a Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy (DRTS). DRTS is applicable to most parallel distributed computing frameworks. Without affecting the completion of the next stage in the pipeline, DRTS delays the execution of different task sets according to the characteristics of the cluster machines they are sent to, minimizing peak resource contention and achieving resource-efficient interleaved. Integrated into the distributed framework Spark, DRTS divides the execution state of tasks into two stages: non-CPU-intensive and CPU-intensive stages.Based on the above analysis, our scheduling strategy aims to schedule tasks on different nodes at optimal times. When an execution task utilizes network resources for data fetching in the non-CPU-intensive phase, we schedule parallel tasks to execute alongside it, staggering the utilization of the cluster's resources and reducing resource competition. Therefore, the contributions of this paper can be summarized as follows:

(1) On the basis of fully considering whether there is resource contention in the cluster, this paper proposes a Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy (DRTS). This algorithm prioritizes the scheduling of jobs with long execution times according to the obtained relationship between task execution time, execution machine, and stage. Additionally, it performs interleaved execution of long and short tasks for the tasks in the stage with the longest execution time.

(2)For other parallel stages, the task execution time is calculated (including CPU-intensive and non-CPU-intensive stage times), and they are scheduled at the appropriate times. Continuous negative feedback is applied based on the task completion effectiveness, resulting in significant improvements in practical applications.

(3)Implemented a prototype model of DRTS on Spark 2.4.0 and conducted several experiments to evaluate its performance. The experiments show that DRTS enhances resource utilization and reduces job completion time.

2. Related Work

In big data processing frameworks, task scheduling is a critical step to ensure efficient resource utilization and rapid job completion. Spark's task scheduling module primarily consists of the DAGScheduler and TaskScheduler. These two components are responsible for partitioning user-submitted computational tasks into different stages according to a Directed Acyclic Graph (DAG) and assigning the computational tasks within these stages to different nodes in the cluster for parallel computation. Moreover, based on the various transformations and actions of RDDs, Spark enables users to implement strategies using complex topologies without significantly increasing the learning cost. However, because tasks are executed in parallel, this can lead to frequent usage of system resources during certain periods while they remain relatively idle at other times, thereby causing resource contention issues.

In addition to traditional scheduling strategies such as First-In-First-Out (FIFO) or Fair Scheduling (FAIR), which employ techniques like delay scheduling [24] to improve cluster performance, many studies have focused on addressing resource contention issues. For example, Xu et al. [23] proposed and developed the middleware Dagon, which, by considering and analyzing the dependency structure of jobs and heterogeneous resources, enables reasonable task allocation. They designed a sensitivity-aware task scheduling mechanism to prevent Executors from waiting for location-insensitive tasks for long pe-

riods and implemented cache eviction and prefetch strategies based on the priority of stages. Pan [15]proposed a task scheduling strategy for heterogeneous storage clusters that classifies tasks based on data locality and storage type. This approach redefines the priorities of different types of tasks according to storage device speeds and data locality to reduce task execution time. Lu et al. [14] argued that different stages have varying performance and resource requirements for different tasks, which could lead to longer overall task completion times. As a result, they proposed a task scheduling algorithm based on a greedy strategy, which balances job distribution across nodes to efficiently complete job scheduling tasks in heterogeneous clusters.

In contrast to the aforementioned works, DRTS performs task-level scheduling with finer granularity of control. DRTS assigns different priorities to tasks allocated to each node, ensuring that they are scheduled at the appropriate times, thus minimizing task completion times. Recently, Shao et al. [16] designed DelayStage, which arranges the execution of stages in a pipelined manner to minimize the completion time of parallel stages and maximize the performance of resource interleaving. However, this scheduling algorithm is coarse-grained, and simply delaying the submission time of stages still results in contention for CPU and network resources, affecting cluster execution efficiency. Duan et al. [5] argued that adding more computational resources may not significantly improve data processing speed and proposed a resource pipeline scheme aimed at minimizing job completion time. They also investigated an online scheduling algorithm based on reinforcement learning, which can adaptively adjust to resource contention. However, the reinforcement learning-based scheduler is currently only applicable to the Spark processing framework, and its compatibility with other data processing frameworks has not yet been determined.

In addition, many studies focus on stage-level scheduling strategies, while DRTS operates at the task level. DRTS addresses resource contention issues that arise from subtle time gaps in stage scheduling by performing fine-grained task execution analysis and using appropriate algorithms to schedule tasks at the optimal time, thereby interleaving system resource usage. For example, in [17], it was proposed to invoke new network-intensive tasks during non-network stages, executing two tasks in a pipelined manner by sharing the same CPU core. In [12], the design of Symbiosis, an online scheduler, allows for predicting resource utilization before task initiation and refilling compute-intensive tasks when launching network-intensive ones. In contrast, DRTS breaks tasks into CPU-intensive and network-intensive phases and achieves interleaved resource utilization by appropriately delaying task execution.

Hu et al. [10] pointed out that traditional scheduling strategies do not consider job size and designed a Shortest Job First (SJF) scheduling algorithm to avoid large jobs from blocking smaller ones. Zhang et al. [23] proposed a task scheduling method in heterogeneous server environments, based on data affinity, to minimize the maximum task completion time. He et al. [9] introduced a network-aware scheduling method, SDN, to eliminate communication barriers between the cluster computing platform and the underlying network. Zhang et al. [25] optimized scheduling using hierarchical algorithms and node scheduling algorithms, incorporating dynamic factors such as task runtime and CPU utilization on work nodes. Fu [6] utilized a bipartite graph model to propose an optimal location-aware task scheduling algorithm to reduce execution time delays and network congestion caused by cross-node data transfers.

Although a significant amount of research has been devoted to optimizing Spark's task scheduling strategies, most of the work has focused on the stage level, and some limitations remain: the details at the stage level are not fully utilized, leading to continued resource contention; dynamic resource adjustment is not performed, resulting in imbalanced resource utilization during high-concurrency task submissions; and some scheduling algorithms based on specific technologies (such as reinforcement learning) are only applicable to certain frameworks, lacking broad applicability.

To address these limitations, this paper proposes a resource-interleaved task scheduling algorithm (DRTS). DRTS overcomes the shortcomings of existing stage-level scheduling strategies by employing fine-grained task-level scheduling and dynamic resource management. It provides more efficient resource utilization and shorter job completion times, significantly improving the overall performance of Spark clusters.

3. Dual-phase Task Scheduling Strategy for the Spark Platform

When task scheduling is not properly managed, CPU and network resources are not fully utilized, leading to longer task execution times. To improve the resource utilization and job execution efficiency of the Spark platform, it is essential to address the resource contention issue.

This paper designs and implements a resource-interleaved task scheduling strategy (DRTS). The approach first greedily selects stages with the longest execution times for scheduling, in order to avoid delays in job completion. Then, by analyzing the logs generated during the job execution, detailed information for each task corresponding to the nodes is extracted. For the stage with the longest execution time, tasks are executed alternately in a way that switches between long and short tasks within the node. For the remaining stages, appropriate algorithms are used to ensure that tasks are executed at the optimal time, with continuous feedback to adjust the scheduling timing. This aims to interleave node resource usage and minimize job completion time.

3.1. Task Scheduling Optimization Strategy based on Resource Interleaving

Each task submitted by the user forms an RDD in a DAG. If an RDD requires a shuffle during the transformation process, the DAG is divided into different stages. Due to the shuffle, these stages cannot be computed in parallel, as the subsequent stages depend on the results of the preceding stages. Therefore, we divide the DAG at the boundaries of parallel stages and represent this with the symbol φ_m.

As shown in Fig. 3, the downstream shuffle stages in this splitting path are not always connected to the stages in the next splitting path. To avoid resource contention, it is necessary to appropriately delay the start time of such tasks, which will be explained in detail in Alg. 2.

In a Spark job, the Spark framework prioritizes nodes with high data locality to execute tasks, aiming to minimize data transfer overhead and enhance overall performance. While the specific scheduling of tasks is not fully determined before job execution, the Spark scheduler dynamically assigns and schedules tasks based on real-time cluster state and task demands. By analyzing log information generated during job execution, detailed task information is extracted from each node, including execution time, data volume, etc.,

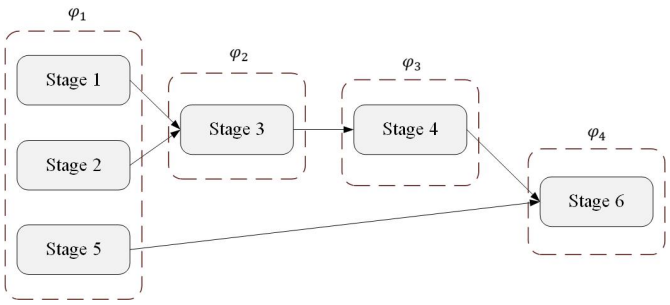


Fig. 3. Illustration of the separation of parallel phases in DAG style. Based on topological ordering, we separate the parallel stages, denoted by φ_m. The first set of parallel phases contains Stage 1, Stage 2, and Stage 5, so they can be placed in one path, i.e.,φ_1.. Stage 3 is used as the second execution path, i.e., φ_2. Stage 4's execution path is φ_3, and Stage 6's execution path is φ_4

and the scheduling node is recorded to ensure subsequent tasks are scheduled to the same node. In the case of φ_1, tasks within it are scheduled at optimal times, thereby staggering resource usage. The completion time of the longest stage in φ_1 determines the start time of φ_2. However, if stages within φ_1 are executed simultaneously, it can lead to significant resource contention, affecting execution efficiency. As shown in Fig. 4. (left), when tasks assigned to nodes execute simultaneously, it may result in severe resource contention.

We constructed an analytical model to simulate the scheduling of parallel tasks in a DAG job, determining the optimal timing for submitting parallel tasks within the job. The primary objective is to develop a scheduler that facilitates the scheduling of parallel tasks at optimal times to interleave various types of resources, thereby enhancing resource utilization and minimizing job completion time.

3.2. Task Time Statistics Strategy Based on Data Fetching and Processing

To address task resource contention, scheduling tasks at optimal times can minimize completion time by efficiently managing CPU resource requirements across different phases of the cluster. A group of Spark jobs comprises parallel computation stages, represented as $S = \{S_1, S_1, ...S_n\}$. The parallel computation phases are submitted first when there are sufficient computational resources, and each parallel phase must wait until all parallel phases have completed their computations before submitting the next phase. specifically, a stage is divided into individual tasks based on parallelism. Subsequently, the DAGScheduler submits these tasks to the TaskScheduler. We use $T_{StageId\#TaskId}$ to denote the tasks within each stage, indicating the stage number and the task number within that stage. For each task, the execution process involves several stages: initially, it requires significant network resources to fetch data; subsequently, the fetched data is processed, requiring high CPU usage; ffnally, the processed result data is written to disk. We classify the execution of a task into a non-CPU-intensive phase, denoted as $F_{i\#j}$, and a CPU-intensive phase, denoted as $P_{i\#j}$. To elucidate the execution time of tasks on the worker node W during the non-CPU-intensive phase, this paper delves into the detailed process of segmenting each task. Specifically, the non-CPU-intensive phase primarily involves

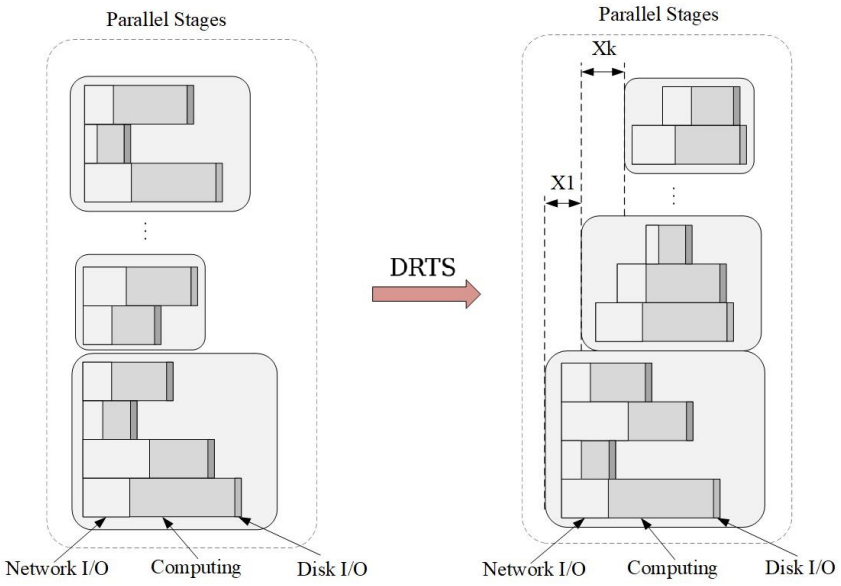


Fig. 4. Three-stage scheduling optimization strategy

data transfer, encompassing data reading and writing. Hence, the execution time of a task during the non-CPU-intensive phase on worker node W can be expressed as follows:

$$FT'_{i\#j} = T_{sr}^{i\#j} + T_{sw}^{i\#j} \tag{1}$$

The first term of Eq.(1) $T_{sr}^{i\#j}$, occurs when a task needs to get data from other nodes or file systems. The second term, $T_{sw}^{i\#j}$,occurs when a task stores the resulting data to disk after completing the computation.

Further, $T_{sr}^{i\#j}$ is calculated as:

$$T_{sr}^{i\#j} = \frac{D_r^{i\#j}}{BW_r^{i\#j}} \tag{2}$$

The process of writing data to disk involves some computation and I/O operations. However, if the data is stored only in memory, it doesn't impose a significant demand on CPU usage. Therefore, disk I/O is categorized as a non-CPU-intensive stage.

Therefore, there is a formula for $T_{sw}^{i\#j}$:

$$T_{sw}^{i\#j} = \frac{N_w^{i\#j} * B_w^{i\#j}}{BW_w^{i\#j}} + T_{dw} \tag{3}$$

$$FT_{i\#j} = \frac{D_r^{i\#j}}{BW_r^{i\#j}} + \frac{N_w^{i\#j} * B_w^{i\#j}}{BW_w^{i\#j}} + T_{dw} \tag{4}$$

Generally, the duration of writing intermediate data to disk is not extensive. For simplicity, we omit the disk I/O time when calculating the non-CPU-intensive time. Thus, we have:

$$FT_{i\#j} = \frac{D_r^{i\#j}}{BW_r^{i\#j}} + \frac{N_w^{i\#j} * B_w^{i\#j}}{BW_w^{i\#j}} \tag{5}$$

In a task, the computation time of a non-CPU intensive phase can be used to estimate the computation time of a CPU intensive phase. Given Spark's log messages provide records for the entire task completion time, noted as $T_t^{i\#j}$, once we compute the non-CPU-intensive time for the jth task in the ith stage, we can subtract this from the total task execution time to obtain the CPU-intensive time:

$$PT_{i\#j} = T_t^{i\#j} - FT_{i\#j} \tag{6}$$

We calculated the time consumption of tasks in both CPU-intensive and non-CPU-intensive phases by combining the online and offline methods described above.

4. Algorithm Implementation of Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy

In this section, we present a Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy (DRTS). The aim of DRTS is to stagger the utilization of CPU and non-CPU resources on the worker node, thereby reducing resource contention and minimizing task completion time. Initially, we compute the time set during which a task performs data fetching and processing. Subsequently, we greedily determine the optimal scheduling time for the task using the resource polling task scheduling algorithm.

4.1. Dual-stage Task Time Estimation Based Algorithm

To facilitate finer task scheduling, DRTS splits a task's execution phase into two stages: the CPU-intensive phase and the non-CPU-intensive phase. During the non-CPU-intensive phase, tasks undertake data fetching operations, which consume significant network resources and disk I/O. Conversely, the CPU-intensive phase involves extensive computation. The time intervals during which the task resides in the non-CPU-intensive and CPU-intensive phases during execution are determined by Alg. 1. The primary execution steps are divided into the following segments:

(1) We initialize two empty maps, each with NodeId as the key. The value associated with each NodeId is a map with StageId as the key and task fetch or process time as the value. These maps represent the data fetching phase (X_fetch) and data processing phase ($X\_process$) of tasks corresponding to different stages assigned to different nodes.

(2) The execution times of the data fetching phase and data processing phase during task execution, obtained from Eq. (5) and Eq. (6), are added to the result set.

(3) By traversing the collection of tasks in stage and considering the execution order and branching, the data fetching time and data processing time of different tasks are calculated and recorded in the collections X_fetch and $X\_process$.

Algorithm 1 Phase-based Task Time Estimation

Input: The total amount of data for the i#jth task $D_r^{i\#j}$ network bandwidth BW_w , the time taken to execute the completed task $T_t^{i\#j}$, the collection of execution information for various stages obtained from historical data, denoted as set S.

Output: Time maps X_fetch,$X\_process$

Initialize: $X\_fetch = \{\emptyset\}$, $X\_process = \{\emptyset\}$
for all S_i in S **do**
  for all $T_{i\#j}$ in S_i **do**
    Calculate the task data fetching time $FT_{i\#j} \leftarrow eq.(4) and eq.(9)$
    $X\_fetch.add\,(FT_{i\#j})$
    Calculate the task data processing time $PT_{i\#j} \leftarrow eq.(10)$
    $X\_process.add\,(PT_{i\#j})$
  end for
end for
return $FT_{i\#j}, PT_{i\#j}$

In Spark's default task scheduling mechanism, when the number of partitions exceeds the number of tasks running concurrently, the vCPU tasks are executed simultaneously. Only when the current task completes execution and there are available CPU cores, can the next task commence. Task execution demands various resources including CPU and network, and this demand is dynamic. Hence, task execution is divided into data fetching and data processing phases. Running other tasks at suitable times during task execution can mitigate resource contention and enhance resource utilization.

4.2. Task Scheduling Algorithm Based on Maximizing Resource Interleaving

According to the detailed information of the corresponding tasks in each node, the tasks in each node are first grouped according to different stages and sorted according to the execution time; then, since the stage with the longest execution time affects the completion time of the whole pipeline, it is prioritized and staggered according to the length of the completion time of the tasks in the historical data; finally, for other parallel stages, due to the existence of deadline, it is only necessary to ensure the completion of its execution before the completion of the longest stage, so its tasks can be scheduled at the right time to stagger the utilization of cluster resources.

The data fetching time $FT_{i\#j}$, data processing time $PT_{i\#j}$ of $Task_{i\#j}$ is obtained from Algorithm 1. During task execution, a large amount of network resources are needed in the data fetching phase, while a large amount of memory is needed for computation in the data processing phase. The goal of DRTS is to mitigate resource contention when executing tasks in the CPU-intensive phase, during which the cluster requires substantial network resources and disk I/O. In accordance with Spark's default task scheduling strategy, multiple tasks will be executed in parallel, and multiple tasks simultaneously multiple tasks compete for resources at the same time. At this point, Algorithm 2 schedules parallel tasks at optimal times, leveraging staggered time intervals to utilize cluster resources efficiently and reduce resource contention.

To achieve the objective of minimizing the job execution time, the following algorithm is designed. The algorithm can be divided into the following steps:

Algorithm 2 Delay-Aware Resource-Efficient Interleaved Task Scheduling Strategy

Input: Time maps X_{fetch}, X_{process} and the initial set of parallel stages S, the downstream shuffle S_i of S_k.

Output: Path and delay time set X.

Initialize: $X \leftarrow \{\}$ and the set of execution path P according to the job's DAG.
Sort the parallel S in descending order
for all S_i in S **do**
    Sort $T_{i\#j}$ in S_i in descending order, and group by target_node
    X.put $(i,$ new List$())$
end for
for all S_i in S **do**
    for all $different\_node\_task$ in S_i **do**
        for all $T_{i\#j}$ in all $different\_node$ **do**
            if isLongestStage(node.StageId) **then**
                Target machines execute tasks in an interleaved manner of long and short tasks firstly
            else
                Base_Time $= \left(S_{S_i} - S_{S_{i+1}}\right)/2$
                if $S_k \notin \varphi_m$ **then**
                    Base_Time $= (S_{S_k} - \max(S_{\text{parent}_{S_k}}))/2$
                    $S_i = S_i +$ Base_Time
                    for all $\hat{x_k} \in [0, T_{i\#j} - T_{i\#(j+1)}]$ **do**
                        if $PT_{i\#j} < T_{i\#(j+1)}$ **then**
                            $\hat{x_k} \leftarrow T_{i\#j} - T_{i\#(j+1)}$
                            $\hat{x_k} \leftarrow \hat{x_k} - \triangle x$
                        else
                            $\hat{x_k} \leftarrow FT_{i\#j}$
                            $\hat{x_k} \leftarrow \hat{x_k} \pm \triangle x$
                        end if
                        $x_k \leftarrow \hat{x_k} +$ Base_Time
                        X.get(i).add$(T_{i\#j}, x_k)$
                    end for
                end if
            end if
        end for
    end for
end for

(1) According to the parallelism of stage, it is partitioned according to different paths, and the partitioned paths are φ_m. Based on the allocation of parallelizable stage in each node, traversal is conducted to obtain the set of tasks in the node belonging to different partitioned paths.

(2) When the downstream shuffle of $S_i \notin \varphi_m$, the scheduling time of S_i is adjusted accordingly. This adjustment involves finding the downstream shuffle of S_i, denoted as S_k, by analyzing the DAG in the log file. Consequently, the delayed scheduling time interval of S_i is determined to be [0, S_{S_k} -max($S_{parent_{S_k}}$)]. where S_{S_k} refers to the start time of S_k, and max($S_{parent_{S_k}}$) refers to the maximal completion time in the parent stage of S_k as shown in Figure 3,and Stage5 belongs to this situation.

(3) Since the stage with the longest execution time significantly impacts the completion time of the entire job, it is prioritized for scheduling. Therefore, it is necessary to first sort φ_m in descending order.

(4) $Task_{i\#j}$ are grouped according to i in each target node according to the above results. They are then arranged in both descending and ascending time orders to enable interleaved execution.

(5) For a stage, if it is the longest among several parallel stages, each node prioritizes the execution of tasks assigned to that stage. Tasks of that stage are then executed according to the alternation of long and short time tasks.

(6) For the other stages that can be parallelized, since their overall execution time is not long, the start time of the execution of the back-ordered stages has little to do with the completion time of those stages. The adjusted completion time must not exceed the completion time of the first stage after sequencing. Therefore, the task execution of subsequent stages is initially delayed by a certain period of time, with the delay time initially set as the middle value of $S_i - S_{i+1}$.

(7) The execution of tasks corresponding to each stage on each node is performed in the order in (3), and the execution of tasks is performed in the order of the results in (4), and the execution of tasks that are at the back of the ordering is postponed at an appropriate time. But the delay time of the subsequent task must be within [0, $S_i - S_{i+1}$-$T_{i\#j}$]. The reason is that when the Task is delayed for too long, it may lead to a marginal overall completion time of the task, which is contrary to the DRTS policy of minimizing the job completion time.

(8) Ensure that after the execution of the task with the longest stage, the data fetching time $FT_{i\#j}$ in the other tasks is used as a cumulative count of the initial value. This ensures that the delay time of each task is different, and constant feedback is provided to adjust this value.

(9) Divide the delay time of the task into blocks of time Δx (e.g., each block has an interval of 200 ms). When $PT_{i\#j} < T_{i\#(j+1)}$, $\hat{x_k}$ iterates over its upper and lower ranges [0, $T_{i\#j} - T_{i\#(j+1)}$]; When $PT_{i\#j} >= T_{i\#(j+1)}$,$\hat{x_k}$ iterates over its upper and lower ranges [$FT_{i\#j}$, $T_{i\#j} - T_{i\#(j+1)}$]. There is a candidate task scheduling time $\hat{x_k}$ in each iteration . Where $T_{i\#j}$ is the previous task of $T_{i\#j+1}$ at the same target machine.

(10) After continuous feedback and adjustments based on the results of the last execution, DRTS finally determines the optimal value of the delayed task execution time x_k to greedily minimize the execution time of parallel tasks.

DRTS schedules parallel tasks in a pipelined manner, which effectively reduces resource interleaving during task execution and improves the resource utilization of the cluster.

5. Experiments

In this section, a comprehensive evaluation of the DRTS strategy is presented. Three workloads ConnectedComponents, CosineSimilarity and TriangleCount are used as benchmarks to evaluate the performance. The evaluation results are as follows:

(1) Accelerated the workload of the benchmark suite, reducing the average job completion time by 7.51% to 15.64%.

(2) By utilizing cluster resources in a staggered manner, the CPU utilization and network utilization of the cluster are improved by 5.83% and 10.38%, respectively.

5.1. Experiment Setting

Cluster Configurations In this section, to validate the performance of DRTS, we conduct extensive experiments on a Spark 2.4.0-based cluster with one master node and six worker nodes. We evaluated the performance of task scheduling using three representative DAG-style data analysis workloads as benchmarks. The cluster configuration is shown in Tab.1. The deployment method is Spark on Standalone, and Spark's parameters are configured as shown in Tab. 2.

Table 1. Configuration of Cluster

Type	Configuration
Number of Nodes	1 master,6 workers
CPU Number	4
RAM	32G
Hard disk	500G
Environment	Centos7.0, Spark2.4.0, Hadoop 2.6.0
Digital meter	2500W, 10A
JDK	JDK 1.8

Table 2. Configuration of Spark parameters

Type	Configuration
Executor cores	2
Executor memory	4G
Executor number	12

Workload Detail Description In the cluster, we chose three representative Spark benchmark workloads: ConnectedComponents, CosineSimilarity, and TriangleCount, where ConnectedComponents and TriangleCount are from Spark GraphX. There are 5 stages and 11 stages, respectively. CosineSimilarity comes from Spark MLlib and has five stages. Specifically, the experimental ConnectedComponents application has 11GB of synthetic data, the CosineSimilarity application uses 34GB of synthetic data, and TriangleCount uses synthetic data from 1 million users and 20 million followers. The workload specifications are summarized in Tab. 3.

Base Line The scheduling algorithms we compared in these experiments are as follows:
Spark: the default scheduling strategy FIFO in Spark.

Table 3. Workload

Wordload	Specification
ConnectedComponents	11GB synthetic input data
CosineSimilarity	34GB synthetic input data
TriangleCount	1 million users and 20 million followers of synthetic input data

DelayStage[16]: it is a stage delay scheduler for big data parallel computing frameworks (e.g. Spark). It optimizes resource utilization by overlapping cluster resources during parallel phases to minimize completion time. Additionally, it schedules phase execution in a pipelined manner to maximize the performance benefits of resource interleaving.

5.2. Performance Results

In order to validate the performance of DRTS, a series of experiments were conducted on different datasets using the benchmarks in Tab .3. Each benchmark was tested 15 times and the final results were averaged over the tests. This experiment compares the DRTS scheduling strategy with Spark's default scheduling strategies FIFO and DelayStage scheduling strategies.

Job Execution Time We first compare the job execution time of different schedulers. As shown in Fig. 5, we can see that DRTS shortens the job execution time by 8.01% to 14.46% compared to Spark's default scheduling strategy. DRTS is a delayed scheduler operating at the task level across different nodes. When tasks are executed concurrently, it can lead to resource contention issues. By scheduling parallel tasks at the right time, DRTS effectively interleaves the use of cluster resources, thereby enhancing overall resource utilization. Moreover, the task execution is divided into two phases, when the data is being acquired, scheduling a CPU-intensive task at the right time to greedily realize the resource interleaving between tasks can also reduce the total execution time of parallel tasks.

Additionally, the performance of DRTS is also enhanced (3.18% - 6.48%) compared with DelayStage due to the following reasons: DRTS schedules parallel tasks assigned to each node at optimal times, allowing tasks in the parallel stage to be delayed based on the state of the target machine. On the other hand, DelayStage only delays the overall stage without considering that tasks within the stage are assigned to different machines, resulting in uniform delays across all tasks. DRTS considers more factors and offers finer granularity compared to DelayStage.

Upon closer examination of the job completion times depicted in Fig. 5, DRTS demonstrates superior performance compared to Spark's default FIFO scheduling approach. The figure illustrates the job completion times for the three benchmarks: ConnectedComponents, CosineSimilarity, and TriangleCount. Specifically, DRTS improves performance by 8.01% for ConnectedComponents, reduces job completion time by 13.51% for CosineSimilarity, and enhances performance by 14.46% for TriangleCount. This improvement can be attributed to the nature of FIFO scheduling, which prioritizes earlier submitted tasks for execution, causing subsequent tasks to wait until prior ones are completed. When early

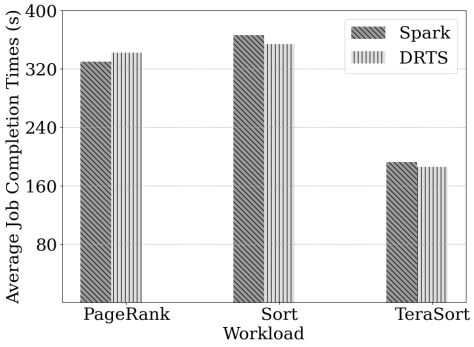


Fig. 5. Comparison of Spark and DRTS across different workloads, including PageRank, Sort, and TeraSort

tasks do not utilize resources efficiently, it may result in overall low resource utilization. By comparing these three benchmarks with DelayStage, DRTS demonstrates performance improvements of 3.18%, 6.05%, and 6.48%, respectively. This is because DelayStage only delays the stage without considering granularity such as different target machines or tasks on different target machines. In contrast, DRTS operates at the task level, sorting tasks within stages with longer execution times based on target machine execution times. It prioritizes stages with the longest execution times for scheduling while ensuring that execution scheduling delays are applied to different target machines for different tasks based on varied execution conditions. This approach alternately utilizes cluster resources to avoid resource contention, ultimately achieving the goal of reducing job execution time.

In the Spark framework, parallelism refers to the number of data blocks processed simultaneously during task execution, and it's a manually configured parameter. Excessive parallelism may over-consume cluster resources or result in frequent task startups, increasing overhead. Conversely, insufficient parallelism may underutilize cluster resources, prolonging task execution. Configuring parallelism appropriately poses challenges for developers, as improper configurations can increase job completion time, impacting cluster performance.

DRTS can mitigate performance degradation resulting from suboptimal parallelism configurations. To observe the impact of different parallelism levels on DRTS, we conducted a series of experiments. The results, depicted in Fig. 6, demonstrate that DRTS effectively reduces job completion time across varying degrees of parallelism.

Resource Utilization Effectiveness To assess whether our DRTS policy enhances resource utilization, we conducted further observations on CPU utilization and network throughput of a worker node while executing various workloads. As depicted in Fig. 7. and 8, the DRTS policy optimizes DelayStage by efficiently utilizing idle time during stage execution, thereby notably enhancing CPU utilization and network throughput. Compared to Spark's default scheduler, DRTS executes tasks in a two-stage resource interleaving manner, effectively improving system resource utilization. On average, CPU utilization improves by 4.43% to 8.77%, and network throughput utilization improves

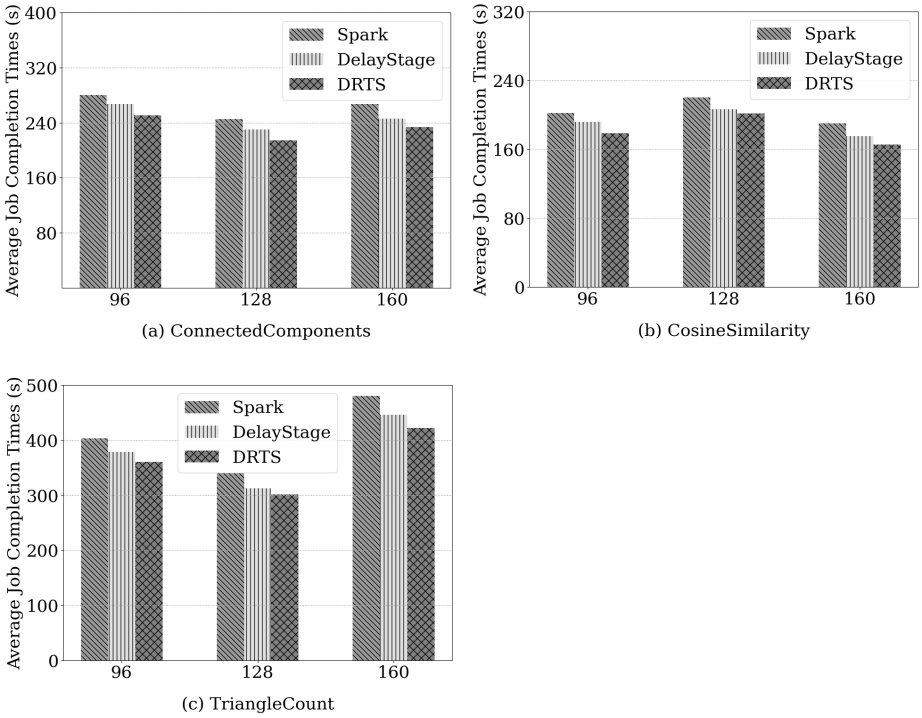


Fig. 6. Comparison of Spark, DelayStage, and DPRS at different levels of parallelism

by 5.19% to 8.84% with DRTS. Additionally, DRTS enhances resource utilization compared to DelayStage, with CPU utilization improving by 8.42% to 14.64% and network throughput improving by 10.43% to 12.98%. This improvement stems from DRTS's finer granularity in scheduling tasks to the same target machine at optimal times, whereas DelayStage simply delays stage scheduling, potentially resulting in idle resources on some machines. By scheduling CPU-intensive tasks at opportune moments during data acquisition, DRTS effectively fills resource gaps during CPU idle periods, thereby enhancing job operational efficiency through optimal resource utilization.

Table 4. Average of network throughput (MB/s)

	Spark	DelayStage	DRTS
ConnectedComponents	18.88	19.86	20.85
CosineSimilarity	136.21	148.25	152.14
TriangleCount	23.66	25.32	26.73

Further, we compute the average values of network throughput and CPU utilization of a work node while executing these four workloads, as summarized in Tab. 4. and 5. respectively.

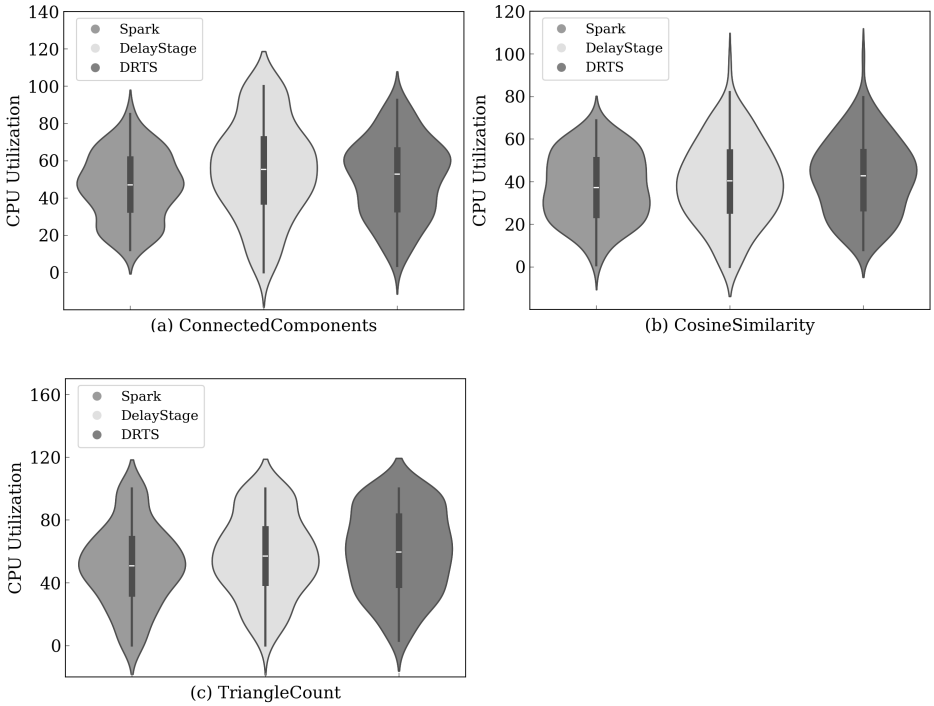


Fig. 7. CPU utilization under the three workloads: ConnectedComponents, CosineSimilarity, and TriangleCount

Clearly, DRTS achieves higher and more stable CPU utilization and network throughput compared to Spark's default scheduling strategy. In more detail, DRTS improves network utilization by 11.42% and network throughput by 11.7% over Spark, and also improves CPU utilization and network throughput by 6.33% and 7.02%, respectively, compared to DelayStage.

6. Conclusion

In this paper, we address the task scheduling problem within job execution scenarios. Underutilized assigned tasks can significantly prolong job execution times, thereby impacting cluster performance. To mitigate this issue, we propose a strategy focusing on

Table 5. Average CPU utilization (%)

	Spark	DelayStage	DRTS
ConnectedComponents	46.22	48.89	50.11
CosineSimilarity	37.51	39.17	41.71
TriangleCount	50.88	55.34	58.33

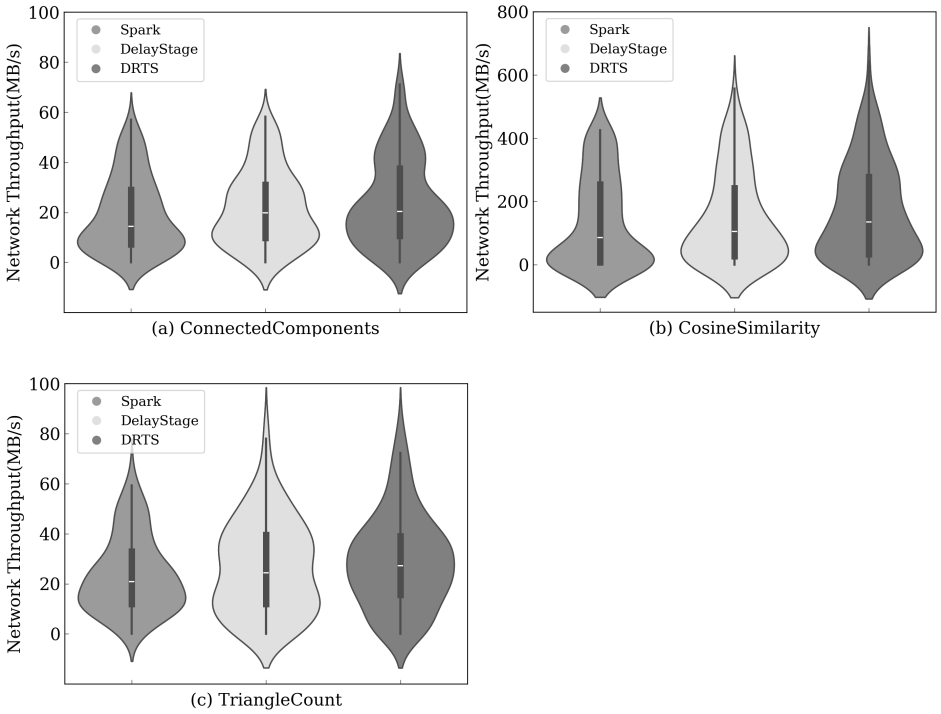


Fig. 8. Network throughput under the three workloads: ConnectedComponents, CosineSimilarity, and TriangleCount

cross-utilization of resources to minimize job completion times. Firstly, we provide a comprehensive theoretical analysis of the task scheduling problem and devise an algorithm to compute task execution times for both data acquisition and processing phases. Subsequently, we introduce a task scheduling algorithm based on resource polling, employing a delayed scheduling approach at the task level across different nodes. For stages with extended execution times, tasks are scheduled on target machines in an alternating pattern of long and short task durations. For other stages, appropriate scheduling algorithms are employed to ensure tasks are executed at optimal times, thereby facilitating cross-resource utilization within the cluster. Finally, we conduct extensive experiments across three benchmarks using various datasets. Our experimental results demonstrate that the proposed DRTS approach effectively harnesses cluster resources and reduces job completion times.

Acknowledgments. This work was supported by a National Natural Science Foundation of China, Major Research Program, 92367302, Integrated System and Validation of Industrial Internet Based on the New Architecture of AllFactor On-Demand Collaborative Interconnection.

References

1. Apache flink — stateful computations over data streams. https://flink.apache.org.

2. Apache spark - unified engine for large-scale data analytics. https://spark.apache.org.
3. Apache storm is a free and open source distributed realtime computation system. https://storm.apache.org.
4. Dhawalia p, kailasam s, janakiram d. chisel: A resource savvy approach for handling skew in mapreduce applications[c]//2013 ieee sixth international conference on cloud computing. ieee, 2013: 652-660.
5. Duan y, wang n, wu j. accelerating dag-style job execution via optimizing resource pipeline scheduling[j]. journal of computer science and technology, 2022, 37(4): 852-868.
6. Fu z, tang z, yang l, et al. an optimal locality-aware task scheduling algorithm based on bipartite graph modelling for spark applications[j]. ieee transactions on parallel and distributed systems, 2020, 31(10): 2406-2420.
7. Gu r, tang y, tian c, et al. improving execution concurrency of large-scale matrix multiplication on distributed data-parallel platforms[j]. ieee transactions on parallel and distributed systems, 2017, 28(9): 2539-2552.
8. "hadoop," 2021. [online]. https://hadoop.apache.org.
9. He x, shenoy p. firebird: Network-aware task scheduling for spark using sdns[c]//2016 25th international conference on computer communication and networks (icccn). ieee, 2016: 1-10.
10. Hu z, li b, qin z, et al. job scheduling without prior information in big data processing systems[c]//2017 ieee 37th international conference on distributed computing systems (icdcs). ieee, 2017: 572-582.
11. Hu z, li d. improved heuristic job scheduling method to enhance throughput for big data analytics[j]. tsinghua science and technology, 2021, 27(2): 344-357.
12. Jiang j, ma s, li b, et al. symbiosis: Network-aware task scheduling in data-parallel frameworks[c]//ieee infocom 2016-the 35th annual ieee international conference on computer communications. ieee, 2016: 1-9.
13. Li x, ren f, yang b. modeling and analyzing the performance of high-speed packet i/o[j]. tsinghua science and technology, 2021, 26(4): 426-439.
14. Lu s x, zhao m, li c, et al. time-aware data partition optimization and heterogeneous task scheduling strategies in spark clusters[j]. the computer journal, 2023: bxad017.
15. Pan f, xiong j, shen y, et al. h-scheduler: Storage-aware task scheduling for heterogeneous-storage spark clusters[c]//2018 ieee 24th international conference on parallel and distributed systems (icpads). ieee, 2018: 1-9.
16. Shao w, xu f, chen l, et al. stage delay scheduling: Speeding up dag-style data analytics jobs with resource interleaving[c]//proceedings of the 48th international conference on parallel processing. 2019: 1-11.
17. Tang z, xiao z, yang l, et al. a network load perception based task scheduler for parallel distributed data processing systems[j]. ieee transactions on cloud computing, 2021, 11(2): 1352-1364.
18. Tang z, zeng a, zhang x, et al. dynamic memory-aware scheduling in spark computing environment[j]. journal of parallel and distributed computing, 2020, 141: 10-22.
19. Wang j, gu h, yu j, et al. research on virtual machine consolidation strategy based on combined prediction and energy-aware in cloud computing platform[j]. journal of cloud computing, 2022, 11(1): 50.
20. Wang j, yu j, song y, et al. an efficient energy-aware and service quality improvement strategy applied in cloud computing[j]. cluster computing, 2023, 26(6): 4031-4049.
21. Wang j, yu j, zhai r, et al. gmpr: a two-phase heuristic algorithm for virtual machine placement in large-scale cloud data centers[j]. ieee systems journal, 2022, 17(1): 1419-1430.
22. Xu g, xu c z, jiang s. prophet: Scheduling executors with time-varying resource demands on data-parallel computation frameworks[c]//2016 ieee international conference on autonomic computing (icac). ieee, 2016: 45-54.

23. Xu y, liu l, ding z. dag-aware joint task scheduling and cache management in spark clusters[c]//2020 ieee international parallel and distributed processing symposium (ipdps). ieee, 2020: 378-387.
24. Zaharia m, borthakur d, sen sarma j, et al. delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling[c]//proceedings of the 5th european conference on computer systems. 2010: 265-278.
25. Zhang x, li z, liu g, et al. a spark scheduling strategy for heterogeneous cluster[j]. computers, materials continua, 2018, 55(3).
26. Dean J, G.S.: Mapreduce: simplified data processing on large clusters. In: Proceedings of the 1st International Conference on Preparing ComSIS Articles. pp. 107–113. Communications of the ACM (2008)
27. Islam M T, Karunasekera S, B.R.: Performance and cost-efficient spark job scheduling based on deep reinforcement learning in cloud computing environments. In: IEEE Transactions on Parallel and Distributed Systems. pp. 107–113. Communications of the ACM (2021)

Yanhao Zhang is currently studying for a master's degree. In 2018, he received a bachelor's degree in Software Engineering from Shangqiu Normal University. Her research interests include parallel and distributed computing and Spark.

Congyang Wang is currently studying for his master's degree. He received his Bachelor of Engineering degree from Henan University in 2018. His research interests include big data and parallel computing.

Xin He is a professor at the School of Software at Henan University. He received his Master of Science degree in Applied Mathematics from Henan University in 2005. In 2011, he received his Ph.D. degree in Computer System Structure from Xi 'an Jiaotong University. His research interests include mobile computing, cloud computing and big data processing, and computer networking.

Junyang Yu is an associate professor at the School of Software at Henan University. He received his master's degree from the School of Computer and Information Engineering, Henan University in 2007. In 2016, he received his PhD in Software Engineering from Central South University. His research interests include cloud computing, big data, and distributed systems.

Rui Zhai received his Ph.D. degree in computer science from University of Chinese Academy of Sciences in 2015. Currently, he is a lecturer at the Software School of Henan University. His research interests include machine learning, federated learning, and GNN.

Yalin Song is an associate professor at the School of Software, Henan University, and holds a Ph.D. degree from Tongji University. His research interests include cognitive computing and computer vision.

Received: August 31, 2024; Accepted: January 20, 2025.

Computer Science and Information Systems 22(3):859–879 https://doi.org/10.2298/CSIS240930019Z

Identification and Detection of Illegal Gambling Websites and Analysis of User Behavior

Zhimin Zhang[1], Dezhi Han[1], Songyang Wu[2,⋆], Wenqi Sun[2], and Shuxin Shi[1]

[1] College of Information Engineering, Shanghai Maritime University,
201306 Shanghai, China
zhangzhimin@stu.shmtu.edu.cn
dzhan@shmtu.edu.cn
shishuxin@stu.shmtu.edu.cn
[2] Network Security Center, The Third Research Institute of the Ministry of Public Security,
200031 Shanghai, China
wusongyang@stars.org.cn
sunwenqi@gass.ac.cn

Abstract. Illegal gambling websites use advanced technology to evade regulations, posing cybersecurity challenges. To address this, we propose a machine learning method to identify these sites and analyze user behavior accurately. The method extracts key data from post messages in a real-world network environment, generating word vectors via Word2Vec with TF-IDF, which are then downscaled and feature-extracted using a Stacked Denoising Auto Encoder (SDAE). Next, this paper uses Agglomerative Clustering, improved through a combination of distance caching and heap optimization, to initially cluster post-template websites of the same type by clustering them into the same cluster. Then, multiple algorithms are integrated within each website cluster to cluster users' different operational behaviors into different clusters based on the cosine similarity consensus function voting secondary clustering. Results show improved detection of illegal gambling sites and classification of user activities, offering new insights for combating these sites.

Keywords: Gambling websites, post messages, feature extraction, illegal website identification, cluster analysis.

1. Introduction

At present, the rapid development of gambling websites has brought great harm to society. These websites not only jeopardize the financial security of individuals but also undermine social order and contribute to the spread of criminal activities. Due to their hidden and transnational nature, gambling websites are complicated to regulate and combat. Gambling websites usually utilize advanced network technologies to hide their true intentions through sophisticated encryption and camouflage means, making it difficult for traditional regulatory means to be effective. By attracting users to engage in gambling activities, these websites reap substantial illegal profits and legalize these profits through money laundering and other means. Users who participate in illegal gambling often face significant financial risks and may even lose all their money as a result of indulgence in

⋆ Corresponding author

gambling, bringing severe negative impacts on families and society. In addition, illegal gambling may be closely linked to other criminal activities such as fraud, extortion, and violence, further contributing to social instability[1]. With technological advances, these websites have become increasingly covert and highly adaptable, making it difficult for traditional regulatory measures to cope. Users' operational behavior on these websites not only affects their financial and psychological health but also threatens the security of the entire online environment. In the face of these challenges, there is an urgent need to develop and optimize new identification and detection methods. Traditional detection methods, however, often struggle to cope with the complexity of illegal gambling websites due to the constant changes in technology and strategies. Therefore, it is crucial to effectively combat illegal gambling activities by exploring more accurate and efficient identification and detection methods using advanced data analytics and machine learning techniques to gain insights into users' operational behaviors on illegal gambling websites and reveal their potential commonalities and differences. In this paper, we propose a machine learning-based gambling website identification method that can accuriately identify and detect illegal gambling websites and analyze the different operating behaviors of users in different types of gambling websites.The main research and contributions of this paper are summarized below:

(1)By crawling post messages and extracting critical information in a real-world network environment, critical information such as cookie parameters, request body parameters, request line parameters, or keywords are grouped into two datasets, DataSet1 and DataSet2, through WebName and Host links. The critical information is reduced and features extracted using the Term Frequency-Inverse Document Frequency(TF-IDF) weighted Word2Vec method as well as through an improved Stacked Denoising Auto Encoder (SDAE) for dimensionality reduction and feature extraction to obtain stable features of critical information in post messages.

(2)The DataSet1, i.e., Cookie parameter is initially clustered with the same type of websites using a cohesive Agglomerative Clustering algorithm improved by combining distance caching and heap optimization, and the identification and detection of different types of illegal gambling websites is achieved through the stable features of different types of post message template websites obtained. The experimental results show that the adopted method performs well in terms of accuracy and stability and can successfully realize the identification and detection of illegal gambling websites.

(3)Further clustering of DataSet2, i.e., request body parameters and keywords in the request line, within each website clustering cluster, K-means, DBSCAN, Agglomerative Clustering, OPTICS, and Gaussian Mixture Models clustering algorithms are selected, and by evaluating the performance of each method as well as its adaptability to the present experimental By evaluating the performance of each method and its adaptability to the data of this experiment, the consensus function based on cosine similarity votes for integrated clustering to further classify the different behaviors of the same type of websites.

The remainder of the paper is organized as follows: in Section. 2, the paper reviews current research advances and technical approaches in the field of illegal gambling website detection. Section. 3 describes the overall architecture of this paper and the process of dataset production, including data sources, data cleaning, and feature extraction methods, and describes in detail the clustering methods and algorithms used, specifically including preliminary cohesive clustering of the same type of post message template websites using

distance caching combined with heap optimization for improved Agglomerative Clustering, and multiple algorithms integrated clustering of user behaviors based on the voting of consensus functions.Section. 4 shows the experimental results and analysis to evaluate the performance of the methods proposed for identifying and detecting illegal gambling websites in comparison with different models, focusing on profile coefficients, accuracy, and recall and showing the overall results obtained in this paper.Section. 5 summarizes the main contributions of this paper and proposes future research directions, suggesting further optimization of the model to improve its effectiveness in practical applications.

2. Related work

In recent years, many research results have been published on the identification and detection of illegal gambling websites and their user behavior. These researches mainly focus on network traffic analysis, deep learning, multi-view clustering, text analysis, behavioral pattern detection, etc., which promote developing and applying technologies in this field.

In network traffic analysis, Kong et al. (2020) proposed a hybrid model based on convolutional neural network (CNN) and long short-term memory network (LSTM) for network traffic classification. This method combines CNN's extraction of spatial features and LSTM's processing of time-series data, and especially performs well in encrypted traffic detection, effectively distinguishing between legitimate and illegitimate traffic[2].Mu et al. (2022) further developed this idea by proposing a hybrid intrusion detection model based on CNN-LSTM and attention mechanism, which combines the attention mechanism with an enhanced feature extraction capability, making the model high accuracy and robustness when dealing with complex intrusion detection tasks[3].Alshingiti et al. (2022), on the other hand, developed a phishing website detection system based on CNN, LSTM, and LSTM-CNN, which efficiently categorizes phishing website URLs with an accuracy of more than 99% through deep learning[4].

In terms of multi-view clustering, Alnemari and Alshammari (2023) proposed a feature extraction based phishing domain name classification system using algorithms such as Support Vector Machines (SVMs) and decision trees[5]. The method significantly improves the detection accuracy of phishing domain names by analyzing domain name features such as length and character structure.Chen et al. (2023), on the other hand, explored the application of graph convolutional networks (GCN) and feature fusion techniques in multiview learning, which can efficiently extract multiview features and improve the clustering accuracy[6].Huang et al. (2023) proposed a depth-weighted multiview clustering method based on self-supervised graphical attention networks. weighted multiview clustering method, which especially performs well when dealing with complex data structures[7].

In the field of text analytics, Chen et al. (2020) proposed an automated detection system that combines visual and textual content for identifying pornographic and gambling websites, extracting web page HTML text features through Doc2Vec and combining them with visual content for categorization, which ultimately achieves more than 99% accuracy[8].Wang et al. (2022) proposed a multimodal data fusion framework that improves the accuracy of gambling website detection by combining text features extracted from images and OCR and fusing multiple data using a self-attention mechanism improves the accuracy of gambling website detection[9].Sun et al. (2023) apply domain cer-

tificate information and textual analysis to enhance gambling domain name recognition, achieving improved accuracy in domain classification[10].

In terms of user behavior clustering, Singh et al. (2021) proposed a clustering-based e-commerce webpage recommendation system, which improves the personalization and accuracy of recommendation by analyzing user behavior data[11].Li et al. (2021) combined K-mean clustering and hybrid particle swarm optimization algorithms to segment e-commerce users and optimize the effect of precision marketing[12].Liu (2022) used machine learning to to develop a personalized recommendation system based on user behavioral data, which significantly improves recommendation accuracy and user experience[13].

These research results demonstrate the application of various technical means in identifying gambling websites and their user behaviors, covering a wide range of areas from network traffic analysis to behavioral pattern detection and providing valuable references for improving detection accuracy and effectiveness. However, illegal gambling websites are becoming more and more covert and adaptable, and the technical means continue to evolve. Therefore, exploring new techniques, such as machine learning and data mining, to effectively detect and cluster analyze gambling websites has become a hot and challenging topic in current research. This paper aims to explore the characteristics and laws of illegal gambling websites by effectively clustering the POST messages of illegal gambling websites and analyzing them in depth to provide strong technical support for the regulatory authorities and help them better identify and respond to illegal behaviors.

3. Methodology

3.1. Overall architecture

In this paper, by crawling illegal gambling website post messages in the real-world network environment for data cleaning and extracting critical information, cookie parameters are divided into one group, request body parameters and keywords in the request line are another group, which are compiled into two datasets through WebName and Host links. Respectively, we use the Word2Vec method with TF-IDF weighting and SDAE for feature extraction to obtain stable features of critical information in post messages. Agglomerative Clustering, an improved cohesive hierarchical clustering algorithm combining distance caching and heap optimization, was used to initially cluster the same type of post-message template websites for DataSet1, and further Clustering was performed for DataSet2 within each cluster through OPTICS, Gaussian Mixture Models, Agglomerative Clustering, K-means, and DBSCAN multiple algorithms based on the cosine similarity of the consensus function voting integrated clustering, through the obtained different types of websites and different behaviors of the stability of the characteristics of the different types of illegal gambling websites to achieve different types of illegal gambling websites identification and detection, and further to obtain the different behaviors of the user in different types of illegal gambling websites. The overall architecture of this paper is depicted in Fig. 1.

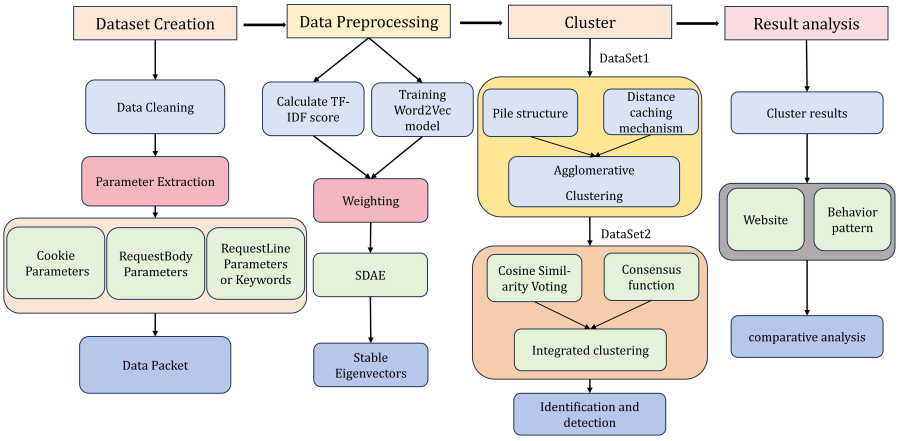


Fig. 1. Overall Architecture Diagram

3.2. Dataset production and data preprocessing

Dataset production Based on the collected website information, this paper simulates the gambling operation behaviors of real users at illegal gambling websites in a real-world network environment. For these gambling websites, the usual operation behaviors include registering, logging in, binding bank cards, recharging, placing bets, consulting customer service, and participating in website activities. In this paper, we collect the POST request messages generated by each behavior and mark the collected data with unique tags for subsequent analysis.POST request message is a method used in the HTTP protocol to send data to the server. Unlike GET requests, POST requests include the data in the request body rather than passing it through a URL. A typical POST request message consists of three parts: the request line (which specifies the method, destination URL, and HTTP protocol version), the request header (which contains information describing the request and the client, such as Host, Content-Length, User-Agent, Content-Type, etc.), and the request body (which contains the data to be sent to the server and is typically used to submitting form data, uploading files, etc.).The POST request message is essential for communication between illegal gambling websites and users. It contains rich request data, such as form information submitted by users, files, JSON data, etc. The gambling user's specific operation and behavioral mode can be identified by analyzing the parameters and data content in the POST request, such as the user's betting records and query records on the illegal gambling website. These data are used for feature extraction. Key features that reflect user behavior and website characteristics are extracted by analyzing the parameters and contents in different requests to provide a basis for subsequent analysis. Fig. 2 shows the original format of a particular post message collected.

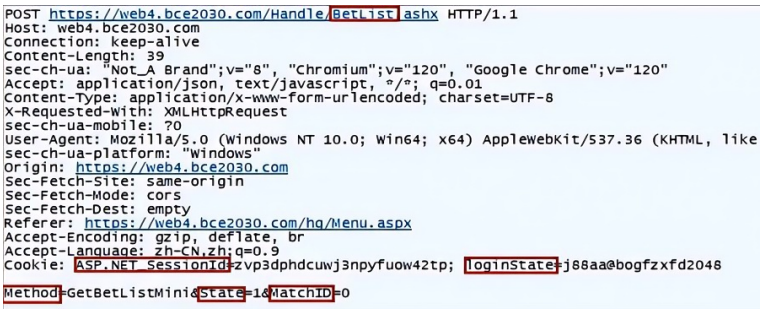


Fig. 2. Post message format and information of each part

In a POST message, the cookie parameter is transmitted in the header of the HTTP message, which may contain the user's login information, identifiers for tracking user activity, session tokens, and so on, which can help identify user behavior patterns or find the mechanisms used by these sites to track users, such as the above figure labeled 'ASP.NET_SessionId'. The request body of a POST request is the data contained in the HTTP request and is usually used to submit a form or upload data. In illegal gambling sites, the parameters in the POST request body may include sensitive data such as the user's betting amount, selected stakes, and payment information. These parameters can reveal how the user interacts with the gambling platform, and analyzing these parameters can help to understand the user's behavior, such as the 'Method', 'State' and 'MatchID' labeled in the figure above. Keywords in the request line usually refer to specific paths in the URL or query string parameters that represent actions performed by the user or application. The keywords in the request line can indicate specific functions accessed by the user, such as logging in, placing bets, withdrawing money, and so on. By analyzing these keywords, it is possible to determine the intent of the user's action and identify the specific interaction patterns of a gambling website, such as the 'BetList' labeled in the figure above.

Among all the collected POST request messages, this paper tags each message with the WebName, Host, and behavior. It extracts critical data, including cookie parameters, parameters in the request body, and representative keywords in the request line. Since in the real-world network environment, the exact behavior of the same website may generate multiple identical POST request messages, in order to ensure the clarity and rationality of the data, these messages are first de-duplicated, and the garbled parameter problem generated by some post request messages is handled specially. According to the characteristics of the post message and the data form characteristics of this dataset, this paper will be data grouping[14], the same site under the cookie parameters merged and de-emphasized data for a group, all the different behaviors of the site parameters in the request body and the request line with a representative of the keyword for another group, the two sets of data according to the WebName and Host links, made into illegal gambling sites The two sets of data are linked according to WebName and Host to create two datasets of illegal gambling websites, DataSet1 and DataSet2, which are convenient for subsequent work.

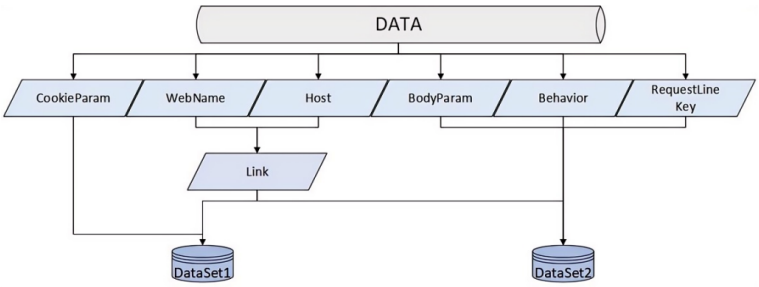


Fig. 3. Schematic representation of each part of the dataset

Word2vec feature extraction with TF-IDF weighting TF-IDF is a weighting technique commonly used in text analysis to measure the importance of a term in a document. The core idea is that the combination of the Term Frequency (TF) of a term in a particular document and the rarity of its occurrence in all documents (Inverse Document Frequency (IDF)) can effectively differentiate representative words from common irrelevant words.Word2Vec is a word embedding model that maps words to a Word2Vec, which maps words into a high-dimensional vector space. It can learn the semantic similarity between words through training.Word2Vec learns the vector representations of words in two ways: CBOW (Continuous Bag of Words) and the Skip-gram model. In this paper, we mainly use the Skip-gram model, and the goal of the training is to capture the co-occurrence probability of words between the target word and its context by maximizing the co-occurrence probability of words. The training objective is to capture the semantic relationship between words by maximizing the co-occurrence probability between the target word and its context words. In order to obtain the weighted word vectors, the Word2Vec vectors of each word are weighted and averaged according to their TF-IDF values[15],as shown in Eqs. 1, 2, 3.

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t)\,. \tag{1}$$

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c<j<c, j\neq 0} \log p(w_{t+j} \mid w_t)\,. \tag{2}$$

$$v_d = \sum_{t\in d} \text{TF-IDF}(t,d)\cdot v_t\,. \tag{3}$$

Eqs. 1 represents the TF-IDF value of the word in the document, $\text{TF}(t,d)$ is the weight of the term t in the document d, and $\text{IDF}(t)$ is the inverse document frequency of the term t. The co-occurrence probability of the target word in its context is maximized by Eqs. 2, where w_t is the target word, w_{t+j} is the target word's contextual word, and c is the size of the contextual window; and Eqs. 3 represents the final word vector of the word, v_t is the Word2Vec vector of the word t, and v_d is the TF-IDF-weighted Word2Vec word vector of the document.

In addition, to eliminate the effects between different feature magnitudes and transform the data to the same scale, the two parts of the data are normalized. The normal-

ization step ensures that the features are within the same magnitude range, improving the model's performance and the comparability of the results. Through the above method, statistical and semantic features can be comprehensively utilized to extract feature vectors with high differentiation and representativeness so that data analysis and model training can be carried out more effectively[16].

Obtaining stable features by SDAE dimensionality reduction In this paper, refer to Xin et al. to add the attention mechanism to the model[17], and weight the word vectors obtained from the Word2vec model with TF-IDF through the SDAE model, so as to further obtain the stable features of the above data.The SDAE model is initialized with specified input dimensions and his. Eden layer dimensions and the Leaky ReLU activation function are used. The SDAE model is trained by performing forward propagation, loss computation, backpropagation, and weight updating through the training function. The tensor is moved from the GPU to the CPU and converted to a Numpy array to obtain more stable and representative feature vectors[18]. The SDAE model self-encoder consists of two parts: encoder and decoder, and the specific implementation process is shown in Eqs. 4, 5, 6.

$$h^{(k)} = f(W_e^{(k)} h^{(k-1)} + b_e^{(k)}) . \tag{4}$$

$$z^{(k)} = g(W_d^{(k)} h^{(k)} + b_d^{(k)}) . \tag{5}$$

$$L(x, z) = \frac{1}{n} \sum_{i=1}^{n} (x_i - z_i)^2 . \tag{6}$$

Eqs. 4 computes the implied representation h of the input data compressed by the encoder, W_e is the weight matrix of the encoder, b_e is the bias vector of the encoder, and f is the Leaky ReLU activation function; Eqs. 5 computes the reconstructed z of the decoder that returns the implied representation h back to the original input, W_d is the weight matrix of the decoder, b_d is the bias vector of the decoder, and g is the activation function. The SDAE measures the difference between the reconstructed output z and the original input x via the loss function $L(x, z)$ in Eqs. 6.

SDAE stacks multiple denoising self-encoders layer by layer to build a deep network. After the layer-by-layer pre-training, all layers are connected and fine-tuned using labeled data to optimize the whole network[19]. In this paper, through the SDAE model, the noise in the dataset is effectively removed to obtain a purer feature representation. The complex features of the data are captured through a multilayer nonlinear transformation, and both parts of the data are downscaled to one-tenth of the original, respectively, while retaining the essential features to reduce the computational overhead.

3.3. Clustering methods

Website clustering methods Agglomerative Clustering is a bottom-up clustering method. The method starts by treating each data point as a separate cluster and then gradually

merges the most similar clusters until all data points have been merged into a single cluster or a predetermined number of clusters has been reached. The traditional cohesive hierarchical clustering algorithm needs to recalculate the distance between clusters each time the clusters are merged, and this process has a significant computational overhead, which makes the algorithm inefficient, especially when the dataset size is large.Therefore, several improvements to the algorithm are proposed, as illustrated in Fig. 4.

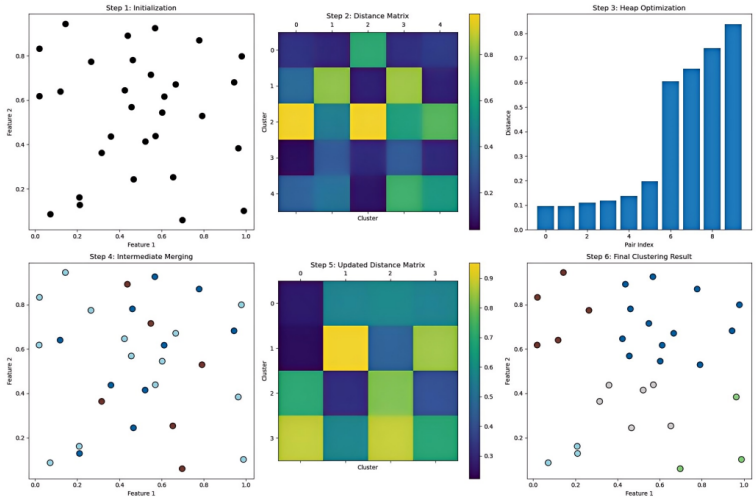


Fig. 4. Implementation of Improved Agglomerative Clustering Algorithm Combining Distance Caching and Heap Optimization

In order to improve the efficiency of the algorithm, this paper introduces a heap structure to store the distances between clusters based on traditional Agglomerative Clustering. Heap structure is an efficient data structure, and this paper uses the minimal heap to store the distance between clusters. During each cluster merging process, the heap structure can quickly retrieve the cluster pair with the smallest distance with a complexity of O(1) and only needs to update the distance associated with the merged clusters after the merger, thus avoiding recalculating the distances between all clusters. The complexity of updating the heap is (logN) (where N is the number of clusters), which greatly reduces the computation time. The introduction of the heap structure makes the merging clusters more efficient each time. In addition, a distance caching mechanism is introduced in this paper to reduce further the repetitive distance computation in the cluster merging process. In traditional methods, the distances between the new cluster and other clusters need to be recalculated after each cluster merge, and these distances may be calculated multiple times in subsequent iterations. To avoid this problem, this paper uses a distance caching mechanism to store the computed inter-cluster distances in a cache. Each time the distance needs to be calculated, it first checks whether the corresponding distance value already exists in the cache; if it does, the cached value is used directly; if not, it is calculated and stored. Through this mechanism, the overhead of repeated computation is greatly reduced, espe-

cially when dealing with large-scale datasets, which can significantly improve the overall efficiency of the algorithm.

This paper combines a further distance caching mechanism with heap optimization; firstly, the distances between all cluster pairs are computed in the initialization phase and stored in the distance cache. After each cluster merger, the new cluster pairs' distances are updated. Then, a minimal heap is used to store the initial clusters and their distance values, and the cluster pairs with the smallest distances are extracted from the heap each time a merge occurs; the distance cache is updated, and the distance values of the new clusters are reinserted into the heap. This approach reduces repeated computations and accelerates the cluster pair-finding process through the heap's prioritization mechanism. The algorithm complexity is reduced from $O(n^3)$ of the original cohesive clustering to $O(n^2 \log n)$, which improves the efficiency of the cohesive hierarchical clustering algorithm, makes it more suitable for large-scale datasets, and enhances the speed of the clustering process and resource[20,21].

Through the improved cohesive hierarchical clustering algorithm combining distance caching and heap optimization, the DataSet1 data, i.e., the cookie parameters of all websites, are clustered to cluster websites of the same type into the same cluster and to design a cluster label for each cluster, i.e., WebsetClusterLabel.The pseudocode is shown in Algorithm 1.

Algorithm 1 Agglomerative Clustering with Distance-Based Heap

Require: Dataset1, num_clusters
Ensure: WebsetClusterLabel
Normalize vectors using `StandardScaler` to obtain `vectors_scaled`.
Compute `dist_matrix` for all points using the pairwise Euclidean distance.
for each pair of points (i, j) in `vectors_scaled` **do**
  Push $(distance, i, j)$ to heap.
end for
Define a cache `dist_cache` to store computed distances.
while the number of clusters is reduced to `num_clusters` **do**
  Merge cluster j into cluster i and delete cluster j from clusters
  Update `dist_cache` and heap with new distances:
  for each remaining cluster k **do**
    if distance between (i, k) is not in `dist_cache` **then**
      Compute the minimum distance between points in clusters i and k
      Update `dist_cache` with the computed distance
    end if
    Push $(distance, i, k)$ to heap
  end for
end while
for each cluster i in clusters **do**
  Assign label i to all points in the cluster.
end for

Behavioral clustering methods After completing website clustering on DataSet1, based on the same data in DataSet1 and DataSet2, i.e., WebName and Host, the WebsetClusterLabel, i.e., the same type of website obtained above, is mapped to DataSet2 through the WebName and Host's Uniqueness is mapped to DataSet2 and clustered again in the same type of website through WebsetClusterLabel identity.

Due to different URLs and behavioral data within each cluster, the request body parameters and keyword data in the request line may exhibit varying distribution characteristics and patterns, often resulting in a disorganized state. This paper performs a finer cluster analysis to refine the data in these clusters[22]. This paper tries to design multiple clustering algorithms after many experiments to ensure the accurate categorization of the data. K-means, DBSCAN, Agglomerative Clustering, OPTICS, and Gaussian Mixture Models (GMM) clustering algorithms are chosen through careful data analysis.The advantages and disadvantages of several clustering algorithms are shown in Table. 1.

Table 1. Advantages and disadvantages of different clustering algorithms

Algorithm	Advantages	Disadvantages
K-means	1. Simple and easy to understand. 2. Efficient for large datasets. 3. Fast computation.	1. Requires specifying the number of clusters (K) in advance. 2. Sensitive to outliers. 3. Assumes spherical clusters.
DBSCAN	1. No need to specify the number of clusters. 2. Handles noise and outliers. 3. Can find clusters of arbitrary shapes.	1. Sensitive to parameters (distance threshold and min points). 2. Struggles with high-dimensional data.
Agglomerative Clustering	1. No need to specify the number of clusters. 2. Can capture hierarchical relationships.	1. High computational complexity. 2. Sensitive to noise and outliers.
OPTICS	1. Handles clusters of varying densities. 2. No need to specify the number of clusters. 3. Reveals clustering structure in the data.	1. High computational complexity. 2. Sensitive to parameter settings.
GMM	1. Capable of modeling clusters with different shapes. 2. Provides soft clustering (probability of membership). 3. Offers a well-interpretable model.	1. Requires specifying the number of clusters. 2. Sensitive to initial conditions. 3. Computationally intensive.

K-means is a center-of-mass-based clustering method suitable for dealing with spherical distributions through iterative optimization to classify the data points into the nearest clusters. K-means is a center-of-mass-based clustering method that divides data points into the nearest clusters through iterative optimization. It is suitable for dealing with globally distributed data, with the advantages of fast computation speed and simple implementation. It realizes clustering by defining core, boundary, and noise points and is suitable for

dealing with non-uniformly distributed data. Agglomerative Clustering gradually merges the most similar clusters by constructing a hierarchical tree that is able to deal with clusters of different shapes and sizes and is suitable for scenarios requiring hierarchical structural analysis.OPTICS is similar to DBSCAN but is able to deal with clusters of different densities better by generating clusters of different shapes. Clusters of different densities reveal the clustering structure in the data by generating ordered reachable graphs.GMM clusters the data again using the Expectation Maximization (EM) algorithm, which is suitable for dealing with data characterized by Gaussian distribution.

In this paper, based on the clustering results generated within each website clustering cluster, the cosine similarity $\text{cosine\_similarity}(i, j)$ between different clusters is calculated, and the cosine similarity between every two clusters is calculated to construct the cosine similarity matrix $\text{similarity\_matrix}[i, j]$ [23]. Based on the inter-cluster similarity matrix $\text{similarity\_matrix}[i, j]$, a "weighted" vote is assigned for each data point according to a threshold value by Eqs. 7, and if the data points are assigned to similar clusters (i.e., the similarity is above a threshold value of 0.5) in both clustering algorithms, a weighted vote is assigned according to Eqs. 8 to get a final voting score $\text{vote}[j]$.

$$\text{cosine\_similarity}(i, j) = \frac{i \cdot j}{\|i\| \|j\|} . \tag{7}$$

$$\text{vote}[j] += \text{similarity\_matrix}[i, j] \quad \text{if} \quad \text{similarity\_matrix}[i, j] > 0.5 . \tag{8}$$

Algorithm 2 Clustering with Multiple Methods and Voting Consensus

Require: Dataset2, WebsetClusterLabel
Ensure: BehaviorLabel

```
Extract ClusterIndices of points with the same WebsetLabel
Normalize the data points in ClusterIndices to obtain CombinedScaled
for each MethodName, MethodClass, ParamGrid in ClusteringMethods do
    for each parameter set params in ParameterGrid(PparamGrid) do
        Configure method using params
        Apply method to CombinedScaled to obtain labels.
    end for
end for
Initialize SimilarityMatrix with shape (NumPoints, NumPoints)
for each (MethodName, labels) in ClusterLabelsList do
    Compute pairwise cosine similarity.
    Update SimilarityMatrix with similarity values
end for
for each point i in CombinedScaled do
    Calculate votes based on similarities in SimilarityMatrix
    Obtain the final BehaviorLabel
end for
```

In this paper, according to the function and advantages and disadvantages of each clustering method, concerning the dataset used in the experiments of this paper, the algorithm with optimal performance and its parameter combinations are selected, and through the several clustering methods mentioned above, each cluster of clustering of websites is in-

tegrated through the consensus function based on the cosine similarity and the voting[24], and finally clusters out the different behaviors of users in different types of websites. The pseudocode is shown in Algorithm 2.

By integrating the clustering method through the consensus function based on similarity voting[25], based on the mapping WebsetClusterLabel of the results clustered from the DataSet and then set a small label for each behavior, i.e., BehaviorLabel, to cluster the same operation behaviors of the user in the same type of website, and finally, based on the obtained results WebsetClusterLabel and BehaviorLabel through the mapping of a certain type of user behavior under a certain type of website, so that the results are more precise, but also to reduce the data due to the different distribution characteristics and patterns of the use of a single clustering method of the results of the randomness brought about by the impact of the results of the final clustering results to improve the stability of the final clustering results.

4. Experiments and results

4.1. Experimental Configuration

The configuration of the experimental environment in this paper is shown in Table. 2.

Table 2. Experimental environment configuration

Name	Configuration Information
Development Language	Python 3.11
Framework	Pytorch 2.1.0 + cuda 12.3
GPU	NVIDIA GeForce RTX 4060 Laptop
CPU	13th Gen Intel(R) Core(TM) i7-13700H 2.40 GHz
Memory	16.0 GB

4.2. Evaluation indexes of experimental effect

This paper evaluates the effectiveness of the experimental clustering algorithm through the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. Eqs. 9 calculates the Silhouette Score (S), which is used to interpret and verify the consistency of the data points in the clusters, where a is the average distance of the sample from the other points in the same cluster, and b is the average distance of the sample from all the points in the nearest cluster. The contour coefficient combines the two aspects of Cohesion within clusters and Separation between clusters. The contour coefficient of each sample is calculated and averaged to evaluate the effect of the whole cluster.Eqs. 10measures the quality of clustering by calculating the similarity between individual clusters and the dispersion of data points within clusters, i.e., Davies-Bouldin Index (DBI), where N is the number of clusters, σ_i, σ_j are the average distances of all the points in clusters i and j from the centers of their respective clusters, and d_{ij} is the distance between the centers of clusters in clusters i and j. The smaller the DBI is, the better the effect of the clustering is

indicated.Eqs. 11 measures the quality of clustering by calculating the ratio of the sum of squares of the inter-cluster and intra-cluster dissociations, i.e., the Calinski-Harabasz Index (CHI), also known as the Variance Ratio Criterion (VRC), where k is the number of clusters, $\mathrm{Tr}(B_k)$ is the trace of the inter-cluster discretization matrix, and $\mathrm{Tr}(W_k)$ is the trace of the intra-cluster discretization matrix. n is the total number of samples. A larger CHI indicates better clustering[26].

$$S = \frac{b-a}{\max(a,b)}. \tag{9}$$

$$DBI = \frac{1}{N}\sum_{i=1}^{N}\max_{j\neq i}\left(\frac{\sigma_i+\sigma_j}{d_{ij}}\right). \tag{10}$$

$$CHI = \frac{\mathrm{Tr}(B_k)/(k-1)}{\mathrm{Tr}(W_k)/(n-k)}. \tag{11}$$

In this paper, the performance of the experimental model is evaluated by Accuracy and Recall. Accuracy is used to evaluate the correctness of the clustering results, and Recall is a measure of the model's ability to recognize samples of the positive class, that is, among all the samples that are actually positive, the proportion of samples correctly recognized as positive by the model, specifically, as in Eqs. 12 and Eqs. 13. TP represents the positive sample predicted to be positive by the model, which can be called the accuracy rate judged to be true. TN represents the negative sample predicted to be negative by the model, which can be referred to as the percentage of correct judgments that are false. FP represents the negative sample predicted by the model to be positive, which can be referred to as the false alarm rate. FN represents the positive sample predicted to be negative by the model, which can be referred to as the underreporting rate.

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}. \tag{12}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{13}$$

4.3. Analysis of results

The overall results obtained from the experiment are shown in Fig. 5. For the improved Agglomerative Clustering clustering method using distance caching in combination with heap optimization in DataSet1, based on the difference in clustering performance and effectiveness depending on the number of clusters, several experiments were conducted to compare the S obtained by the number of clusters clustered out of the data, the DBI and the CHI to select the clustering result with the most appropriate separation for each cluster with the best effect. After clustering the cookie parameters in DataSet1, roughly the same type of post-message templates from illegal gambling sites are clustered in the same cluster. According to the analysis of the obtained experimental results, the current illegal gambling websites are roughly board game websites, virtual lottery and scratch-off websites, sports betting websites, social gambling websites, and virtual sports betting websites. Each type can be further categorized into multiple post-message templates.

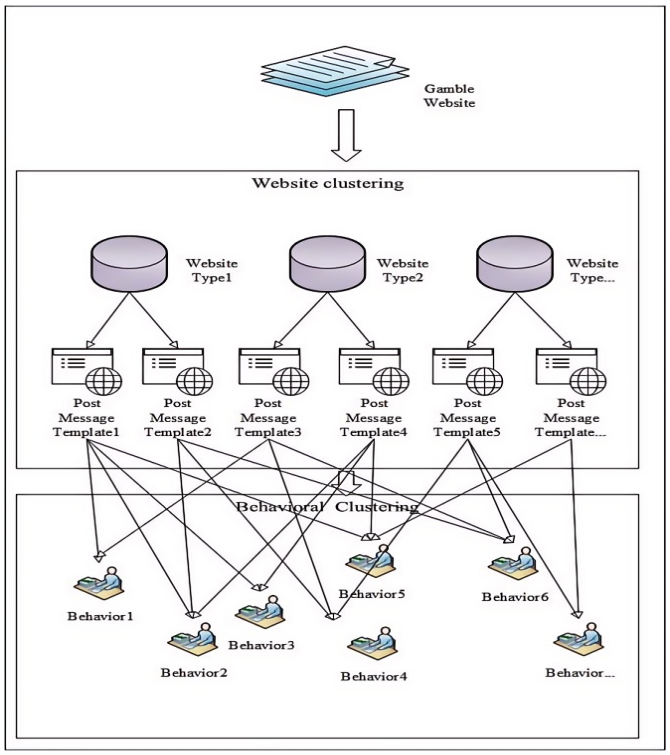
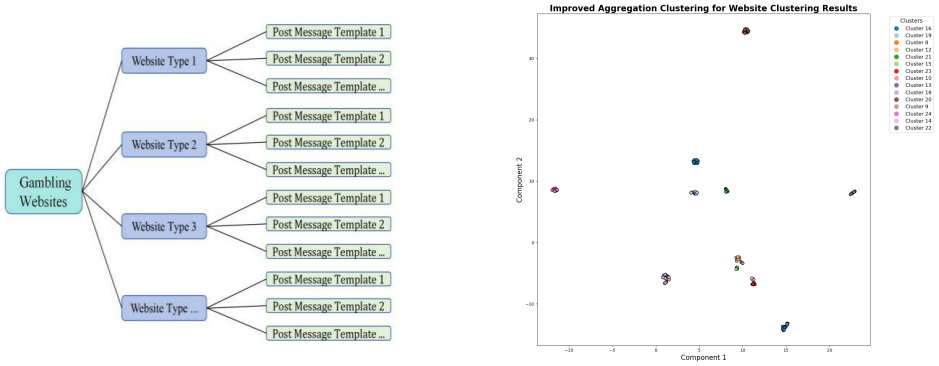


Fig. 5. Overall results of the experiment



(a) Website Message Clustering **(b)** Agglomerative Clustering Result

Fig. 6. Improved Agglomerative Clustering Result Classification and Scatter Plot

Through the above methods, this paper clusters, identifies, and detects illegal gambling websites and successfully detects illegal gambling websites of the same type. Current research commonly uses single clustering methods such as DBSCAN, K-means, and GMM to cluster and analyze the parameters obtained from POST messages. In contrast to the direct use of the above methods for clustering and detecting the parameters in the collected POST messages, in this paper, we use an improved cohesive hierarchical clustering algorithm that combines distance caching and heap optimization for clustering. This method not only improves the clustering of different types of websites, but also significantly improves the accuracy and recall of recognition and detection, shows higher effectiveness, and excels in several performance metrics.The following are the comparison results of several methods with the method mentioned in this paper for clustering detection, and their performance on the three metrics of S, Accuracy, and Recall, respectively, are shown in Table. 3.

Table 3. Comparison results of clustering detection

METHOD	SILHOUETTE SCORE	ACCURACY	RECALL
K-means	0.42	0.96368	0.96896
DBSCAN	0.48	0.95185	0.97533
GMM	0.52	0.97815	0.98907
Ours	0.57	0.98216	0.98942

From the comparison results in Table. 3, the model in this paper can significantly improve the clustering effect by further optimizing and adjusting the clustering model and method by comparing with the previous methods, showing better cluster separation and tightness. The classification and recognition effects are improved. According to the comparison results, the clustering analysis through phased and multi-algorithm can be more effective in identifying and detecting illegal gambling websites.

Since the two data sets of DataSet1 and DataSet2 are linked according to WebName and Host during the previous processing of the data set, the data are clustered in the same cluster under the same WebName and Host after the above clustering. According to DataSet1 and DataSet2 public data WebName and Host, the data of DataSet2 through the above clustering generated WebsetClusterLabel through the WebName and Host grouping, so that the same WebsetClusterLabel, that is, the same type of post message template illegal gambling WebsetClusterLabel, i.e., the same type of post message template illegal gambling website is grouped within the same group. Then, for each group of data, the same type of post message template illegal gambling websites are integrated and clustered according to the consensus function based on cosine similarity of multiple algorithms of K-means, DBSCAN, Agglomerative Clustering, OPTICS, and GMM clustering.

According to Fig.7, the percentage of Silhouette Score, DBI, and CHI effects for integrated clustering within clusters of the same type of post message templates on illegal gambling websites shows that the integrated clustering through multiple algorithms based on the cosine similarity of the consensus function is more effective to further categorize

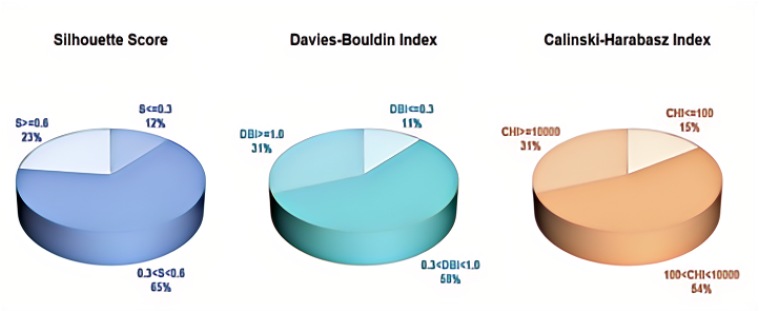


Fig. 7. Integration clustering effect percentage

the same post message templates of the same type of post message templates of illegal gambling websites with different behavioral approaches. In the real-world network environment, with different types of websites, users can realize different operations and behaviors. In this paper, we first cluster out the same type of post message template illegal gambling websites, and then cluster the different types of user operation behaviors together in the cluster through the integrated clustering of consensus functions based on cosine similarity of multiple algorithms and finally obtain the results of a certain type of website under a certain type of user operation of a certain behavior.

Based on the analysis of the experimental results in this paper, users of sports betting websites often place real-time bets while events are in progress. They analyze them based on past performance to make betting decisions, possibly exchanging betting advice and strategies through social interactions. Users of board game sites participate in various poker games and tournaments and may use statistical tools to analyze their performance. On lottery and instant game sites, users typically make random bets or select specific combinations of numbers; their participation is usually less frequent, the betting process is more straightforward, and users frequently check lottery results and manage winning information. Users interact with friends in games on social gambling sites, share game progress and strategies, and usually bet using virtual currency rather than real money. In contrast, users of virtual sports betting sites bet on simulated sporting events, such as virtual soccer or horse racing. They make quick bets and track the progress of the virtual event in real-time.

5. Conclusion and Outlook

In recent years, with the rapid development of the Internet, the rampant illegal gambling websites pose a severe challenge to the socio-economic and legal order of countries. In order to effectively curb online gambling behaviors, this paper examines website and behavioral characteristics through critical information from post messages obtained in a real-world network environment. The experimental results show that the different operational behaviors of users under different types of websites are obtained by first clustering all gambling websites using an improved cohesive hierarchical clustering algorithm combining distance caching and heap optimization to obtain the same type of post message template websites, and then integrating clustering of the same type of websites by

similarity-based consensus function. This multilevel clustering method shows high computational efficiency and classification accuracy when dealing with large-scale data and excels in indicators such as Silhouette Score, accuracy, and recall. It can effectively identify and detect gambling websites and user operating behaviors. By comparing various clustering methods and parameter combinations, this multilevel clustering method not only improves the accuracy of clustering but also enhances the robustness of the model to better adapt to changes in data. The approach in this paper can help identify and monitor illegal activity patterns, optimize risk assessment and alert systems, improve anti-fraud strategies, and support compliance and legal enforcement. In this way, standard features of the website and user behaviors can be revealed, and the efficiency of data analysis can be improved, thus effectively combating illegal gambling activities, protecting users' rights and interests, and maintaining the security of the online environment.

Although this study has made some progress in identifying and detecting illegal gambling websites, it still faces many challenges. First, it is not easy to acquire and label the data. Due to gambling websites' hidden nature and dynamic changes, obtaining high-quality datasets becomes a significant challenge. Second, regarding feature extraction and selection, effectively extracting and selecting features to improve the clustering effect still needs to be continuously optimized. Finally, the existing clustering algorithms have efficiency problems when dealing with large-scale and high-dimensional data, which need further optimization to improve performance. Future research should pay more attention to the practicality and efficiency of the technology, Such as blockchain and dynamically searchable encryption based data storage and sharing solutions provide secure data storage and privacy protection [27,28],to cope with the increasingly complex online gambling behavior and ensure the healthy development of cyberspace. Further development of automated data acquisition and cleaning tools, exploration of more advanced feature extraction techniques, and optimization of the computational performance of existing clustering algorithms are recommended to address these challenges better [29,30].

Acknowledgments. This work was Supported by Key Lab of Information Network Security, Ministry of Public Security,the National Natural Science Foundation of China under Grants 61672338, the Natural Science Foundation of Shanghai under Grant 21ZR1426500. .

References

1. Ghelfi, M., Scattola, P., Giudici, G., Velasco, V.: Online gambling: A systematic review of risk and protective factors in the adult population. In: Proceedings of the Journal of Gambling Studies. vol. 39, pp. 1–27 (2023)
2. Kong, X., Wang, C., Li, Y., Hou, J., Jiang, T., Liu, Z.: Traffic classification based on cnn-lstm hybrid network. In: International Forum on Digital TV and Wireless Multimedia Communications. pp. 401–411. Springer Singapore, Singapore (2021)
3. Mu, J., He, H., Li, L., Pang, S., Liu, C.: A hybrid network intrusion detection model based on cnn-lstm and attention mechanism. In: International Conference on Frontiers in Cyber Security. pp. 214–229. Springer Singapore, Singapore (2021)
4. Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q.E.U., Saleem, K., Faheem, M.H.: A deep learning-based phishing detection system using cnn, lstm, and lstm-cnn. Electronics 12(1), 232 (2023)
5. Alnemari, S., Alshammari, M.: Detecting phishing domains using machine learning. Applied Sciences 13(8), 4649 (2023)

6. Chen, Z., Fu, L., Yao, J., Guo, W., Plant, C., Wang, S.: Learnable graph convolutional network and feature fusion for multi-view learning. Information Fusion 95, 109–119 (2023)
7. Huang, Z., Ren, Y., Pu, X., Huang, S., Xu, Z., He, L.: Self-supervised graph attention networks for deep weighted multi-view clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 7936–7943 (2023)
8. Chen, Y., Zheng, R., Zhou, A., Liao, S., Liu, L.: Automatic detection of pornographic and gambling websites based on visual and textual content using a decision mechanism. Sensors 20(14), 3989 (2020)
9. Wang, C., Zhang, M., Shi, F., Xue, P., Li, Y.: A hybrid multimodal data fusion-based method for identifying gambling websites. Electronics 11(16), 2489 (2022)
10. Sun, G., Ye, F., Chai, T., Zhang, Z., Tong, X., Prasad, S.: Gambling domain name recognition via certificate and textual analysis. The Computer Journal 66(8), 1829–1839 (2023)
11. Singh, H., Kaur, P.: An effective clustering-based web page recommendation framework for e-commerce websites. SN Computer Science 2(4), 339 (2021)
12. Li, Y., Chu, X., Tian, D., Feng, J., Mu, W.: Customer segmentation using k-means clustering and the adaptive particle swarm optimization algorithm. Applied Soft Computing 113, 107924 (2021)
13. Liu, L.: e-commerce personalized recommendation based on machine learning technology. Mobile Information Systems 2022(1), 1761579 (2022)
14. Qiao, M., Wei, L., Han, D., et al.: Efficient multi-party psi and its application in port management. Computer Standards & Interfaces 91, 103884 (2025)
15. Jiang, T., Jia, L., Wan, C.M., et al.: The text modeling method of tibetan text combining word2vec and improved tf-idf. In: Proceedings of 2020 4th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE 2020). vol. 3, p. 8. IOP Publishing (2020)
16. Zhang, T., Wang, L.: Research on text classification method based on word2vec and improved tf-idf. In: Advances in Intelligent Systems and Interactive Applications: Proceedings of the 4th International Conference on Intelligent, Interactive Systems and Applications (IISA2019). pp. 199–205. Springer International Publishing (2020)
17. Xin, X., Han, D., Cui, M.: Daaps: A deformable-attention-based anchor-free person search model. Computers, Materials & Continua 77(2) (2023)
18. Ni, Q., Fan, Z., Zhang, L., Nugent, C.D., Cleland, I., Zhang, Y., Zhou, N.: Leveraging wearable sensors for human daily activity recognition with stacked denoising autoencoders. Sensors 20, 5114 (2020)
19. Fernández-García, M.E., Sancho-Gómez, J.L., Ros-Ros, A., Figueiras-Vidal, A.R.: Complete stacked denoising auto-encoders for regression. Neural Processing Letters 53, 787–797 (2021)
20. Shkaberina, G., Verenev, L., Tovbis, E., Rezova, N., Kazakovtsev, L.: Clustering algorithm with a greedy agglomerative heuristic and special distance measures. Algorithms 15, 191 (2022)
21. Ali, M.A., PP, F.R., Abd Elminaam, D.S.: An efficient heap based optimizer algorithm for feature selection. Mathematics 10, 2396 (2022)
22. Ezugwu, A.E., Ikotun, A.M., Oyelade, O.O., Abualigah, L., Agushaka, J.O., Eke, C.I., Akinyelu, A.A.: A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence 110, 104743 (2022)
23. Iffath, N., Mummadi, U.K., Taranum, F., Ahmad, S.S., Khan, I., Shravani, D.: Phishing website detection using ensemble learning models. In: AIP Conference Proceedings. vol. 3007, p. 1. AIP Publishing (2024)
24. Chen, G.: Scalable spectral clustering with cosine similarity. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 314–319. IEEE (2018)
25. Pho, K.H., Akbarzadeh, H., Parvin, H., et al.: A multi-level consensus function clustering ensemble. Soft Computing 25, 13147–13165 (2021)

26. Alizade, M., Kheni, R., Price, S., Sousa, B.C., Cote, D.L., Neamtu, R.: A comparative study of clustering methods for nanoindentation mapping data. Integrating Materials and Manufacturing Innovation 13, 526–540 (2024)
27. Li, J., Han, D., Wu, Z., et al.: A novel system for medical equipment supply chain traceability based on alliance chain and attribute and role access control. Future Generation Computer Systems 142, 195–211 (2023)
28. Li, J., Han, D., Weng, T.H., et al.: A secure data storage and sharing scheme for port supply chain based on blockchain and dynamic searchable encryption. Computer Standards & Interfaces 91, 103887 (2025)
29. Han, D., Pan, N., Li, K.C.: A traceable and revocable ciphertext-policy attribute-based encryption scheme based on privacy protection. IEEE Transactions on Dependable and Secure Computing 19(1), 316–327 (2022)
30. Han, D., Zhu, Y., Li, D., Liang, W., Souri, A., Li, K.C.: A blockchain-based auditable access control system for private data in service-centric iot environments. IEEE Transactions on Industrial Informatics 18(5), 3530–3540 (2022)

Zhimin Zhang is currently pursuing the M.S.degree with the School of Information Engineering, Shanghai Maritime University, Pudong, China. Her main research topic is network security.

Dezhi Han received the B.S. degree in applied physics from the Hefei University of Technology, Hefei, China, in 1990, and the M.S. and Ph.D. degrees in computing science fromthe Huazhong University of Science and Technology, Wuhan, China, in 2001 and 2005,respectively. He is currently a Professor with the Department of Computer, Shanghai Mar-itime University, Pudong, China, in 2010. His current research interests include cloud and outsourcing security, blockchain, wireless communication security, network, and information security.

Songyang Wu is a researcher and director of the Cyber Security Center of the Third Research Institute of the Ministry of Public Security (MPS) and also serves as the deputy director of the National Engineering Research Center for Network Security Level Protection and Security Technologies and the executive deputy director of the Key Laboratory ofthe Ministry of Public Security of the Ministry of Information Network Security, etc. He received his B.S. degree in Computer Science and Technology from Tongji University in2005 and his Ph.D. in Computer Application from Tongji University in 2011. He joined the Center of the Ministry of Public Security in the same year. He received his PhD in Computer Application from Tongji University in 2011. He joined the Network Security Center of the Third Research Institute of the Ministry of Public Security in the same year.His research interests include cybercrime investigation, electronic data forensics, ampledata security, and artificial intelligence security.

Wenqi Sun, Associate Researcher and Research Engineer at the Cyber Security Center of the Third Research Institute of the Ministry of Public Security; she received her B.S.degree in Computer Science and Technology from Northeastern University in 2010 and her Ph.D. degree in Computer Science and Technology from Tsinghua University in 2016. She joined the Cyber Security Center of the Third Research Institute of the Ministry of Public Security in 2018; her current research interest is in cybercrime investigation.

Shuxin Shi received an M.S. in Computer Science and Technology from Shanghai Maritime University, Pudong, China, in 2024 and is currently pursuing a Ph.D. in the Schoolof Information Engineering. His current research interest is network security.

Received: September 30, 2024; Accepted: January 17, 2025.

Computer Science and Information Systems 22(3):881–905 https://doi.org/10.2298/CSIS241022050S

Classification and Forecasting in Students' Progress using Multiple-Criteria Decision Making, K-Nearest Neighbors, and Multilayer Perceptron Methods

Slađana Spasić[1] and Violeta Tomašević[2]

[1] University of Belgrade – Institute for Multidisciplinary Research
Kneza Višeslava 1, 11030 Belgrade, Serbia
sladjana@imsi.bg.ac.rs
[2] Singidunum University, Faculty of Informatics and Computing
Danijelova 32, 11000 Belgrade, Serbia
vitomasevic@singidunum.ac.rs

Abstract. The research paper addresses students' performance in higher education. It proposes using the MCDM method - Promethee II to assess students' knowledge and the K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) methods for grade classification. The main goals are tracking and diagnosing students' knowledge levels, predicting their outcomes, and providing tailored recommendations. It helps to identify students at risk of not passing the course and evaluates teaching methods. This encourages student engagement and progress during the course. The research demonstrates the suitability of Promethee II, MLP, and KNN methods for effectively monitoring, classifying, and predicting students' progress during the semester, enhancing the objectivity of the assessment process.

Keywords: Promethee II, MLP, KNN, student's grades mark classification, student's achievement forecasting, Matthews Correlation Coefficient, Class Balance Accuracy

1. Introduction

The primary objective of higher education is to provide students with academic and professional knowledge in specific areas, which is evaluated based on the grades they achieve in exams. Educational Data Mining is a new field that examines academic performance to improve educational effectiveness [8]. Predicting student academic performance has been a major focus in the field of education. In the past decade, there has been a growing interest in understanding student performance in learning management systems due to recent advancements in artificial intelligence, data mining, and the increasing influence of outcome-based theory in education. Developed models have generally analyzed student data to predict various forms of learning outcomes, such as student achievements, dropout and at-risk rates, and feedback and recommendations. Study [34] analyzed relevant research from this period 2010-2020, showed that learning outcomes were predominantly measured by class standings and achievement scores, and regression and supervised machine learning models were commonly used to categorize student performance. Among these models, Neural Networks and Random Forests (RF) exhibited the highest prediction performance, while linear regression models fared the worst [34]. In higher education,

extensive research is conducted on student academic performance to address challenges such as underachievement and university dropout rates [14]. It is important to evaluate student performance in a course. This not only helps to determine their success, but also enables the identification of students who may be at risk of dropping out. Recent studies have shown that dropout rates in some European countries were between 14.7% to 34.1% in 2014 [9], while in Latin America, dropout rates are as high as 50%, leading to delayed completion of higher education for many students [2]. To address this, it is important to develop effective predictive models to identify at-risk students in a timely manner and provide them with personalized feedback and support.

Academicians measure student success in various ways, including final grades, grade point averages, and socio-economic aspects. Computational efforts, particularly those using data mining and learning analytics techniques, aim to improve student performance [6]. The timely prediction of student performance can help identify low-performing students and enable early interventions by educators, such as advising, progress monitoring, intelligent tutoring systems development, and policymaking [39]. The advanced methods utilized in learning analytics to predict student success are broadly categorized into supervised learning, unsupervised learning, data mining, and statistical approaches [21, 37]. Each category encompasses a plethora of intelligent algorithms, such as Artificial Neural Networks, Support Vector Machine, K-Nearest Neighbor (KNN), and RF. The factors influencing student performance are extensively researched, encompassing both academic (e.g., pre-admission scores and entry qualifications) and non-academic factors (gender, ethnicity, parents' socioeconomic status, emotional intelligence and resilience) [24, 15, 33]. With the increased use of distance, online, and hybrid learning, especially during the COVID-19 pandemic, it is important to develop fair assessment methods.

A recent study aimed to build a model using data mining techniques to test, predict, and understand the academic performance of IT students [23]. Students engage in planned activities to enhance their knowledge and achieve academic success. It is crucial to develop a method to predict overall performance and identify at-risk students early in the course and provide valuable feedback to teachers on the effectiveness of their teaching [30].

Until 2013, most studies used statistical methods and linear programming rather than neural network methods for academic achievement classification [24, 30]. The first study on predicting academic achievement compared four mathematical models, including multiple regression, multi-layer perceptual network, radial basis model, and support vector machines [22]. The focus of our research was on creating objective assessment methods and predicting learning outcomes to enhance teaching and learning techniques. To ensure that students are given appropriate feedback on time, it would be useful to monitor their progress throughout the semester. To address this problem, our proposed solution involves diagnosing a student's current level of knowledge, predicting their expected final outcome based on that state, and providing appropriate recommendations to the student. Additionally, it is crucial to continuously evaluate our teaching methods to ensure that they are appropriate and effective, and make any necessary changes to improve the learning experience for all participants. If a significant number of students are not achieving the expected level of progress, proactive steps will be taken to reevaluate our teaching methods and make any necessary changes to ensure that all participants receive the best possible education.

The primary objective of our research was to monitor the activities and progress of course participants throughout the semester to ensure they were on track to complete the course. If a student is not making sufficient progress, they should be advised to increase their efforts. To address this issue, we have employed a combination of Multiple-Criteria Decision-Making (MCDM) alongside classification methods, specifically KNN and Multi-Layer Perceptron (MLP). In previous research, this problem was addressed by applying various modern techniques, or a combination of these techniques, to data obtained through traditional methods of monitoring student success. These conventional methods typically summarize students' achievements in course activities in a scaled format. The innovative aspect of the proposed approach is introducing a more sophisticated way to monitor students' progress, utilizing multi-criteria analysis to gather quality input data for the prediction process. For this purpose, we employ the outranking-based Promethee II method [10]. This method was chosen from a wide range of MCDM methods because it gives an opportunity for a precise and detailed definition of the decision-maker's attitude towards different decision criteria.

The Promethee II method utilizes weighting coefficients and various types of preference functions assigned to the criteria. Unlike traditional approaches, this method enables lecturers to express their attitudes towards course activities in greater detail. By assigning weight coefficients, lecturers can favor or disfavor specific activities and establish their relative importance in the overall evaluation process. Additionally, by selecting the appropriate preference function and setting its thresholds, lecturers can accurately convey their personal views regarding specific activities.

Our proposal is to assess students' objective progress throughout the semester using the Promethee II method. At designated time points, we will monitor the results achieved by students on various course activities. The set of results obtained by a student at a specific moment represents their current state, meaning that throughout the semester, students pass through states that are time-dependent. The progress made by a student between two consecutive states can be quantified using the Promethee II method by treating states as alternatives and course activities as decision criteria. By comparing the net flows (which represent the output of the Promethee II method) of successive states, we can determine how much the next state is objectively better or worse than the previous one. This novel application of the Promethee II method is distinct because, in earlier applications, the alternatives were not time-dependent; instead, they represented a set of options that could address the problem. Although the temporal aspect was introduced into the Promethee II method in prior research [5], it was done in a different context, focusing on dynamic threshold settings to accommodate the decision-maker's temporal preferences.

Based on the results obtained using the Promethee II method, progression functions are generated that clearly illustrate the dynamics of student advancement throughout the semester. These functions are discrete and quantitatively describe the level of progress at specific time points and will be utilized as inputs for the MLP and KNN classifiers. The MLP and KNN classifiers will be compared to determine which one performs better. In practice, they can be used alternatively. Recent research [1, 40] has identified these classifiers as the most suitable options for this type of classification. Selecting the right algorithm is crucial for creating an effective predictive model. For example, authors of [4] found that logistic regression outperformed RF and KNN when predicting student dropout rates. Additionally, study [7] identified MLP, Logistic Regression, Support Vec-

tor Machines, and RF as the most accurate algorithms for various STEM programs at a Brazilian university. Despite various research reports discussing the adequacy and accuracy of different classification methods, we chose MLP due to its accurate predictions. Additionally, we selected the KNN model because it allows us to group students with similar performance levels in the course. This approach enables us to form groups of 3 to 20 students, allowing for personalized attention. As a result, weaker students can improve their progress, while more successful students can enhance their knowledge through more advanced lessons.

The generalizability of these models poses a significant challenge, as it is crucial to ensure that models trained on one group can also be effectively applied to others. While some researchers have experimented with ensemble methods [12], a one-size-fits-all approach is not practical. Differences in instructional context [18] and student demographics [31] can influence the effectiveness of a model. Therefore, it is essential to develop customized models for each degree program while recognizing that predictive performance may vary over time. Regularly assessing these models in their specific contexts is vital for effectively reducing dropout rates.

To implement the proposed solution in real-world educational settings, the lecturer needs to define several key components for the course:

1. Course activities – These are the decision-making criteria.
2. Lecturer's attitude – This involves determining the weight of each activity in the overall process and identifying which deviations in the earned points are significant, along with the degree to which they matter (preference functions).
3. Monitoring dynamics – This refers to the specific moments when progress will be tracked.

This information serves as the input data for software that utilizes the Promethee II method, a simplified and easily implementable version of the decision-making tool. The software then generates progress functions based on the results of multi-criteria analysis. Subsequently, these progress functions are input for another software application that implements various classification methods. For this research, data analysis and predictions were performed using Microsoft Excel and IBM SPSS Statistics 25 software (IBM, USA), which provided the environment for executing the proposed procedure.

The effectiveness of the approach improves as the volume of input data for the classifiers increases, leading to more reliable predictions. A larger set of input data is generated when there is a larger group of students in the course and when the system is used over a longer period, such as when the same lecturer teaches the course for several years. If needed, this data can be filtered by different time periods.

The progress of each student in the course is tracked using an appropriate progress function. When a new student enrolls, a new progress function is created, which contributes to the dataset used for predictions. This makes the system scalable in terms of the number of students, enhancing performance as the number increases. Additionally, the system can be expanded by adding new subjects. The input data can vary from subject to subject, depending on the lecturer, since the subjects are mutually independent. However, if this is the case, modifications will be necessary in the Excel implementation, although it is template-based.

The system is also scalable concerning the classification methods employed. The progress functions are generated independently of the classification method, allowing them to serve as input data for different classifiers.

To validate our approach experimentally, we should conduct a longitudinal study that follows at least 2-3 generations of students across one or more courses. Furthermore, we need to statistically test the hypothesis that the distribution of grades varies between students who experienced traditional teaching methods and those who were taught using new approach.

After the introduction in section 1, section 2 covers materials and methods, explaining the participants' data, problem statement, proposed solution, and implementation. Section 3 details the results of the Promethee II method application, classification, and statistical analysis. Section 4 contains the discussion and conclusion.

2. Material and Methods

2.1. Participants and Data Set

This research includes 400 anonymous students' data at Singidunum University in Belgrade, Serbia. The real data come from the learning activities of the students who attended the subject of Computer Architecture and Organization, which is performed in the first year of undergraduate studies at the Faculty of Informatics and Computing and the Faculty of Technical Sciences over three consecutive academic years (2021/22-2023/24). Only data related to teaching activities within the mentioned course were analyzed and all personal data was excluded. The students are registered under numerical codes to keep their identities anonymous. The data included in this analysis is represented with continuous numerical values, only the grade mark is discretized. The students' data has the following organization: student ID, attendance, activity, homework, test, and grade mark from 5 to 10. Data is collected for at five different time points (t_2–t_6). The grade mark is the output value based on the teacher's criterion. This study was conducted by the consensus on the design of the study of the Singidunum University.

2.2. Problem Statement

Let O be an object in the system S which at time t is in the state s_t. The object represents a student who attends the subject. The state s_t is determined by four parameters: attendance (p_1), activity (p_2), homework (p_3), and test (p_4). It can be represented as $p_1(s_t)$,…,$p_4(s_t)$. Object state parameters are recorded in six times in moments t_j, $j = 1, \ldots, 6$ (moment t_1 represents the initial state in which all parameters are zero, i.e., $p_q(s_1) = 0$, $q = 1, \ldots, 4$), and thus the state matrix of the object $M_{6\times 4}$ is obtained:

$$M = \begin{bmatrix} p_1(s_1) & p_2(s_1) & p_3(s_1) & p_4(s_1) \\ p_1(s_2) & p_2(s_2) & p_3(s_2) & p_4(s_2) \\ p_1(s_3) & p_2(s_3) & p_3(s_3) & p_4(s_3) \\ p_1(s_4) & p_2(s_4) & p_3(s_4) & p_4(s_4) \\ p_1(s_5) & p_2(s_5) & p_3(s_5) & p_4(s_5) \\ p_1(s_6) & p_2(s_6) & p_3(s_6) & p_4(s_6) \end{bmatrix} \tag{1}$$

The recording timing aligns with teaching activities and does not need to be evenly spaced. State s_6 indicates if the desired outcome is achieved. If not, it suggests ineffective teaching or insufficient learning. Early identification is crucial to avoid negative outcomes.

It is important to note that the selection of parameters p_1–p_4 was based on the structure of the subject for which the study was conducted. However, the proposed solution is flexible and can be easily adapted to meet the needs of any subject. The lecturer can define an arbitrary number of different parameters based on the content and scope of the material being studied, the group of students enrolled in the course, the assessment goals, and any personal or other requirements relevant to the course.

Our model can incorporate additional data to predict the final success or failure of a course. Personal factors, such as socioeconomic status, demographic details, psychological aspects, and academic performance, can be included in the Promethee II analysis as predictors of success. While many researchers have extensively examined the factors influencing student success, our university currently considers only a limited set of data during enrollment. These data include name, age, gender, place of birth, educational background, citizenship, pre-enrollment results, and entry qualifications. In this paper, it is theoretically feasible to use these data as predictors of success, with their inclusion at the additional time point (t_0). Also, it is possible to include the initial test at time point t_1. Specifically, we could utilize the input data collected during the enrollment process, which is available in the university's information system, similar to the approach taken by [1]. However, we refrained from using these data in this analysis due to personal data protection regulations. Importantly, if individual student consent were obtained, our model could potentially include these additional data. Including them might improve the accuracy of our predictions. According to the findings presented in the paper [25], variables related to grade point average, socioeconomic factors, and course completion rates could positively impact our model's effectiveness.

Additionally, the lecturer can establish the dynamics of monitoring student progress, including the timing and number of evaluations, in line with the teaching methods used in the course to achieve the desired outcomes.

2.3. Proposed Problem Solution

The proposed solution aims to identify the state of the object that indicates an unfavorable outcome based on the teacher's experiences with the teaching procedure over the past two years. Valid conclusions drawn from these experiences are applied to real student or group data to identify a critical state and make an appropriate decision. Fig. 1 shows the structure of the proposed system model.

Let S^{400} be the system of 400 objects which consists of two disjunctive subsystems S^{300} (students who have attended the subject in the previous three years) and S^{100} (new students from current academic year) of 300 and 100 objects, respectively. All of these systems are extensions of the system S (defined in Sec. 2.2). The teaching activities of the defined procedure are applied to the objects of the S^{400} system. The data subset of S^{300} provides training and the data subset of S^{100} test data with a ratio of 3:1.

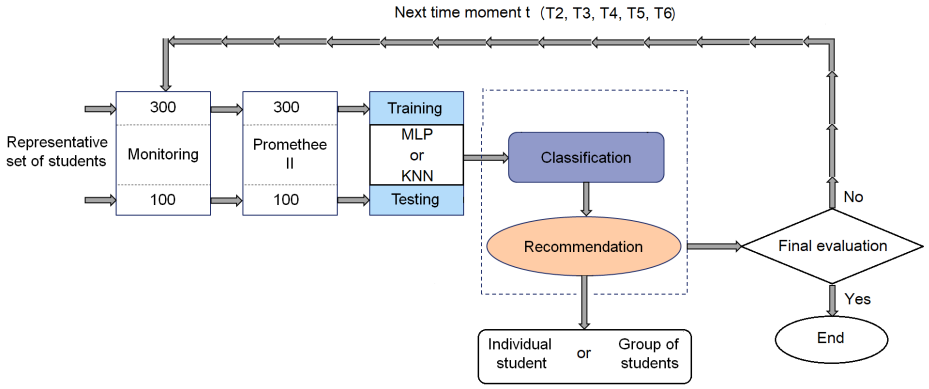


Fig. 1. The structure of the proposed system model

2.4. System Model Implementation

Monitoring. During the course, monitoring was done by recording the state matrices of all participants M_i, $i = 1, \ldots, 400$. All state parameters (attendance, activity, homework and test) are cumulative. Therefore, each state parameter at time t is the sum of the value of that parameter at time $t-1$ and the result achieved between these time points. Attendance at lectures (p_1) was scored with a maximum of 5 points (1 point between every two recordings, i.e., $5 \cdot 1$). Class activity (p_2) was assessed by the lecturer in the range of 0–10 points ($5 \cdot 2$ points). Students could win a maximum of 35 points ($5 \cdot 7$ points) on homework (p_3), and up to 50 points ($5 \cdot 10$ points) on tests (p_4).

For example, the state matrix from (1) with the results achieved by the O_{83} student is given by (2). The state matrices have a first row of zeros because we assumed that all students start the course with the same level of knowledge. However, if we were to conduct an entry test, this row would be filled with numbers greater than or equal to zero.

$$M_{83} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.5 & 0 & 4 & 7 \\ 1.5 & 1 & 8 & 14 \\ 2.5 & 2 & 11 & 21 \\ 3.5 & 3.5 & 16 & 29 \\ 4.5 & 5.5 & 20 & 35 \end{bmatrix} \tag{2}$$

Multi-criteria Analysis. Multi-criteria analysis was conducted on each state matrix M_i. The objective is to evaluate how much each state of the object O_i is better or worse than the previous one. Since every two consecutive states from M_i are compared, the total number of comparisons is 2000. The states are compared using the Promethee II method. Five evaluation tables $T_i(j)$, $j = 1, \ldots, 5$, are created for the object O_i. The table $T_i(j)$ establishes a connection between the set of alternatives representing the two successive states through which the object O_i passed $A = \{s^i_{j+1}, s^i_j\}$ and the set of criteria representing the state parameters $C = \{p_1, p_2, p_3, p_4\}$ (Table 1).

Table 1. Evaluation table

$T_i(j)$	p_1	p_2	p_3	p_4
s^i_{j+1}	$p_1(s^i_{j+1})$	$p_2(s^i_{j+1})$	$p_3(s^i_{j+1})$	$p_4(s^i_{j+1})$
s^i_j	$p_1(s^i_j)$	$p_2(s^i_j)$	$p_3(s^i_j)$	$p_4(s^i_j)$

The nature of the introduced criteria is such that we strive to maximize them because they positively contribute to the desired goal. As all criteria are not equally important, they have been assigned priorities.The test (p_4) has the highest priority 5 because it directly reflects acquired knowledge. Homework (p_3) weights 2.5 because it reflects knowledge, but homework is not time-critical, and allows external assistance. Activity (p_2) has priority 1 as it reflects the lecturer's objective impression of the course participant, while attendance (p_1) weights 0.5 as it only reflects the physical presence in the classes. These priorities have been normalized to obtain non-negative relative weight coefficients (w_q, $q = 1, \ldots, 4$) for the criteria, with the sum of these coefficients equaling 1.

In general, different priorities can be assigned for various parameters depending on the academic discipline of the course, the method of assessing student's knowledge, and the set of criteria. For example, a teacher of foreign languages or art history history might prioritize attendance and participation more than a teacher of mathematical analysis would. In contrast, for students in a mathematical analysis course, greater emphasis may be placed on homework and tests.

Our approach allows lecturer the flexibility to define evaluation parameters according to their preferences. Based on their extensive teaching experience, the lecturer associates with each criterion one of the six preference functions recommended by the authors of the Promehtee II method (*Usual*, *U-Shape*, *V-Shape*, *Level*, *Linear*, and *Gaussian*), the one he considers most suitable for that parameter. Otherwise, the Promethee II method allows the addition of new preference functions, so their set can be expanded if necessary.

The preference function reflects the analyst's attitude towards the value difference between the two alternatives.

For the criterion p_q, $q = 1, \ldots, 4$, the difference for alternatives s^i_{j+1} and s^i_j is calculated as

$$d_q(s^i_{j+1}, s^i_j) = p_q(s^i_{j+1}) - p_q(s^i_j)\,. \tag{3}$$

The analyst uses thresholds to indicate the significance of the difference and to what degree it matters to him. This shows his preference for alternatives based on a specific criterion. This can be represented by the preference function in the form of a graphical dependence of the preference towards the alternative s^i_{j+1} in relation to the alternative s^i_j, denoted by $P = P_q(s^i_{j+1}, s^i_j)$, and the difference $d = d_q(s^i_{j+1}, s^i_j)$ from (3).

Among the six preference functions proposed by the author of the Promethee II method, three are associated with the criteria p_q, $q = 1, \ldots, 4$ (Fig. 2).

Criteria p_1 and p_2 are associated with *Level* preference function because the difference of up to 20% in attendance (0.2 out of 1 point) and up to 25% in activity (0.5 out of 2 points) is considered small enough to be neglected. Significant preference is given only at differences of 80% for presence (0.8 out of 1 point) and 75% for activity (1.5 out of 2 points). Criterion p_3 was assigned a *V-Shape* preference function with a threshold 7 that shows that any difference in the number of points on homework is important with linear

growth. *Gaussian* preference function with a threshold 5 is associated with criterion p_4 as is often used to assess the success of results achieved in exams.

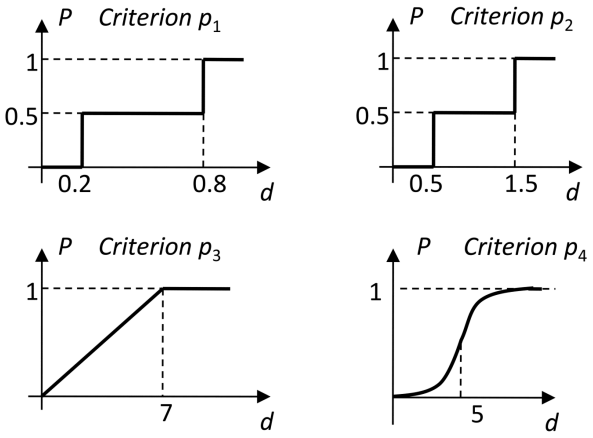


Fig. 2. Preference functions: p_1 – attendance, p_2 – activity, p_3 – homework, and p_4 - test

The Promethee II method is applied five times over the state matrix M_i, once for each evaluation table $T_i(j)$, $j = 1, \ldots, 5$. In general, the Promethee II method compares alternatives from a set of alternatives L based on m criteria by calculating:

aggregated preference index

$$\pi(a_i, a_l) = \sum_{j=1}^{m} P_j(a_i, a_l) \cdot w_j \, , \forall a_i, a_l \in L \tag{4}$$

positive (or outgoing) outranking flow

$$\Phi^{+}(a_i) = \frac{1}{n-1} \sum_{x \in L} \pi(a_i, x) \tag{5}$$

negative (or ingoing) outranking flow

$$\Phi^{-}(a_i) = \frac{1}{n-1} \sum_{x \in L} \pi(x, a_i) \tag{6}$$

net outranking flow

$$\Phi(a_i) = \Phi^{+}(a_i) - \Phi^{-}(a_i) \tag{7}$$

and then ranks the alternatives in order of decreasing values of Φ.

Due to the specificity of the analyzed problem (comparison of only two alternatives), the implementation of the Promethee II method is reduced to the calculation of the aggre-

gated preference index from (4):

$$\pi(s_{j+1}^i, s_j^i) = \sum_{q=1}^{4} P_q(s_{j+1}^i, s_j^i) \cdot w_q \tag{8}$$

Namely, in the considered case for $n = 2$ and $L = A$, from (5) it follows

$$\Phi^+(s_{j+1}^i) = \pi(s_{j+1}^i, s_j^i)\,. \tag{9}$$

As $d_q(s_j^i, s_{j+1}^i) \leq 0$ due to the cumulative nature of the criteria, it follows that $P_q(s_j^i, s_{j+1}^i) = 0$ and $\pi(s_j^i, s_{j+1}^i) = 0$. From (6) and (7), it follows that

$$\Phi(s_{j+1}^i) = 0 \text{ and } \Phi(s_{j+1}^i) = \Phi^+(s_{j+1}^i) = \pi(s_{j+1}^i, s_j^i)\,. \tag{10}$$

The index π shows how much the alternative s_{j+1}^i is better than the alternative s_j^i.

As a result of multi-criteria analysis performed on the state matrices of all participants, the M_π^{400} matrix was obtained. This matrix has the following appearance:

$$M_\pi^{400} = \begin{bmatrix} \pi(s_2^1, s_1^1) & \pi(s_3^1, s_2^1) & \pi(s_4^1, s_3^1) & \pi(s_5^1, s_4^1) & \pi(s_6^1, s_5^1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.5336 & 0.6169 & 0.5772 & 0.7662 & 0.6105 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \pi(s_2^{400}, s_1^{400}) & \pi(s_3^{400}, s_2^{400}) & \pi(s_4^{400}, s_3^{400}) & \pi(s_5^{400}, s_4^{400}) & \pi(s_6^{400}, s_5^{400}) \end{bmatrix} \tag{11}$$

Numerical values in the matrix correspond to the O_{83} course participant. As expected, due to the cumulative nature of state parameters, elements of matrix M_π^{400} are in the range $[0, 1]$.

Progress Functions. Progress functions are generated based on the M_π^{400} matrix. For each object O_i, a discrete cumulative progress function f_p^i is formed, which shows how that object progresses towards the set goal over time. The function is defined as follows:

$$\begin{aligned} &f_p^i(t_1) = 0 \\ &f_p^i(t_j) = \sum_{r=1}^{j-1} \pi(s_{r+1}^i, s_r^i) = f_p^i(t_{j-1}) + \pi(s_j^i, s_{j-1}^i), \quad 1 < j \leq 6 \end{aligned} \tag{12}$$

For illustration, the graph of the f_p^{83} function is shown in Fig. 3.

Classification Models Application. We employed MLP and KNN classificators to draw conclusions from the input values provided for the training set and to predict the grade marks on the test set. The students' grades ranged from 5 to 10 (from poor to excellent), requiring the use of a multi-class model. Data analysis and predictions were carried out using IBM SPSS Statistics 25 software (IBM, USA).

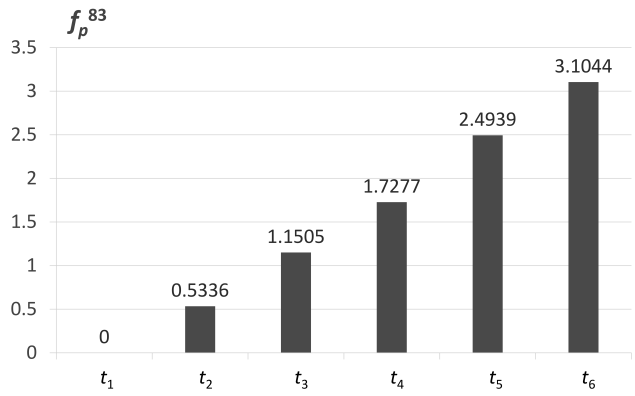


Fig. 3. O_{83} student progress function

Multi-layer Perceptron. The specific type of neural network called a Multi-layer Perceptron (MLP) was introduced by Frank Rosenblatt who is widely acknowledged as a pioneer in the training of neural networks, including multi-layer perceptron [38]. The MLP uses feedforward architecture and has one or more non-linear hidden layers between the input and output layers. A detailed explanation can be found in [35]. MLP can be used in predicting or classifying problems. It is a supervised method that means the results of prediction can be compared with the values of the target variables. MLP learns a function

$$f : R_m \to R_n \tag{13}$$

by training on a dataset, where m is the dimension of the input vector and n is the dimension for output vector. For a given set of m-dim input vectors and n-dim target variables, a nonlinear function for classification can be found.

In our investigation, the data set comprises student information from three consecutive academic years (2021/22 to 2023/24). We assigned cases to split the data into training (75%) and test (25%) sets. Before selecting hyperparameters, we conducted a series of classification and prediction tests that guided our decisions. We utilized predefined activation functions to evaluate the best options. Input variables were specified as covariates, and they are rescaled by default to improve network training. The automatic architecture of a network with a single hidden layer was selected with the best number of units in the hidden layer and the default activation functions for the hidden and output layers. We employed the hyperbolic tangent activation function in the hidden layer, while the output layer used the "Softmax" activation function.

Avoiding overfitting is essential in machine learning to ensure that a model generalizes well to unseen data. We used various techniques to achieve this. Our sufficiently large dataset captures the underlying patterns, and more data could enhance generalization. We split the dataset into training and testing sets, training the model on the former and evaluating it on the latter. The independent variable importance analysis was conducted to assess the significance of each predictor.

When learning neural network is completed, an MLP model is created that will be used to classify data. Classification is based on the values of progress function in milestones

$$f_p^i(t_j), \quad j = 2, 3, 4, 5, 6\,. \tag{14}$$

This used the input vector with variables T2 to T6 as features. Target variable was the student's grade mark ranged from 5 to 10. During training, we monitored the model's performance on the validation set. The algorithm used one consecutive step with no decrease in error as a stopping rule, a maximum of 15-minute training, and a maximum relative change in training error of 0.0001.

K-Nearest Neighbors. The K-Nearest Neighbors (KNN) model is widely used in machine learning for classification and regression. The initial concepts of the KNN model were introduced by [16] and further developed by [13]. KNN is a machine learning technique that can be used for supervised classification of cases based on their similarity to other cases. "Neighbors" are similar cases close to each other, while different cases are far from each other. The distance between the two cases is a measure of their difference and could be Euclidean or another distance. In this paper, Euclidean distance given in (15) is used to measure the similarity between neighbors:

$$d(x, y) = \sum_{i=0}^{N} \sqrt{(x_i^2 - y_i^2)} \tag{15}$$

The free parameter K is the number of nearest neighbors to be examined. The distance of the new case from each of the cases in the model is calculated, then the classifications of the most similar cases are added up and the new case is placed in the category that contains the largest number of nearest neighbors.

We implemented a KNN classification model, using variables T2 to T6 as features and the grade mark as the target variable. The value of the parameter K was selected based on the error rate indicated in the error log graph (Fig. 4), which demonstrated that the smallest error rate occurred when $K = 3$. Consequently, we chose $K = 3$ as the optimal parameter value for our KNN model.

Statistical Analysis and Measures for Multi-class Classification. Exploratory and data analyses were performed using the IBM SPSS Statistics 25 software (IBM, USA). The values of progress function in milestones (T2, T3, T4, T5, and T6) were used as input variables for the final grade prediction. There were 300 observations in the training group and 100 in the test group. Correlation analysis for paired samples was applied to test the correlation between the observed and predicted grade at the significance level $P < 0.05$.

Various metrics are available to assess the effectiveness of any multi-class classifier, making them valuable for comparing different models and analyzing a model's behavior when adjusting parameters [19]. These metrics are derived from the Confusion Matrix, providing essential information about algorithm and classification rule performance. We evaluated MLP and KNN methods for multi-class classification using a gold standard, measuring the level of agreement between observed and predicted grade marks.

Class Balance Accuracy (CBA) is a measure that aims to balance precision and recall for each input class [32]. Let is G the set of class labels with cardinality k, where $i, j =$

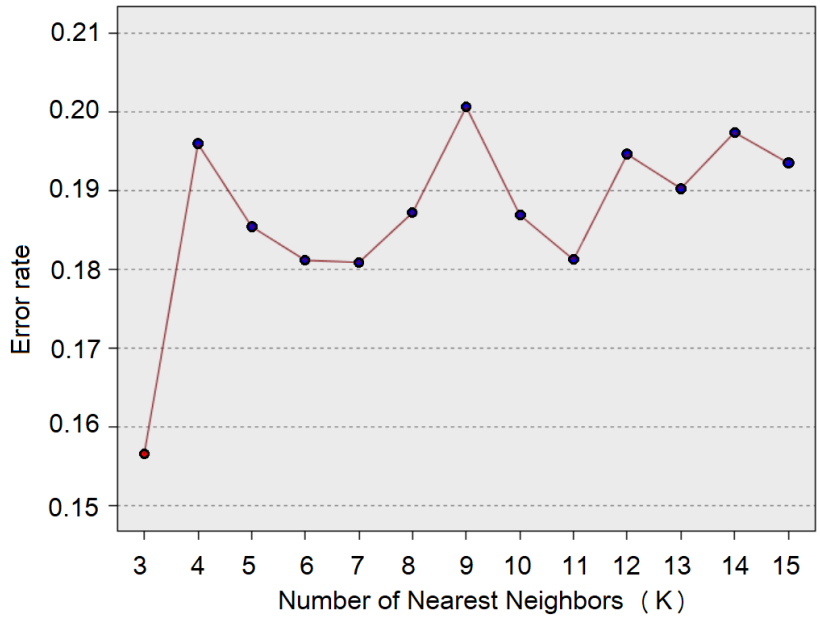


Fig. 4. K selection error log

$1, 2, \ldots, k$, and k is number of classes. The confusion matrix C^k is a k-dim square matrix or contingency table with elements c_{ij} representing the number of cases with true label i classified into group j. For confusion matrix C^k, $k > 2$, CBA is defined as

$$\text{CBA} = \frac{\sum_{i=1}^{k} \frac{c_{ii}}{\max(t_i, p_i)}}{k}, \tag{16}$$

where

$$t_j = \sum_{i=1}^{k} c_{ij} \tag{17}$$

$$p_i = \sum_{j=1}^{k} c_{ij} \tag{18}$$

$$c = \sum_{i=1}^{k} c_{ii} \tag{19}$$

$$s = \sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} \tag{20}$$

t_j is the number of times class j truly occurred (column total),
p_i is the number of times class i was predicted (row total),

c is the total number of samples correctly predicted, and
s the total number of samples.

Brian W. Matthews introduced the Matthews Correlation Coefficient (MCC) [29, 28] that became widely used as a measure to evaluate the performance of Machine Learning techniques, with some adaptations for multi-class scenarios [11]. The MCC has a range of $[-1, 1]$. Values close to 1 indicate very accurate predictions, suggesting a strong positive correlation between the predicted values closely matching the actual classification. An MCC of 0 suggests no correlation between the variables, indicating that the classifier is randomly assigning units to classes without any connection to their true values [19]. MCC can also be negative, indicating an inverse relationship between true and predicted classes. While this is undesirable, it often occurs due to modeling errors. A strong inverse correlation suggests that the model has learned how to classify the data but consistently switches all the labels.

In the multi-class case, the MCC can be defined in terms of a confusion matrix C^k for k classes and according to (17)–(20):

$$\text{MCC} = \frac{c \cdot s - \sum_{j=1}^{k} p_j \cdot t_j}{\sqrt{\left(s^2 - \sum_{j=1}^{k} p_j^2\right)\left(s^2 - \sum_{j=1}^{k} t_j^2\right)}} \tag{21}$$

In the multi-class case, the calculation of Cohen's Kappa (Kappa) is similar to Matthews Correlation Coefficient [17]. Referring to multi-class confusion matrix C^k and according to (17)–(20):

$$\text{Kappa} = \frac{c \cdot s - \sum_{j=1}^{k} p_j \cdot t_j}{s^2 - \sum_{j=1}^{k} p_j \cdot t_j} \tag{22}$$

Kappa allows the comparison of two models with the same accuracy but different Cohen's Kappa values. The Kappa and MCC statistics both range from -1 to $+1$, and their interpretation is as follows:

$[0.00, 0.09]$ agreement equivalent to chance,
$[0.10, 0.20]$ slight agreement,
$[0.21, 0.40]$ fair agreement,
$[0.41, 0.60]$ moderate agreement,
$[0.61, 0.80]$ substantial agreement,
$[0.81, 0.99]$ near perfect agreement, and
1.00 perfect agreement.

Negative values can be understood in the context of the MCC statistic.

3. Results

The classification was based on the progress function values $f_p^i(t_j)$, $j = 2, 3, 4, 5, 6$, at milestones t_2, t_3, t_4, t_5, and t_6, denoted as T2, T3, T4, T5, and T6 input variables. The accuracy of MLP and KNN classification was validated using a confusion matrix and statistical measures.

The MLP classification results are given in Table 2.

Table 2. MLP classification

MLP	Training	Predicted grade						Correct	Test	Predicted grade						Correct
Input var.	Observed	5	6	7	8	9	10		Observed	5	6	7	8	9	10	
T2, T3	5	45	10	2	1	0	0	77.60%	5	11	1	0	0	0	0	91.70%
	6	25	21	14	2	0	0	33.90%	6	8	9	6	0	0	0	39.10%
	7	0	8	32	14	0	0	53.30%	7	0	4	13	9	0	0	50.00%
	8	0	1	12	37	3	2	67.30%	8	0	0	3	14	0	0	82.40%
	9	0	0	1	8	20	7	55.60%	9	0	0	0	6	6	0	50.00%
	10	0	0	0	0	3	26	89.70%	10	0	0	0	0	1	9	90.00%
	Overall %	25.30%	13.30%	20.30%	20.70%	8.70%	11.70%	60.30%	Overall %	19.00%	14.00%	22.00%	29.00%	7.00%	9.00%	62.00%
T2, T3, T4	5	45	13	0	0	0	0	77.60%	5	11	1	0	0	0	0	91.70%
	6	10	44	8	0	0	0	71.00%	6	4	16	3	0	0	0	69.60%
	7	0	11	38	11	0	0	63.30%	7	0	3	16	7	0	0	61.50%
	8	0	0	8	37	10	0	67.30%	8	0	0	1	15	1	0	88.20%
	9	0	0	0	8	24	4	66.70%	9	0	0	0	3	9	0	75.00%
	10	0	0	0	0	5	24	82.80%	10	0	0	0	0	1	9	90.00%
	Overall %	18.30%	22.70%	18.00%	18.70%	13.00%	9.30%	70.70%	Overall %	15.00%	20.00%	20.00%	25.00%	11.00%	9.00%	76.00%
T2,T3, T4, T5	5	54	4	0	0	0	0	91.30%	5	11	1	0	0	0	0	91.70%
	6	10	48	4	0	0	0	77.40%	6	1	22	0	0	0	0	95.70%
	7	0	4	51	5	0	0	85.00%	7	0	1	21	4	0	0	80.80%
	8	0	0	4	48	3	0	87.30%	8	0	0	2	13	2	0	76.50%
	9	0	0	0	2	31	3	86.10%	9	0	0	0	2	10	0	83.30%
	10	0	0	0	0	2	27	93.10%	10	0	0	0	0	1	9	90.00%
	Overall %	21.30%	18.70%	19.70%	18.30%	12.00%	10.00%	86.30%	Overall %	12.00%	24.00%	23.00%	19.00%	13.00%	9.00%	86.00%
T2,T3, T4, T5,T6	5	58	0	0	0	0	0	100.00%	5	12	0	0	0	0	0	100.00%
	6	0	62	0	0	0	0	100.00%	6	0	23	0	0	0	0	100.00%
	7	0	1	59	0	0	0	98.30%	7	0	0	26	0	0	0	100.00%
	8	0	0	0	55	0	0	100.00%	8	0	0	0	17	0	0	100.00%
	9	0	0	0	0	36	0	100.00%	9	0	0	0	0	12	0	100.00%
	10	0	0	0	0	0	29	100.00%	10	0	0	0	0	0	10	100.00%
	Overall %	19.30%	21.00%	19.70%	18.30%	12.00%	9.70%	99.70%	Overall %	12.00%	23.00%	26.00%	17.00%	12.00%	10.00%	100.00%

These results indicate that the accuracy achieved using the input variables T5 and T6 was between 93% and 94.3% (not shown). There was a small variation in accuracy when using the independent variables T4-T6 and T3–T6, with accuracies ranging from 92% to 94.3%. In all three cases, T6 was found to be the most important independent variable in the MLP classification. When using independent variables T2–T5, the accuracy was between 86.0% and 86.3%. However, when using T3–T5 variables, the accuracy slightly decreased to 84%–84.3%, with T5 being determined as the most important independent variable in the MLP classification (Table 2). Using T2–T4 or T3–T4 as independent variables resulted in a further decrease in classification accuracy to 70.3%–78.0%, with T4 being identified as the most important input variable. The highest correct classification rate for the training set was 99.7% when all five input variables were used, while the test set had a predicted grade of 100% (Table 2). Fig. 5A shows a chart of run example of one-hidden-layer MLP with three variables in the input layer and the student's grade in the output layer. The corresponding Receiver-operating characteristic (ROC) curves

and values of the Area Under the Curve (AUC) are plotted for six classes (Fig. 5B). The dependent variable has six categories, so each curve treats the category at issue as the positive state versus the aggregate of all other categories. It can be seen that the best results are achieved for grades 5 and 10: AUC was approximately 1 which represents an ideal measure of separability. However, AUC values are also large for the other classes (0.918-0.963).

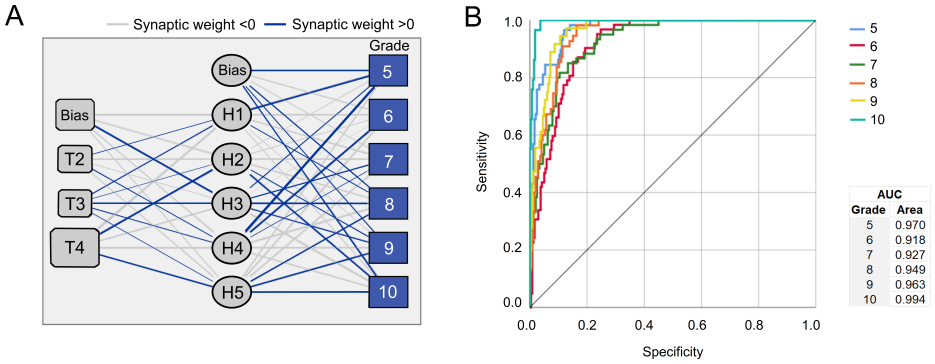


Fig. 5. A. MLP classifier: Classification is based on the values of progress function in milestones: T2, T3, and T4. A hidden layer activation function was hyperbolic tangent, while the output layer activation function was "Softmax". The target variable/class was the student's grade marks ranging from 5 (drop down) to 10 (excellent). B. ROC curves and corresponding AUC values

The results of the KNN classification (Table 3) indicated poorer performance compared to the MLP classifier. The highest accuracy was attained using the input variables T2–T6 (93.0%–94.0%). There was a slight variation in accuracy for the input variables T4–T6 (92.3%–94%), T5–T6 (91%–92%), and T3–T6 (90.7%–93%) (not displayed). When the independent variables were T2–T5, the accuracy ranged from 78.0%–79.0%, while for T3–T5, it ranged from 77.0%–83.0%.

It has been observed that in both MLP and KNN classifiers, the variable T6 has a significant impact on classification accuracy. However, while T6 is important for grading, it is not sufficient for predicting grades and guiding students towards their desired grade in a timely manner. Although the accuracy is slightly lower when T6 is not used as an input variable in classification, the results of the classification based on T2–T5 can be very influential in predicting student success in the final grade. When attempting to predict a student's success on the exam (at moment t_6) using input variables T2–T5, better results are obtained by using the MLP (86.0%) compared to the KNN classifier (79.0%). Although the difference of 7% accuracy is not negligible, the KNN classification offers the advantage of comparing the achievements of an individual student with those of their three to five nearest neighbors, allowing us to form groups of students with similar achievements and monitor their progress over time (Fig. 6B). Fig. 6 depicts how a focal case, student #83, would be classified using the three nearest neighbors ($K = 3$). KNN showed lower accuracy in general, except in cases of T2–T4 and T3–T4 input variables. This suggests a potential advantage of the MLP classifier.

Table 3. KNN classification

KNN	Training	Predicted grade						Correct	Test	Predicted grade						Correct
Input var.	Observed	5	6	7	8	9	10		Observed	5	6	7	8	9	10	
T2, T3	5	32	22	3	1	0	0	55.20%	5	8	4	0	0	0	0	66.70%
	6	15	33	11	3	0	0	53.20%	6	4	14	5	0	0	0	60.90%
	7	6	18	23	14	0	0	38.30%	7	0	4	14	8	0	0	53.80%
	8	0	2	19	24	6	4	43.60%	8	0	1	6	10	0	0	58.80%
	9	0	0	2	10	16	8	43.60%	9	0	0	2	3	7	0	58.30%
	10	0	0	0	0	9	20	69.00%	10	0	0	0	1	2	7	70.00%
	Overall %	17.70%	25.00%	19.30%	17.00%	10.30%	10.70%	49.30%	Overall %	12.00%	23.00%	27.00%	22.00%	9.00%	7.00%	60.00%
T2, T3, T4	5	46	12	0	0	0	0	79.30%	5	11	1	0	0	0	0	91.70%
	6	12	41	9	0	0	0	66.10%	6	3	18	2	0	0	0	78.30%
	7	1	16	29	14	0	0	48.30%	7	0	4	14	8	0	0	53.80%
	8	0	0	9	39	7	0	70.90%	8	0	0	1	16	0	0	94.10%
	9	0	0	0	8	22	6	61.60%	9	0	0	0	3	9	0	75.00%
	10	0	0	0	0	5	24	82.80%	10	0	0	0	0	0	10	100.00%
	Overall %	19.70%	23.00%	15.70%	20.30%	11.30%	10.00%	67.00%	Overall %	14.00%	23.00%	17.00%	27.00%	9.00%	10.00%	78.00%
T2,T3, T4, T5	5	47	11	0	0	0	0	81.00%	5	11	1	0	0	0	0	91.70%
	6	8	50	4	0	0	0	80.60%	6	2	20	1	0	0	0	87.00%
	7	0	9	43	8	0	0	71.70%	7	0	1	17	8	0	0	65.40%
	8	0	0	4	45	6	0	81.80%	8	0	0	3	14	0	0	82.40%
	9	0	0	0	7	23	6	63.90%	9	0	0	0	4	7	1	58.30%
	10	0	0	0	0	3	26	89.70%	10	0	0	0	0	0	10	100.00%
	Overall %	18.30%	23.30%	17.00%	20.00%	10.70%	10.70%	78.00%	Overall %	13.00%	22.00%	21.00%	26.00%	7.00%	11.00%	79.00%
T2, T3, T4, T5, T6	5	55	3	0	0	0	0	94.80%	5	12	0	0	0	0	0	100.00%
	6	4	56	2	0	0	0	90.30%	6	0	21	2	0	0	0	91.30%
	7	0	4	55	1	0	0	91.70%	7	0	0	24	2	0	0	92.30%
	8	0	0	1	54	0	0	98.20%	8	0	0	1	16	0	0	94.10%
	9	0	0	0	4	31	1	86.10%	9	0	0	0	1	11	0	91.70%
	10	0	0	0	0	1	28	96.60%	10	0	0	0	0	0	10	100.00%
	Overall %	19.30%	21.00%	19.30%	19.70%	10.70%	9.70%	93.00%	Overall %	12.00%	21.00%	27.00%	19.00%	11.00%	10.00%	94.00%

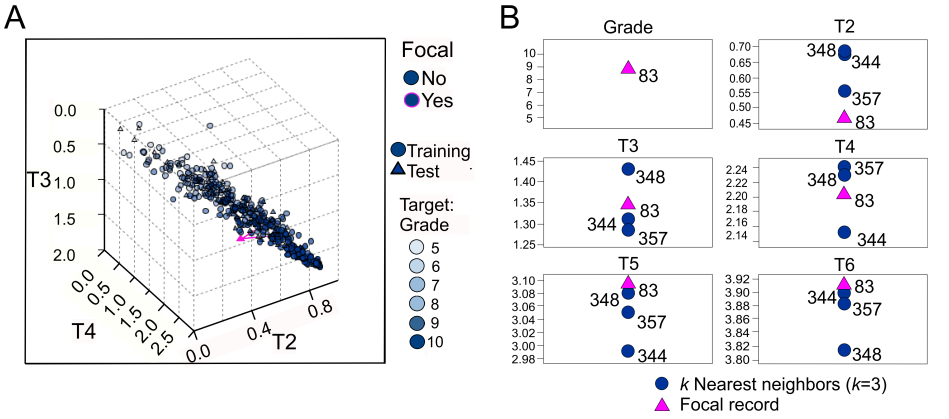


Fig. 6. KNN classification is based on the input values: T2, T3, T4, T5, and T6 with the target variable the student's grade mark. A. Three-dimensional projection of the five-dimensional predictor space representing training and test sets. The pink triangle represents the focal case (student #83) connected to its three nearest neighbors (pink lines); $K = 3$. The classification accuracy was from 93% to 94%; B. Peers chart of the focal record and its three nearest neighbors. The left upper panel shows the final grade (9) of student #83. Panels present the progress function values of the focal case and three nearest neighbors in milestones

Statistical parameters are shown in Tables 4 and 5.

Table 4. Measures for multi-class classification by MLP

Input variables	Set	Pearson's Correlation Coefficient	Cohen's Kappa statistics of agreement	CBA - Class Balance Accuracy	Matthews Correlation Coefficient	Classification quality (agreement)
T2, T3	Training	0.894	0.523	0.567	0.526	Fair
	Test	0.915	0.537	0.559	0.548	
T2, T3, T4	Training	0.941	**0.639**	**0.693**	**0.644**	Substantial
	Test	0.946	**0.707**	**0.716**	**0.712**	
T2, T3, T4, T5	Training	0.973	**0.834**	**0.850**	**0.834**	Near perfect
	Test	0.968	**0.828**	**0.832**	**0.829**	
T2, T3, T4, T5, T6	Training	0.999	**0.996**	**0.995**	**0.996**	Perfect
	Test	1.000	**1.000**	**1.000**	**1.000**	
T3, T4	Training	0.940	**0.640**	**0.689**	**0.640**	Substantial
	Test	0.950	**0.730**	**0.765**	**0.731**	
T3, T4, T5	Training	0.969	**0.810**	**0.834**	**0.810**	Near perfect
	Test	0.963	**0.803**	**0.840**	**0.803**	

Table 5. Measures for multi-class classification by KNN

Input variables	Set	Pearson's Correlation Coefficient	Cohen's Kappa statistics of agreement	CBA Class Balance Accuracy	Matthews Correlation Coefficient	Classification quality (agreement)
T2, T3	Training	0.853	0.299	0.479	0.299	Slight to fair
	Test	0.878	0.505	0.589	0.505	
T2, T3, T4	Training	0.933	0.447	**0.651**	0.447	Moderate to substantial
	Test	0.951	**0.732**	**0.742**	**0.732**	
T2,T3, T4, T5	Training	0.956	0.539	**0.740**	0.539	Moderate to substantial
	Test	0.952	**0.740**	**0.733**	**0.740**	
T2, T3, T4, T5, T6	Training	0.986	**0.915**	**0.913**	**0.915**	Near perfect
	Test	0.986	**0.926**	**0.927**	**0.926**	
T3, T4	Training	0.931	0.519	0.607	0.519	Moderate to substantial
	Test	0.951	**0.731**	**0.757**	**0.731**	
T3, T4, T5	Training	0.953	**0.720**	**0.734**	**0.720**	Substantial
	Test	0.961	**0.790**	**0.824**	**0.790**	

Pearson's Correlation Coefficients showed highly and significantly correlated predicted and observed grades using MLP and KNN methods. Most metric values for multi-class classification performed using MLP were higher than those obtained with KNN.

4. Discussion and Conclusion

Our research primarily focuses on classifying student achievement and its implications on teaching strategies and the learning environment. We emphasize student-centered support

and data-driven strategies to provide personalized feedback to learners. Analyzing student groups helps identify deviations and signals the need to adapt teaching methods for better outcomes.

Based on the classification in [34], we have outlined some features of our research study. We collected performance data from blended learning environments with study focuses on STEM fields. We monitored the academic outcomes of 400 students, which is a common scale for this type of study. We utilized formative and summative assessments, employing predictive modeling and direct methods similar to the approaches used in studies that utilize statistical models and neural networks. This research discusses the classification of students' achievement, specifically predicting their final grades. Our initial assumption implies that all students start learning the subject without prior knowledge. This approach can be adjusted by dividing the class into smaller groups, where an entrance test can be used to assess prior knowledge and incorporate the test result as the value of the progress function at time t_1.

The teaching practice should be modified following the final success prediction of every single student or group. After each time point and prediction of the final grade, students can receive work instructions through direct conversation, which will help them master the course tasks more effectively. It is essential to identify the segment with the weakest results that requires improvement.

For example, let us consider student #83. His progress function value was the lowest among his three closest peers at the time moment t_2 (Fig. 6B). Analyzing his data reveals high performance in class attendance, engagement, and homework completion but poor tests performance. The student was advised to focus on improving this area, with the opportunity for individual consultations with the professor to aid his progress. Following this guidance, student #83 demonstrated better results in knowledge checks and success predictions at times t_3 and t_4 compared to his three closest neighbors. Further recommendations for student #83 were provided based on the prediction results at time t_4. The Fig. 6B indicates that the student embraced the teacher's recommendations, leading to his transformation from the lowest to the highest performer in his group at the control check time t_5 and final exam.

If the majority of the group shows poorer results at milestones t_2 or t_3, it is essential to adjust the teaching approach for the entire group. First, it should be analyzed which specific areas are yielding poor results and tailor the teaching strategies accordingly. For instance, if class attendance is low, the reasons behind this issue should be investigated. This is particularly important when employing a combination of in-person and online teaching methods.

If it becomes evident that students attending in-person classes perform better on tests, efforts should be made to encourage students to attend live classes. Additionally, the online learning experience should be enhanced by addressing the needs and preferences of the students.

If the overall engagement of the group is lacking, it could be considered increasing interaction during lessons to stimulate student participation and awarding for successfully completed homework. If a significant number of students are struggling with their test results, additional consultative sessions for extra support should be organized.

When using MLP and KNN, there was no difference of more than one grade between the observed and predicted grades, except for T2–T3, which means the error will not

exceed one in the grade scale. The prediction accuracy for the input variables T2–T3 is fair because it is early to predict the outcome, given the timeline. For example, some students may perform well at the beginning but later their performance declines. Conversely, some students may not perform well initially but later show improvement. When the values of the progress function at subsequent time points are included, the accuracy of grading and predicting the outcome of the exam increases. Upon comparison, it is evident from Tables 4 and 5 that KNN results in one level lower classification quality than MLP. MLP performed better in prediction, while KNN facilitated the formation of smaller groups for comparative monitoring, despite being less accurate in one level of prediction.

The advantages and weakness of the proposed approach are as follows:

- By including the Promethee II method, the process of monitoring students' progress offers a high level of flexibility. Lecturers can customize various parameters, including course activities, activity priorities, types of preference functions for each activity, and thresholds for these preference functions.
- The data used for predictions is of high quality. It not only reflects students' success but also incorporates the lecturer's attitudes, which can significantly influence outcomes based on their expertise in the subject.
- Implementing the solution is straightforward because the Promethee II method only requires comparisons between two consecutive states, simplifying the overall approach.
- Progress functions can serve as input data for different classifiers.
- The solution is broadly applicable, as it can be adapted to various subjects.
- The proposed method's weakness is its lower accuracy during the initial phase of predicting the final student success.

4.1. Comparison with Existing Approaches

In a review paper [27] that covered publications from 2007 to 2016, it was found that Decision Tree, Rule-based, and Naive Bayes techniques were used in the majority of works to predict students' academic performance. Neural Network and KNN algorithms were used in fewer studies, which was an additional reason to apply these methods. The input data included academic and socioeconomic predictors of success. The meta-analysis determined that the average accuracy is higher when using KNN (87%) compared to using Neural Network (78.7%). Our results showed the opposite, MLP had better performance than KNN.

The authors in [1] used five different machine learning techniques (Naive Bayes, KNN, SVM, XG-boost, and MLP) to predict individual student results. MLP achieved the highest accuracy of 86.25% as in our study when using T2–T5 variables, while other classifiers achieved around 80% accuracy. In [40], they employed seven different classifiers (SVM, KNN, Logistic Regression, Decision Tree, AdaBoost, MLP, and Extra Tree Classifier) to classify students' final grades and they achieved a final accuracy of 81.73%.

Researchers developed a model to analyze IT students' academic performance using data mining techniques: WEKA software and a J48 decision tree to classify success grades [23]. Kappa statistics for this prediction ranged from 0.9070 to 0.9582, but other measures for multi-categorical classification were not assessed. According to Kappa statistics this classification was near perfect. Also in [8], data mining classification techniques such

as J48 Decision Tree, KNN, and MLP were successfully used with the WEKA tool to identify patterns between students' initial grades upon entering the university and their grades at graduation.

The algorithm FlexNSLVOrd, developed by [20], predicts student performance in distance courses by analyzing their online interactions with e-learning platforms using fuzzy systems and ordinal classification. Despite being slower than other algorithms, it has shown better performance in studies.

The study [36] examined dropout rates in online learning environments using lasso and ridge logistic regression. They developed a predictive model based on data from 32,593 students across 22 courses, analyzing 173,912 assessment records. This study focused on early dropout rate predictions at intervals of 30, 45, 60, 90, and 120 days within a course lasting approximately 240 days. They found that the model's AUC improved from 0.549 and 0.661 in the early phase to 0.681 and 0.869 by mid-term. Initially, student demographics and course characteristics were significant predictors, but as the course progressed, student activity became more important. The primary difference between this study and ours is that the former analyzed data from multiple students across several courses over approximately 240 days. In contrast, our research focused on predictions made at five points during a shorter, intensive course. Despite methodological differences, both studies conclude that early predictions are significantly less accurate than those made later.

The article [25] presented a predictive model aimed at identifying students at risk of dropping out during the early stages of their university studies. The researchers analyzed data from 30,576 students enrolled in Higher Education Institutions between 2000 and 2020. They examined the significance of various factors related to dropouts, categorizing them by faculty, degree program, and semester across different predictive models. The findings indicate that variables such as Grade Point Average (GPA), socioeconomic factors, and course pass rates significantly influence the model, regardless of the semester, faculty, or program. Additionally, the study revealed a noteworthy difference in predictive power between Science, Technology, Engineering, and Mathematics (STEM) programs and humanistic programs. Our research was focusing specifically on STEM program.

A critical aspect of the model's accuracy lies in the predictor variables chosen for analysis. Typically, these variables are drawn from students' academic records. One widely used variable is the GPA, often analyzed alongside other grades. For example, the research [26] included both overall GPA and term GPA to predict student dropout rates, while also considering factors such as gender, ethnicity, enrollment status (full-time or part-time), academic classification (freshman or sophomore), and age. Their findings indicated that, although GPA is significantly associated with dropout rates, other variables can also yield strong predictive results.

Additionally, [3] investigated various features influencing student dropouts, including demographic information, family background (such as parents' educational levels), pre-enrollment attributes (like high school GPA and admission test scores), financial circumstances, enrollment details, and academic performance metrics. They specifically analyzed students' GPA, the percentage of credits passed, dropped, and failed, as well as the total credit hours attempted. Their findings indicated that the most significant variables affecting dropout rates were high school GPA, overall GPA, the percentage of failed credits, and various financial factors. In our prediction model, we have used only academic

results in accordance with predefined preference functions, without other pre-enrollment attributes.

4.2. Future Work

The practical application of predictive models relies heavily on their accuracy and reliability. Since these models are based on statistical techniques, uncertainties are always a factor. Therefore, it is important to assess the robustness of the model, particularly regarding confidence in its recommendations, which can be enhanced through sensitivity analysis. Our future research will focus on this issue.

The goals of sensitivity analysis are to determine how the output of a system changes when input parameters are modified, as well as to identify which parameters are the most significant predictors. Numerous studies across various fields address sensitivity analysis for different classifiers, including KNN and MLP. However, our system is heterogeneous, comprising a component that prepares data for prediction and another that contains the classifiers, which adds complexity to the problem.

The input data set is diverse and extensive, making it challenging to identify the input variables for sensitivity analysis. Thus far, we have conducted some research on sensitivity related to preference functions, revealing it to be a very complex issue. Consequently, we may also explore the possibility of conducting a sensitivity analysis specifically for classifiers, using the values of the progress function as input parameters.

Future research could incorporate additional features like pre-enrollment attributes and an entrance test to enhance predictions of student progress.

Acknowledgments. Slađana Spasić was supported by the Ministry of Science, Technological Development and Innovation of Republic Serbia [contract number 451-03-136/2025-03/ 200053]. The authors would like to express their gratitude to Dejan Živković for his assistance in preparing the manuscript for publication.

References

1. Ahammad, K., Chakraborty, P., Akter, E., Fomey, U., Rahman, S.: A comparative study of different machine learning techniques to predict the result of an individual student using previous performances. International Journal of Computer Science and Information Security 19, 5–10 (2021)
2. Alvarado-Uribe, J., Mejía, P., A., Masetto, A., H., Molontay, R., Hilliger, I., Hegde, V., Gallegos, J., Díaz, Ceballos, R., H.: Student dataset from tecnologico de monterrey in mexico to predict dropout in higher education. Data 7, 119 (2022)
3. Ameri, S., Fard, M.J., Chinnam, R.B., Reddy, C.K.: Survival analysis based framework for early prediction of student dropouts. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management. pp. 903–912 (2016)
4. Aulck, L., Velagapudi, N., Blumenstock, J., West, J.: Predicting student dropout in higher education. arXiv preprint arXiv:1606.06364 (2016)
5. Banamar, I., Smet, D., Y.: An extension of promethee ii to temporal evaluations. International Journal of Multicriteria Decision Making 7, No. 3/4, 298–325 (2018)
6. Baradwaj, B., Pal, S.: Mining educational data to analyze students' performance. International Journal of Advanced Computer Science and Applications 2, 63–69 (2011)

7. Barbosa Manhaes, L.M., da Cruz, S.M.S., Zimbrao, G.: Towards automatic prediction of student performance in stem undergraduate degree programs. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. pp. 247–253 (2015)
8. Bawah, U., F., Ussiph, N.: Appraisal of the classification technique in data mining of student performance using j48 decision tree, k-nearest neighbor and multilayer perceptron algorithms. International Journal of Computer Applications 179, 39–46 (2018)
9. Behr, A., Giese, M., Teguim, H., Theune, K.: Motives for dropping out from higher education - an analysis of bachelor's degree students in germany. European Journal of Education 56 (2021)
10. Brans, J.P., Mareschal, B., Figueira, J., Greco, S., Ehrogott, M.: Promethee methods. In: Greco, S., Ehrgott, M., Figueira, J.R. (eds.) Multiple Criteria Decision Analysis: State to the Art Surveys, chap. 5, pp. 163–195. Springer, New York (2005)
11. Chicco, D., Jurman, G.: The advantages of the matthews correlation coefficient (mcc) over f1score and accuracy in binary classification evaluation. BMC Genomics 21, No. 6 (2020)
12. Chung, J.Y., Lee, S.: Dropout early warning systems for high school students using machine learning. Children and Youth Services Review 96, 346–353 (2019)
13. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13, No. 1, 21–27 (1967)
14. Daniel, B.: Big data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology 46, No. 5 (2014)
15. Dixson, D., Worrell, F.: Formative and summative assessment in the classroom. Theory Into Practice 55, 14 (2016)
16. Fix, E., Hodges, J.L.: Discriminatory analysis, nonparametric discrimination: Consistency properties. Tech. Rep. Technical Report 4, USAF School of Aviation Medicine, Randolph Field (1951)
17. Fleiss, J., Cohen, J., Everitt, S., B.: Large sample standard errors of kappa and weighted kappa. Psychological Bulletin 72, 323–327 (1969)
18. Gasevic, D., Dawson, S., Rogers, T., Gasevic, D.: Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education 28, No. 1. 68–84 (2016)
19. Grandini, M., Bagli, E., Visani, G.: Metrics for multi-class classification: an overview. arXiv preprint abs/2008.05756 (2020)
20. Gámez-Granados, J.C., Esteban, A., Rodriguez-Lozano, J., F., Zafra, A.: An algorithm based on fuzzy ordinal classification to predict students' academic performance. Applied Intelligence 53, 27537–27559 (2023)
21. Hellas, A., Liao, S., Ihantola, P., Petersen, A., Ajanovski, V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C.: Predicting academic performance: a systematic literature review. In: Proceedings of the Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. pp. 175–199. Larnaca, Cyprus (2018)
22. Huang, S., Fang, N.: Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. Computers & Education 61, 133–145 (2013)
23. Ibrahim, W., Abdullaev, S., Alkattan, H., Oluwaseun, A., Alkattan, H., Alhumaima, A.: Development of a model using data mining technique to test, predict and obtain knowledge from the academics results of information technology students. Data 7, No. 5, 18 (2022)
24. Jesus, C., F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., Rosa-Louro, A.: Using artificial intelligence methods to assess academic achievement in public high schools of a european union country. Heliyon 6 (2020)
25. Jimenez-Macias, A., Moreno-Marcos, M., Merino, P., Ortiz, M., Delgado-Kloos, C.: Analyzing feature importance for a predictive undergraduate student dropout model. Computer Science and Information Systems 20, 50–50 (2022)
26. Kang, K., Wang, S.: Analyze and predict student dropout from online programs. In: Proceedings of the 2nd International Conference on Compute and Data Analysis. pp. 6–12 (2018)

27. Kumar, M., Singh, A., Handa, D.: Literature survey on student's performance prediction in education using data mining techniques. International Journal of Education and Management Engineering 6, 40–49 (2017)
28. Matthews, W., B.: Solvent content of protein crystals. Journal of Molecular Biology 33, No. 2, 491–497 (1968)
29. Matthews, W., B.: Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure 405, No. 2, 442–451 (1975)
30. Maura, E., A., P., Nazeeruddin, E., Nazeeruddin, M., Daqqa, I., Abdelsalam, H., Abdullah, M.: Is initial performance in a course informative? machine learning algorithms as aids for the early detection of at-risk students. Electronics 11, No. 13, 2057 (2022)
31. Moreno-Marcos, P.M., Laet, D., T., Munoz-Merino, P.J., Soom, V., C., Broos, T., Verbert, K., Kloos, D., C.: Generalizing predictive models of admission test success based on online interactions. Sustainability 11, No. 18, 4940 (2019)
32. Mosley, L.: A balanced approach to the multi-class imbalance problem. Phd dissertation, Iowa State University, USA (2013)
33. Mthimunye, K., Daniels, of academic performance, F.P., success andretention amongst undergraduate nursing students: A systematic review. South African Journal of Higher Education 33, 200–220 (2019)
34. Namoun, A., Alshanqiti, A.: Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences 11, 1–28 (2020)
35. Popescu, M.C., Balas, V., Perescu-Popescu, L., Mastorakis, N.: Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems 8, No. 7 (2009)
36. Radovanović, S., Delibašić, B., Suknović, M.: Predicting dropout in online learning environments. Computer Science and Information Systems 18, No. 3, 957–978 (2021)
37. Rastrollo-Guerrero, J., Gomez-Pulido, A., J., Domínguez, A.: Analyzing and predicting students' performance by means of machine learning: A review. Applied Sciences 10, 1042 (2020)
38. Rosenblatt, F.: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1st edn. (1961)
39. Zhang, L., Li, K.F.: Education analytics: Challenges and approaches. In: Proceedings of the 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). pp. 193–198. IEEE, Krakow, Poland (2018)
40. Zulfiker, S., Kabir, N., Biswas, A., A., Chakraborty, P., Rahman, M.: Predicting students' performance of the private universities of bangladesh using machine learning approaches. International Journal of Advanced Computer Science and Applications 11, No. 3, 672–679 (2020)

Slađana Spasić completed her B.Sc. in Numerical Mathematics and Cybernetics in 1991, followed by a M.Sc. in Artificial Intelligence in 2003, and a Ph.D. in Applied Mathematics in 2007, all from the University of Belgrade, School of Mathematics, Serbia. From 2008 to 2013, she served as an assistant professor at the University of Belgrade - Institute for Multidisciplinary Research (IMSI) and Singidunum University. She was promoted to associate professor from 2013 to 2018, and she has been a full professor at both institutions since 2018. Starting in 2024, she will join IMSI. Her research interests include the structural aspects of biological signals and images, as well as the development of statistical and mathematical models in ecology, physiology, and other related fields. She published over 50 scientific papers on SCI list and 2 textbooks.

Violeta Tomašević received her B.Sc., M.Sc., and PhD degrees from the School of Electrical Engineering, University of Belgrade, Serbia, in 1988, 1994, and 2005, respectively.

From 1989 to 2007, she was with the "Mihajlo Pupin" Institute, Belgrade. In 2007, she joined the University Singidunum, Belgrade, where she is currently a Full Professor with the Informatics and Computing Department. Her research interests include cryptanalysis, expert systems, decision support systems, data retrieval and web programming. She published over 60 scientific papers and 3 textbooks.

Received: October 22, 2024; Accepted: April 10, 2025.

Computer Science and Information Systems 22(3):907–926 https://doi.org/10.2298/CSIS240713051T

Image Semantic Segmentation Based on Multi-layer Feature Information Fusion and Dual Convolutional Attention Mechanism

Lin Teng[1], Yulong Qiao[1,⋆], Jinfeng Wang[2], Mirjana Ivanović[3], and Shoulin Yin[1,4]

[1] School of Information and Communication Engineering, Harbin Engineering University
145 Nantong Street, Nangang District, Harbin, 150001, China
qiaoyulong@hrbeu.edu.cn
[2] Weifang Vocational College
Weifang, 261041 China
1057417325@qq.com
[3] Faculty of Sciences, University of Novi Sad, Serbia
mira@dmi.uns.ac.rs
[4] Software College, Shenyang Normal University
Shenyang, 110034 China
yslin@hit.edu.cn

Abstract. Traditional semantic segmentation methods have problems such as poor multi-scale feature extraction ability, weak lightweight backbone network feature extraction ability, lack of effective fusion of context information, resulting in edge segmentation errors and feature discontinuity. In this paper, a novel semantic segmentation model based on multi-layer information fusion and dual convolutional attention mechanism is proposed. In this method, SegFormer network is used as the backbone network, and multi-scale features of encoder output are fused with overlapping features. The feature extraction subnetwork is optimized by constructing the object region enhancement module, and the intermediate feature map is refined adaptively in each convolutional block of the deep network, so as to strengthen the fine extraction of multi-dimensional feature information of complex images. Dual convolutional attention module is used to fusion high-level semantic information to avoid the loss of feature information caused by up-sampling operation and the influence of introducing noise, and refine the effect of target edge segmentation. At the same time, the feature pyramid grid is proposed to process the overlapping features, obtain the context information of different scales, and enhance the semantic expression of features. Finally, the features processed by the feature pyramid grid module are combined to improve the segmentation effect. The experimental results on the public data set show that the proposed method has better performance than the existing methods, and has better segmentation effect on the object edge in the scene.

Keywords: Semantic segmentation, multi-layer information fusion, dual convolutional attention mechanism, feature pyramid grid.

⋆ Corresponding author

1. Introduction

Image semantic segmentation is one of the important research topics in the field of computer vision. Different from object detection and image classification, semantic segmentation processes images at the pixel level, and each pixel is assigned a corresponding label [1,2]. In the field of self-driving, semantic segmentation can segment roads, buildings, pedestrians, obstacles, etc., and give pixel-level annotations for each category. In the medical field, semantic segmentation can segment the location of lesions to assist doctors in diagnosis.

With the development of Convolution Neural Networks (CNN) [3], image semantic segmentation has developed rapidly. Fully Convolutional Networks (FCN) [4] replaced the fully connected layer with the convolutional layer to realize end-to-end semantic segmentation, so that the input image of semantic segmentation did not need a fixed size, which provided a foundation for subsequent methods in the field of semantic segmentation. Reference [5] proposed SegNet (Segmentation Networks), which recorded the positions during the maximum pooling operation during feature extraction and used the position information during up-sampling, but only the maximum position information was saved in this operation, and more information was still lost. The backbone of PSPNet (Pyramid Scene Parsing Networks) proposed in reference [6] adopts ResNet (Residual Networks) and introduced PPM for multi-scale information fusion, which could effectively solve the class confusion problem. DeepLab v3+ was introduced in reference [7] based on DeepLab series models. In the backbone part of this model, deep separable convolution and dilated convolution were adopted to replace traditional convolution and effectively alleviated the problem of parameter number. At the same time, DeepLabV3+ introduced Atrous Spatial Pyramid Pooling (ASPP) [8] to extract multi-scale features, which significantly improved the overall image segmentation accuracy.

The attention-mechanism-based approach is flexible in capturing the connections between global and local information. SENet is a network with pure channel attention mechanism [9], which obtains the weights of different channels through autonomous learning, thus expressing the importance of feature channels through different weights, and modeling the dependencies between channels. CBAM is a simple and effective mixed domain attention module, which processes feature maps in spatial domain and channel domain, and obtains good segmentation results by simple operation. PSANet [10] connects each location in the feature map with other locations through an adaptive learning attention mask to obtain contextual information. At the same time, the bidirectional information propagation path is designed, and the information collected from other locations is used to assist in predicting the current location, which is an efficient method. EMANet effectively aggregates features by introducing a multi-scale information fusion module [11], which uses feature maps at various scales to generate feature fusion results by weighted summation. However, such methods have high computational complexity, easy over-fitting and limited processing of long sequences, so the accuracy of complex images needs to be improved.

Encoder-decoder-based approaches [12,13] typically choose VGGNet [14] as the encoder and then replace its fully connected layer with the decoder structure. FCN solves the problem that convolutional neural networks limit the input image size, but the segmentation results are still rough. SegNet records the spatial position, so that it can accurately recover the image in the up-sampling stage, but the segmentation accuracy of the object

boundary is still not high. UNet [15] introduces fast connections between encoders and decoders, which has achieved good results on medical and remote sensing images, but the application scenarios of this network are limited. DFANer's encoder innovatively uses three sets of Xception networks [16], and the proposed subnet aggregation module optimizes the results. PointRend [17] regards image segmentation as a rendering problem in image processing, refins the rough mask edges generated by the network before, and implements the point-based segmentation module at the adaptive selected position. It can be seen that the accuracy of this kind of network in the segmentation of complex street view images needs to be improved.

The above methods consider the mining of context information and multi-scale information, but do not mine the correlation of semantic information between pixels, resulting in classification errors in complex scenes. Attention mechanism is a kind of mechanism that can establish the dependency relationship between pixels, which is introduced into the semantic segmentation network and plays an important role. SENet (Squeeze and Excitation Networks) was proposed in reference [18], which proved that the attention mechanism could reduce noise while improving classification performance. In reference [19], DANet (Dual Attention Networks) was proposed for semantic segmentation. Self-attention mechanism was also a kind of attention mechanism, self-attention mechanism calculated the similarity between features, and used this to capture the dependency between pixels. However, the computational overhead and memory overhead of self-attention mechanisms were squared complexity. In order to reduce the cost of computation and memory, reference [20] proposed CCNet (Criss-cross Networks), that is, each calculation only considered the row and column where the current pixel was located, and then indirectly connected the global information through the cascade of two cross-attention modules. Reference [21] proposed EMANet (Expectation maximization Attention Networks), which used expectation maximization clustering to optimize the attention mechanism and reduce the computational overhead. Reference [22] proposed EANet(External Attention Networks), which used two linear layers as the K and Q of the attention mechanism to represent the attention mechanism, reducing the computational cost and memory overhead, and the two linear layers could indirectly interact with the global information. Reference [23] proposed the Convolutional Block Attention Module (CBAM), which blended channel attention and spatial attention. Reference [24] rethought the aforementioned attention and proposed coordinate attention, which could achieve better performance improvement with almost no additional computing overhead.

With the application of semantic segmentation in practical projects, researchers have gradually turned their attention to the lightweight models rather than the accuracy of models. EdgeNeXt [25] attempted to combine the advantages of ViTs (Vision Transformers) with traditional convolution by introducing Split Depth-wise Transpose attention (SDT) to efficiently combine the advantages of ViTs and CNN without adding additional parameters and computation. ICNet (Image Cascade Networks) [26] used complex and deep paths to encode small size inputs. The MobileNets series [27,28] used deep separable convolution to replace traditional convolution operations. Xception built an entirely new network architecture using deep separable convolution. The above works focus on the lightweight of the model, which allows the segmentation model to run faster, but the accuracy limits its widespread use in practical applications.

Based on the above analysis, in order to further improve the accuracy of complex image semantic segmentation, a novel semantic segmentation model based on multi-layer information fusion and dual convolutional attention mechanism is proposed, the main contributions are as follows:

1. In this method, SegFormer network is used as the backbone network, and multi-scale features of encoder output are fused with overlapping features.
2. The feature extraction subnetwork is optimized by constructing the object region enhancement module, and the intermediate feature map is refined adaptively in each convolutional block of the deep network, so as to strengthen the fine extraction of multi-dimensional feature information of complex images.
3. The feature extraction subnetwork is optimized by constructing the object region enhancement module, and the intermediate feature map is refined adaptively in each convolutional block of the deep network, so as to strengthen the fine extraction of multi-dimensional feature information of complex images.
4. Dual convolutional attention module is used to fusion high-level semantic information to avoid the loss of feature information caused by up-sampling operation and the influence of introducing noise, and refine the effect of target edge segmentation. At the same time, the feature pyramid grid is proposed to process the overlapping features, obtain the context information of different scales, and enhance the semantic expression of features.
5. Finally, the features processed by the feature pyramid grid module are combined to improve the segmentation effect.

2. Related Works

The semantic segmentation method based on deep learning usually adopts the encoder-decoder structure. Firstly, the image is input to the feature extraction network, and then the extracted semantic features are sent to the decoder to analyze the semantic features and obtain the segmentation graph.

The concept of the attention mechanism stems from the human ability to process external information, that is, when faced with a large amount of information, people will focus their attention on information that is important to them and ignore irrelevant information. In computer science, this mechanism is known as attention, by giving different weights to information, so that the computer can pay more attention to the important parts of the information.

Compared with the traditional attention mechanism, the self-attention mechanism reduces the dependence on external information, is better at capturing the internal correlation of data or features. It can effectively extract global semantic information, thus improving the performance of the model. The self-attention mechanism is calculated as follows:

$$Q = W^Q X; K = W^K X; V = W^V X \tag{1}$$

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2}$$

Where, Q (Query) is query vector matrix), K (Key) is key vector matrix) and V (Value) is value vector matrix), they are important matrices in self-attention mechanism. $\sqrt{d_k}$ is the dimension of the key vector. Specifically, it first multiplies the input vector X by the three corresponding initialization weight matrices W^Q, W^K, and W^V respectively to get Q, K, and V. Second, it calculates the dot product of Q and K to obtain the correlation between each of the two input vectors, that is, the similarity. To avoid the result of the dot product being too large or too small, the similarity is divided by $\sqrt{d_k}$. Then, the results are normalized using the softmax function to get the attention weights. Finally, these attention weights are dotted with V to get the output of the self-attention mechanism. In this way, the self-attention mechanism can extract and utilize the global information in the input sequence, and show excellent performance in processing tasks in fields such as natural language processing.

To further improve the performance of the model, Transformer introduces multi-head attention (MHA). This mechanism extends the original attention mechanism to allow the model to focus on different parts of the input sequence to obtain more contextual information. Specifically, the input word vector is transformed by different linear transformations to generate different Q, K and V, and then the self-attention operation is carried out to get the output result. Then, the output of all the "heads" is spliced, and finally through a linear transformation, the final output of the multi-head attention mechanism is obtained. This process takes place independently in multiple "heads," each with its own unique weight matrix. The advantage of this approach is that the model can focus on different features in different representation subspaces, and then integrate these features to get a richer feature representation, and make full use of computing resources, accelerate the training and reasoning process of the model, so as to improve the overall performance of the model.

3. Proposed Image Semantic Model

In this paper, the SegFormer network encoder structure is adopted to optimize the multi-stage feature fusion mode, and the feature pyramid grid module is designed and implemented to improve the feature discontinuity problem and improve the segmentation accuracy of the network by combining the overlapping convolutional dual convolutional attention mechanism to highlight important feature information.

SegFormer network adopts encoder-decoder structure, encoder structure consists of four stages. Each stage is stacked by multiple Transformer blocks. In each Transformer block, overlap patch embeddings module is used for feature extraction of input images. The extracted features are then used to calculate the feature correlation through the efficient multi-head self-attention module. Finally, the calculated features are passed through the mix-feed forward network (MixFFN) module [29]. SegFormer replaces position encoding in MixFFN and uses a depth-separable convolution with a convolution kernel size of 3×3 to provide position information. In the decoder part of the network, the resolution of the four stages of the encoder is 1/4, 1/8, 1/16 and 1/32 of the original image. The number of channels in the feature graph is uniformly adjusted to 256 by a convolution layer with convolution kernel size of 1×1. Then bilinear interpolation is used to up-sample the feature map to the 1/4 size of the original image, and multiple feature maps of the same size are spliced. Two convolution layers with convolution kernel size of 1×1 are used to complete pixelation-level prediction. Finally, bilinear interpolation is used to

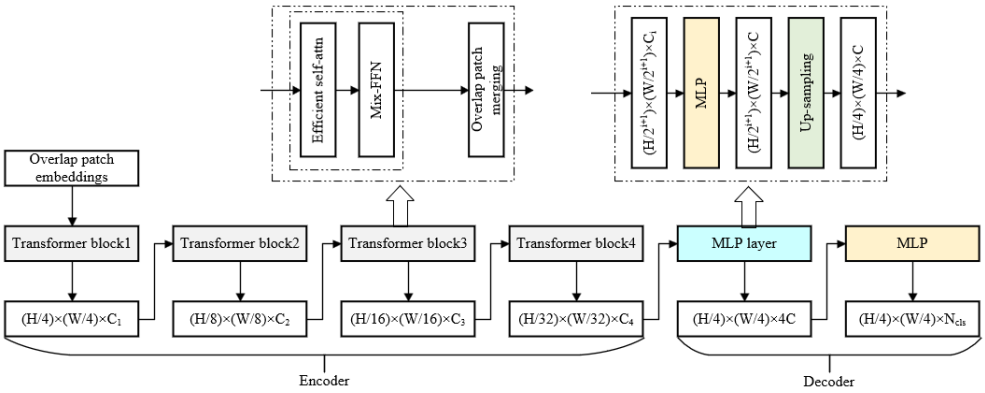


Fig. 1. SegFormer Network structure.

restore the image to the original image size and output the final segmentation result. The segmentation process is shown in Figure 1.

Although the SegFormer network performs well in scene segmentation, it up-samples multiple stage feature maps to the same size at one time in its decoder part, which easily results in insufficient fusion of low-level details and high-level semantic information. At the same time, the high level feature and the low level feature are directly spliced, which inevitably introduces noise and leads to the decrease of segmentation accuracy. In addition, SegFormer network does not introduce multi-scale context processing structure, which is easy to lead to discontinuous features and edge segmentation errors between objects of different scales, resulting in segmentation accuracy problems. Based on the above analysis, this paper proposes innovative improvements in three aspects. The first is to optimize the feature extraction subnetwork by constructing the object region enhancement module (OREM). The second is to use the overlapping feature fusion method to replace the original structure directly connected to the sampling and then merged, and to fuse the low-level features with the high-level features in stages. The dual convolutional attention module is used to calibrate the features of the high-level semantic information before the high-low feature fusion, and to suppress the interference of the low-level redundant information and the noise introduced by the up-sampling process. Third, depthwise separable residual convolution modules are proposed. On this basis, the residual feature pyramid grid (RFPG) is designed and implemented. The semantic information after the fusion of overlapping features is obtained through the feature pyramid grid modules of different scales to obtain the context information of different stages, strengthen the semantic association between features, improve the feature discontinuity problem, and improve the image segmentation accuracy.

3.1. Overlapping Feature Fusion

In this paper, the encoder-decoder structure is followed to build the network model, and the encoder adopts MiT-B2 of SegFormer model as the backbone network. The overall

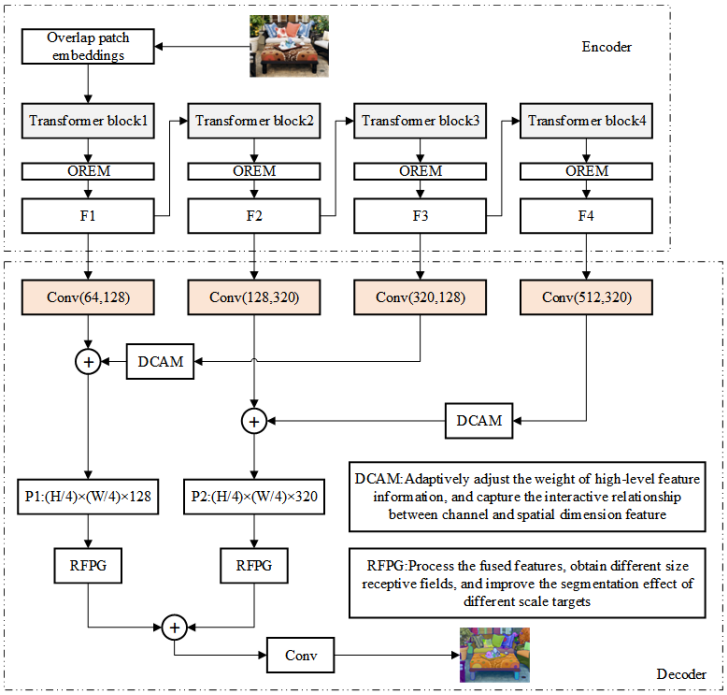


Fig. 2. The proposed model structure in this paper.

architecture of the model is shown in Figure 2. First, an object region enhancement module is established, and the feature maps of each stage are fused with overlapping features to preserve the rich details of low-level features and reduce the interference of noise on high-level semantics. The dual convolutional attention module (DCAM) is used to calibrate the high level feature weights, and then the two parts of the fused results are fed into the RFPG multi-scale feature extraction module with different sizes to extract multi-scale information of different granularity. Then the multi-scale information is merged to improve the segmentation effect of different scales, and the merged multi-scale information is up-sampled by bilinear interpolation to restore the input image size. Finally, the number of channels is adjusted to the number of categories by 1×1 convolutional layer, and the output result of the network is obtained.

As shown in Figure 1, in the decoder part of the original SegFormer network structure, feature graphs of different sizes output by each stage of the encoder are sampled to a unified size, then directly splicing them together, and the final segmentation graph is obtained by adjusting the number of channels through 1×1 convolution. Although this approach has fewer parameters and lower operation cost, it directly fuses low-level detailed features with high-level semantic features, which enricfies the detailed information and inevitably introduces a lot of noise, which affects the segmentation accuracy. For this reason, this paper, based on the MiT-B2 backbone network, carries out overlapping fusion of the extracted feature maps at each stage. As shown in Figure 2, the feature maps of different stages obtained by the backbone network are named F1, F2, F3 and F4, and the overlap-

ping feature fusion is divided into two parts. In P1, the feature map F3 is first adjusted by 1×1 convolution to adjust the number of channels, and the size of the feature map is obtained as $[H/16, W/16, 128]$. Then, the feature map resolution is sampled from 1/16 of the original image to 1/4 of the original image, and the relationship between features is adjusted by the attention module DCAM to enhance the high-level feature channels and spatial information. At the same time, the feature graph F1 is adjusted by 1×1 convolution to get the number of channels, and the size of the feature graph is $[H/4, W/4, 128]$. The two parts of the feature graph are added together to get the output of the first part. Similarly, in the second part of the overlapping fusion, the feature figure F4 is first adjusted to 320 channels by 1×1 convolution, then up-sampled by four times bilinear interpolation, and then added to the feature figure F2 with 320 channels by the attention module DCAM to obtain the output of the second part. Since the feature maps output at different stages of the encoder contain different feature information, the spatial position details of the feature maps F1 and F2 are rich in comparison, which helps to improve the segmentation effect of details such as target edges. However, semantic information is relatively insufficient, while the abstract semantic information contained in feature figures F3 and F4 is rich but lacks spatial details. If multiple feature maps are fused directly, although the detailed information of high-level features is enriched, the low-level feature maps also have a lot of noise in addition to the detailed information, and direct fusion is not conducive to improving the accuracy. The overlapping feature fusion method proposed in this paper fuses the feature maps F1 and F4 with those of the transition stages F3 and F2, and uses dual convolutional attention module DCAM to adaptively adjust the weight of high-level features, which not only realizes the fusion of high-low layer features, but also avoids the interference of lower-level noise on the segmentation results.

3.2. Object Region Enhancement Module (OREM)

The convolutional layer can automatically extract the multi-dimensional feature information from the original data by learning, but the multi-dimensional feature information obtained by most networks is limited. In order to obtain more multi-dimensional feature information and further solve the problem of low precision caused by different object scales in segmentation, this paper constructs an object region enhancement module (OREM), the structure is shown in Figure 3. The OREM module consists of THREE convolution blocks and a DAM, each convolution is connected by a Relu layer and connected by a residual structure, and finally output through a Relu layer [30]. OREM module adaptively refines the intermediate feature graph at each convolution block of feature extraction subnetwork, and dynamically processes the information of each feature graph to help the model learn the features of the data better. At the same time, the features with higher importance can be better expressed, the fine extraction of multi-dimensional features of complex street view images is strengthened, and the sub-network of feature extraction is optimized, which further solves the problem of low segmentation accuracy due to different target scales, occlusion overlap and illumination changes in segmentation, and is suitable for segmentation of complex images.

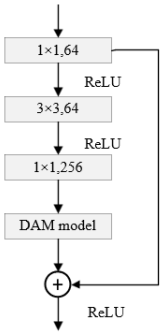


Fig. 3. OREM structure.

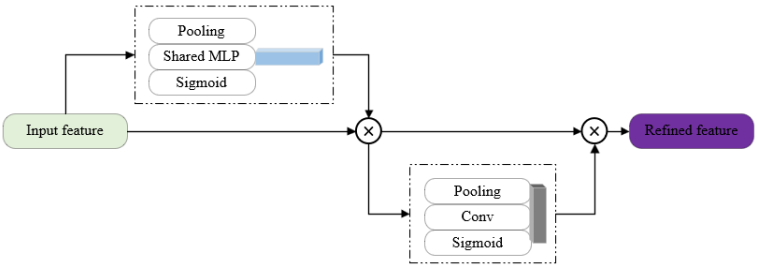


Fig. 4. DCAM structure.

3.3. Dual Convolutional Attention Module (DCAM)

Deep networks can obtain contextual information, but there is no focus on the region and information, resulting in segmentation accuracy has not been a big breakthrough. Therefore, this paper proposes the dual convolutional attention module (DCAM). The DCAM structure is shown in Figure 4. DCAM is a lightweight mixed-domain attention module that improves the accuracy of the network by learning the interrelationships between features. After the feature map is given, DCAM inferences the attention map sequentially along two independent dimensional channel domains and spatial domains, and adaptively weights different features, thereby increasing the weight of useful information and reducing noise and unimportant information.

The role of DCAM in the network is to input the feature map F into the channel attention module, and the input feature map is processed to get the corrected F'. Then the resulting F' is processed with the input F and re-input the spatial attention module to get the corrected feature diagram F”:

$$F' = M_C(F) \otimes F \tag{3}$$

$$F'' = M_S(F') \otimes F' \tag{4}$$

Where C is a channel. M_C is channel attention. M_S is spatial attention. F is the feature graph.

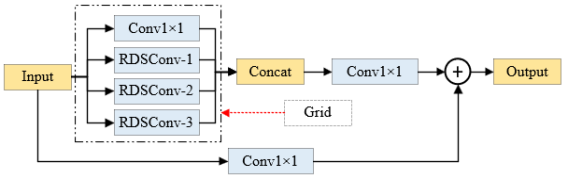


Fig. 5. RFPG module.

3.4. Residual Feature Pyramid Grid (RFPG)

Because the object scale is not uniform in the scene, it is easy to cause the segmentation accuracy is not high, edge segmentation errors and other problems. Therefore, it is very important to obtain multi-scale context information of feature maps to extract features of different scales for improving segmentation accuracy and alleviating feature discontinuity caused by differences between objects of different scales. Therefore, the RFPG module is designed and implemented in this paper. As shown in Figure 5, the RFPG module is mainly built using the depth-separable residual convolution module in this paper. First, the use of depth-separable convolution can reduce the number of parameters in the model, which is conducive to the lightweight of the model, but at the same time, depth-separable convolution separates the spatial dimension operations from the channel dimension operations, which easily leads to the problem of discontinuity between features. Therefore, in this paper, depth-separable residual convolution is proposed to unify the spatial and channel dimensional features, enhance the semantic expression between features without introducing additional parameters, and avoid the phenomenon of gradient disappearance or gradient explosion caused by deepening network hierarchy.

The RFPG module is mainly composed of 1×1 convolution and three RDSConv modules with different expansion rates. The common depth separable convolution operation (DSConv) [31,32] is composed of two parts. First, the input feature graph is separated by the expansion convolution with the convolution kernel size of 3×3, and then the channel dimensions are merged by 1×1 convolution. The specific calculation formula of DSConv is shown in equation (3):

$$output = Conv_{1\times1}(DConv_{3\times3}(input)) \tag{5}$$

$input$ indicates the input feature map of the module. $output$ indicates the output feature map. $Conv1 \times 1$ represents an ordinary convolution with a convolution kernel size of 1×1. $DConv3 \times 3$ represents an expansive convolution with a convolution kernel size of 3×3.

This paper proposes to introduce residual operation into ordinary depth-separable convolution, that is, after performing 3×3 dilated convolution operation on the feature graph, adding the convolution result to the input, and then obtaining the output through 1×1 convolution, which is different from $DSConv$ in that space and channel are operated separately. $RDSConv$ combines spatial dimension operations with channel dimension operations to enhance semantic feature association and alleviate the problems of discontinuous target segmentation and unclear edge segmentation. The specific calculation process of $RDSConv$ is shown in equation (4):

Table 1. RFPG module parameters

Stage	Size of the input feature graph	Expansion rate setting	Input channel number	Output channel number
P1	$(H/4, W/4)$	(1,3,6,9)	128	256
P2	$(H/8, W/8)$	(1,6,12,24)	320	256

$$output = Conv_{1\times1}(input + DConv_{3\times3}(input)) \tag{6}$$

In the RFPG module, the input feature map is concatenated by 1×1 convolution and three $RDSConv$ models with different expansion rates, and the obtained results are concatenated with channel dimensions, and then adjusted to the specified number of channels by 1×1 convolution. At the same time, 1×1 convolution is used to adjust the input feature diagram to the specified number of channels, and the output of the RFPG module is obtained by adding the two together. The specific operation process is shown in Figure 5.

As shown in Figure 2, two parts of output are obtained after overlapping feature fusion of multi-stage encoder feature maps. The first part is obtained by the fusion of feature maps F1 and F3, and the second part is obtained by the fusion of feature maps F2 and F4. In view of the different information contained in these two features, RFPG modules with different expansion rates are used to extract multi-scale features. For the first part, the expansion rate in the RFPG module used is (1,3,6,9); For Part 2, the expansion rate used is (1,6,12,24), and an expansion rate of 1 means that a common convolution of 1×1 is used. The output channel of each convolutional module is set to 256, as shown in Table 1.

For the first part, RFPG module is constructed with a small expansion rate, mainly to avoid the interference of redundant information as far as possible, aiming at the characteristics of rich detailed information in its characteristics. For the second part, which is rich in semantic information, a larger expansion rate is used to increase the receptive field of the network, obtain semantic information at different scales, and improve the segmentation accuracy of the network.

For the two-part feature maps obtained by the fusion of overlapping features, the two-part feature maps with sizes $[H/4, W/4, 256]$ and $[H/8, W/8, 256]$ are obtained after the respective RFPG modules. Then, the feature map of Part 2 is up-sampled with two times bilinear interpolation, and the size of the feature map obtained is $[H/4, W/4, 256]$, which is the same dimension as the feature map of Part 1. Finally, the feature graphs after the fusion of the two parts are added, and the channels of the feature graphs are adjusted by bilinear interpolation, and then the output result of the network is obtained by 1×1 convolution.

3.5. Loss Function

In the field of image semantic segmentation, the loss function has many forms, the most commonly used is the cross-entropy loss function. The function calculates the loss by measuring the similarity between the predicted distribution and the true distribution. The predicted distribution is close to the true distribution, so the smaller the loss function value is smaller. Otherwise it is larger. Its expression is shown in equation (5):

$$Loss_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(p_{ij}) \tag{7}$$

Where N is the number of samples. C is the number of categories. y_{ij} is the label where sample i belongs to class j, with a value of 0 or 1. p_{ij} is the probability that the model predicts the sample to be of class j, with a value between 0 and 1. The limitation of the cross-entropy loss function is that it does not take into account the unbalanced distribution of labels. When the number of pixels of different categories is very different, the training of the loss function will become more difficult. In addition, the cross-entropy loss function [33] only calculates the loss value of each pixel discretely and then averages it, rather than considering the prediction result of the whole image globally. In order to make up for the deficiency of cross entropy loss function, Dice loss function and its variant Tanimoto loss function are introduced, whose expressions are shown in equation (6) and equation (7) respectively:

$$Loss_{Dice} = 1 - \frac{2\sum_{i=1}^{N}\sum_{j=1}^{C}(y_{ij}p_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{C}(y_{ij}+p_{ij})} \tag{8}$$

$$Loss_{Tanimoto} = 1 - \frac{\sum_{i=1}^{N}\sum_{j=1}^{C}(y_{ij}p_{ij})}{\sum_{i=1}^{N}\sum_{j=1}^{C}(y_{ij}+p_{ij}-y_{ij}p_{ij})} \tag{9}$$

The Dice loss function and Tanimoto loss function are numerically equivalent, and both help to solve the problem of difficult training when label distribution is unbalanced. However, when there are more small targets, the loss function is prone to oscillation, and even gradient saturation occurs in extreme cases. Furthermore, as a rule of thumb, a loss function with a quadratic in the denominator is more likely to bring the prediction closer to the true value, regardless of the random initial value of the weight. Therefore, the weighted sum of cross entropy loss function and Tanimoto loss function is selected as the overall loss function, whose expression is shown in equation (8):

$$Loss_{total} = Loss_{CE} + \gamma Loss_{Tanimoto} \tag{10}$$

Where γ is the parameter that balances the effects of the cross-entropy loss function and the Tanimoto loss function, and its value ranges from $(0, +\infty)$.

4. Experimental Results and Analysis

4.1. Data Sets

The performance of the proposed semantic segmentation model is evaluated on two data sets, PASCAL VOC 2012 and Cityscapes. The PASCALvOC 2012 data set is used for training and performance evaluation of the model, and the Cityscapes data set is used for generalization performance testing of the model.

PASCAL VOC 2012 is a widely used public image data set in computer vision. The data set has 21 semantic categories, including 20 object classes and one background class. There are 2913 tagged images. Among them, 2700 images are randomly selected as the

training set, 300 images are selected as the verification set, and the image size of the input semantic segmentation network is set to 512×512.

Cityscapes is one of the well-known data sets of autonomous driving scenarios in urban environments. The data set has 34 semantic categories, of which only 19 are used based on previous work. There are a total of 5000 finely labeled images, each with a resolution of 1024×2048, and in order to be consistent with the VOC2012 data set, 4500 images randomly selected from the data set are used as the training set and 500 images are used as the validation set. The image size of the input semantic segmentation network is set to 512×1024.

4.2. Evaluation Index

In this paper, mean intersection over union (MIoU) and mean accuracy (MAcc) are used to evaluate the experimental results, the expressions are shown in equations (9) and (10):

$$MIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{11}$$

$$MAcc = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{12}$$

Where N represents the total number of categories. TP_i is the number of pixels correctly predicted by category i. FP_i is the number of pixels that predict the other category as category i. FN_i is the number of pixels that predict category i as other categories. $MIoU$ is to calculate the intersection ratio of each semantic class separately, and calculate the average value after summing. $MAcc$ is to calculate the accuracy of each semantic class separately, and calculate the average value after summing.

4.3. Experiment Settings

This experiment is implemented on PyTorch framework, the operating system is Windows11 64-bit operating system, the processor is Intel(R) Xeon(R)Gold 5218R, the graphics card is NVIDIA A10. The memory is 128 GB, and the hard disk is 1TB.

Adam algorithm is used as the optimizer, and the momentum is set to 0.8. Applying the "poly" learning rate strategy, the learning rate gradually decreases with the increase of the number of iterations, and its expression is shown in equation (11).

$$lr - current = lr - initial(1 - \frac{T}{T_{max}})^{mom} \tag{13}$$

Where $lr - current$ is the current learning rate. $lr - initial$ is the initial learning rate, which is defined as 5×10^{-4}. T indicates the current number of iterations and T_{max} indicates the maximum number of iterations. mom is the momentum, it is 0.9. Also, it sets the batch size for each training round to 16, the number of training rounds to 200. In the first 100 rounds of training, the parameters of the backbone network are frozen so that it does not participate in the training, and in the 101-200 rounds of training, the backbone network is thawed.

Table 2. Objective index calculation results

Backbone	MIoU	MAcc
MobileNetv2	77.34/%	86.23/%
ResNet101	80.37/%	89.14/%
Xception	82.74/%	90.50/%
SegFormer	83.78/%	92.61/%

Table 3. Ablation results/%

Method	MIoU	MAcc
Baseline	82.74	90.50
Baseline+OREM	84.49	90.86
Baseline+DCAM	84.73	90.94
Baseline+RFPG	84.58	90.73
Baseline+DCAM+RFPG	85.78	91.45
Baseline+OREM+DCAM+RFPG	85.55	91.42

Table 4. Experimental results of different loss functions/%

Loss function	MIoU	MAcc
$Loss_{CE}$	85.55	91.42
$Loss_{CE} + Loss_{Dice}$	85.73	91.49
$Loss_{CE} + Loss_{Tanimoto}$	85.76	91.54
$Loss_{CE} + 0.5Loss_{Tanimoto}$	85.90	91.64
$Loss_{CE} + 0.3Loss_{Tanimoto}$	86.02	91.73
$Loss_{CE} + 0.25Loss_{Tanimoto}$	85.94	91.57

4.4. Ablation Experiment

Firstly, three deep convolutional neural networks including MobileNetv2, ResNet101 and Xception are selected as the backbone network for the experiment, and the results are shown in Table 2. It can be seen that SegFormer has the highest MIoU and MAcc and the best segmentation effect, so SegFormer is chosen as the backbone network of the model.

Then, ablation experiments are conducted on different modules used in the model, and the results are shown in Table 3. It can be seen that when OREM+DCAM+RFPG module is added on the basis of the baseline model, the $MIoU$ and $MAcc$ of the proposed model are the highest, so this method is adopted as the final network structure.

In the above ablation experiments, the loss function used by the model is the cross entropy loss function. Finally, on the basis of the model in this paper, different loss functions are used for training, and the results are shown in Table 4. It can be seen that when the overall loss function $Loss_{total} = Loss_{CE} + 0.3Loss_{Tanimoto}$, the training effect is the best, indicating that it is easier to use this loss function to converge the model parameters to the optimal value.

4.5. Comparison Experiment results on the PASCAL VOC 2012 data set

This paper is compared with the current popular semantic segmentation models on the PASCAL VOC 2012 data set, and the results are shown in Table 5. It can be seen that

Table 5. Comparison with other methods on the PASCAL VOC 2012 data set/%

Method	MIoU	MAcc
SegNet	60.93	73.53
FCN	62.50	75.31
DeepLab	67.14	76.72
U-Net	73.05	80.69
DeconvNet	74.46	84.77
BiSeNet [34]	79.97	87.65
APCNet	80.67	87.22
PSPNet [35]	82.30	88.68
HRNet [36]	83.91	89.23
DMNet [37]	84.66	90.86
Proposed	**86.02**	**91.73**

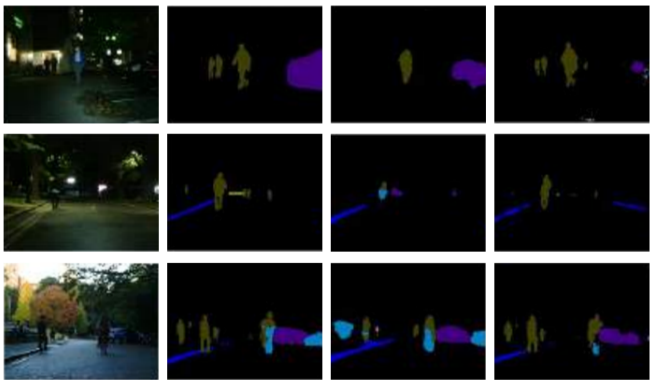


Fig. 6. Comparison of visual results on the PASCAL VOC 2012 data set

the proposed method in this paper achieves the best results on MIoU and MAcc. Compared with DMNet, the MIoU and MAcc of the proposed method increases by 1.32% and 0.87% respectively. Compared with HRNet and PSPNet, MIoU increases by 2.11% and 3.72%, and MAcc increases by 2.50% and 3.05%, respectively. It can be seen that the performance of the proposed method is generally better than that of the current popular semantic segmentation models.

The visualization results of the proposed method and DMNet are shown in Figure 6. The first column is the input image, the second column is the labeled image, and third and fourth column are the segmentation results of DMNet, proposed method, respectively. As can be seen from Figure 6, DMNet makes errors in the segmentation of pedestrians and vehicles, resulting in typical segmentation discontinuities, while none of these errors are found in the proposed method. This is because the proposed method in this paper directly extracts edge features from the original input image and fuses them with the feature image output after sub-sampling by the encoder, so that the feature image has rich shallow spatial information and deep semantic information at the same time. Moreover, the at-

Table 6. Comparison with other methods on the Cityscapes data set/%

Method	MIoU	MAcc
FCN	62.61	70.46
DeepLab	63.18	72.28
U-Net	63.69	71.41
BiSeNet	69.38	76.92
DeepLabv3+	73.87	79.59
PSPNet	74.72	80.77
Proposed	**76.03**	**81.90**

tention mechanism is used to enhance the meaningful information, while other semantic segmentation models do not supplement the spatial details into the feature images after down-sampling. Therefore, the proposed model has significant advantages in object edge segmentation, and the overall segmentation performance is better.

Thereby, by constructing object region enhancement module and dual attention module, the proposed semantic segmentation model in this paper can recover the spatial details of feature images after encoder down-sampling to a certain extent, enhance the accuracy of object edge segmentation, and pay more attention to meaningful information. In addition, by improving the loss function, the parameters of the proposed model in this paper converge more easily to the optimal value, and finally the overall effect of semantic segmentation is improved to some extent. The experimental results show that the proposed model has made remarkable progress in semantic segmentation, especially in object edge segmentation.

4.6. Comparison experiment results on the Cityscapes data set

In order to verify the generalization ability of the model in this paper, Cityscapes data set is used for generalization experiment, and the results are shown in Table 6. The proposed model achieves the best results on MIoU and MAcc, which are 2.16% and 2.31% higher than DeepLabv3+, respectively.

The results of the Proposed visualization with DMNet are shown in Figure 7. It can be seen that the Proposed method can segment the edge part of the object more accurately, the segmentation results are complete and clear, and the overall performance is better.

5. Conclusion

In this paper, a novel semantic segmentation model based on multi-layer information fusion and dual convolutional attention mechanism is proposed to solve the problem of object edge missegmentation and feature discontinuity in scene segmentation. Firstly, the ability to extract multi-dimensional feature information is improved by constructing the object region enhancement module. Combined with the dual convolutional attention mechanism, the high and low level features are fused and the interference of redundant information is suppressed. A pyramid feature grid module combined with residuals is designed and implemented to enhance the semantic expression between features and alleviate the problem of feature discontinuity. The experimental results show that the proposed

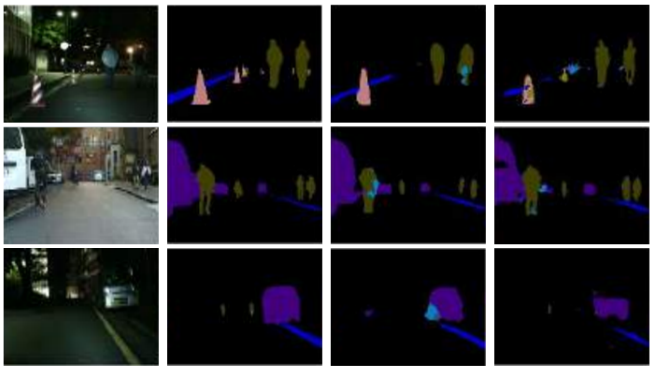


Fig. 7. Comparison of visual results on the Cityscapes data set.

method in this paper can effectively solve the problems of segmentation void and unclear target edge, improve the segmentation accuracy, and the segmentation effect is good. The next step will optimize the lightweight structure and improve the generalization ability of the model in different scenarios.

Acknowledgments. Authors are grateful for the anonymous review by the review experts.

References

1. Luo Z, Yang W, Yuan Y, et al. Semantic segmentation of agricultural images: A survey[J]. Information Processing in Agriculture, 2023.
2. Mo Y, Wu Y, Yang X, et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning[J]. Neurocomputing, 2022, 493: 626-646.
3. Yin S, Wang L, Teng L. Threshold segmentation based on information fusion for object shadow detection in remote sensing images[J]. Computer Science and Information Systems, 2024. doi: 10.2298/CSIS231230023Y.
4. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
5. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
6. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
7. Yuan H, Zhu J, Wang Q, et al. An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots[J]. Frontiers in plant science, 2022, 13: 795410.
8. Lian X, Pang Y, Han J, et al. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation[J]. Pattern Recognition, 2021, 110: 107622.
9. Li X, Li M, Yan P, et al. Deep learning attention mechanism in medical image analysis: Basics and beyonds[J]. International Journal of Network Dynamics and Intelligence, 2023: 93-116.

10. Zhao H, Zhang Y, Liu S, et al. Psanet: Point-wise spatial attention network for scene parsing[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 267-283.
11. Li X, Zhong Z, Wu J, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9167-9176.
12. Asadi A, Safabakhsh R. The encoder-decoder framework and its applications[J]. Deep learning: Concepts and architectures, 2020: 133-167.
13. Qamar S, Ahmad P, Shen L. Dense encoder-decoder Cbased architecture for skin lesion segmentation[J]. Cognitive Computation, 2021, 13(2): 583-594.
14. S. Yin, H. Li, Y. Sun, M. Ibrar, and L. Teng. Data Visualization Analysis Based on Explainable Artificial Intelligence: A Survey[J]. IJLAI Transactionss on Science and Engineering, vol. 2, no. 2, pp. 13-20, 2024.
15. Li X, Chen H, Qi X, et al. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes[J]. IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.
16. Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
17. Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.
18. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
19. Nam H, Ha J W, Kim J. Dual attention networks for multimodal reasoning and matching[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 299-307.
20. Huang Z, Wang X, Huang L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 603-612.
21. Li X, Zhong Z, Wu J, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9167-9176.
22. Katafuchi R, Tokunaga T. LEA-Net: layer-wise external attention network for efficient color anomaly detection[J]. arxiv preprint arxiv:2109.05493, 2021.
23. Agac S, Durmaz Incel O. On the use of a convolutional block attention module in deep learning-based human activity recognition with motion sensors[J]. Diagnostics, 2023, 13(11): 1861.
24. Yang Y, Jiao L, Liu X, et al. Dual wavelet attention networks for image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(4): 1899-1910.
25. Maaz M, Shaker A, Cholakkal H, et al. Edgenext: efficiently amalgamated cnn-transformer architecture for mobile vision applications[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 3-20.
26. Zhao S, Dong Y, Chang E I, et al. Recursive cascaded networks for unsupervised medical image registration[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 10600-10610.
27. Sinha D, El-Sharkawy M. Thin mobilenet: An enhanced mobilenet architecture[C]//2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON). IEEE, 2019: 0280-0285.
28. Qin Z, Zhang Z, Chen X, et al. Fd-mobilenet: Improved mobilenet with a fast downsampling strategy[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1363-1367.
29. Qin Z, Zhang Z, Chen X, et al. Fd-mobilenet: Improved mobilenet with a fast downsampling strategy[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1363-1367.

30. Yin S, Li H, Laghari A A, et al. An anomaly detection model based on deep auto-encoder and capsule graph convolution via sparrow search algorithm in 6G internet-of-everything[J]. IEEE Internet of Things Journal, 2024.
31. Zhang K, Cheng K, Li J, et al. A channel pruning algorithm based on depth-wise separable convolution unit[J]. IEEE Access, 2019, 7: 173294-173309.
32. Dang L, Pang P, Lee J. Depth-wise separable convolution neural network with residual connection for hyperspectral image classification[J]. Remote Sensing, 2020, 12(20): 3408.
33. Li X, Yu L, Chang D, et al. Dual cross-entropy loss for small-sample fine-grained vehicle classification[J]. IEEE Transactions on Vehicular Technology, 2019, 68(5): 4204-4212.
34. Yu C, Wang J, Peng C, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 325-341.
35. Zhou J, Hao M, Zhang D, et al. Fusion PSPnet image segmentation based method for multi-focus image fusion[J]. IEEE Photonics Journal, 2019, 11(6): 1-12.
36. Seong S, Choi J. Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates[J]. Remote Sensing, 2021, 13(16): 3087.
37. Fang K, Li W J. DMNet: difference minimization network for semi-supervised segmentation in medical images[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2020: 532-541.

Lin Teng received the M.A. degree from Shenyang Normal University, Shenyang, China, in 2020. She is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin. Her research interests are image processing and semantic segmentation.

Yulong Qiao was born in 1978. He received the Doctoral degree in engineering from the Harbin Institute of Technology in 2006. In recent years, he has published more than 50 papers in domestic and foreign journals and important conferences, among which more than 20 papers were indexed by SCI, and more than 30 were indexed by EI. His research interests include signal representation and analysis, statistical image processing, image/video processing and applications, and texture analysis and applications. He has won two science and technology awards of Heilongjiang Province. He is a Senior Member of the Chinese Institute of Electronics and IEEE Signal Processing Branch.

Jinfeng Wang received the M.A. degree from Shenyang Normal University, Shenyang, China, in 2016. She currently works at Weifang Vocational College. She has published several academic papers related to her subject. She received several project awards. Her research interests are education management, information processing and big data.

Mirjana Ivanovic (Member, IEEE) has been a Full Professor with the Faculty of Sciences, University of Novi Sad, Serbia, since 2002. She has also been a member of the University Council for informatics for more than 10 years. She has authored or coauthored 13 textbooks, 13 edited proceedings, 3 monographs, and of more than 440 research articles on multi-agent systems, e-learning and web-based learning, applications of intelligent techniques (CBR, data and web mining), software engineering education, and most of which are published in international journals and proceedings of high-quality international conferences. She is/was a member of program committees of more than 200

international conferences and general chair and program committee chair of numerous international conferences. Also, she has been an invited speaker at several international conferences and a visiting lecturer in Australia, Thailand, and China. As a leader and researcher, she has participated in numerous international projects. She is currently an Editor-in-Chief of Computer Science and Information Systems Journal.

Shoulin Yin received the M.A. degree from Shenyang Normal University, Shenyang, China, in 2015. He is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin. His research interests are remote sensing image processing and object detection.

Received: July 13, 2024; Accepted: January 17, 2025.

Computer Science and Information Systems 22(3):927–944 https://doi.org/10.2298/CSIS241201033L

Efficient algorithms for collecting the statistics of large-scale IP address data

Hui Liu[1], Yi Cao[1,⋆], Zehan Cai[1], Hua Mao[2], and Jie Chen[3]

[1] College of Electronic and Information, Foshan Polytechnic,
Foshan 528137, P.R. China
lxyliuhui@163.com
{cycaoyi, lxyliuhui}@hotmail.com
[2] Department of Computer and Information Sciences, Northumbria University,
Newcastle, NE1 8ST, U.K.
hua.mao@northumbria.ac.uk
[3] College of Computer Science, Sichuan University,
Chengdu 610065, P.R. China
chenjie2010@scu.edu.cn

Abstract. Compiling the statistics of large-scale IP address data is an essential task in network traffic measurement. The statistical results are used to evaluate the potential impact of user behaviors on network traffic. This requires algorithms that are capable of storing and retrieving a high volume of IP addresses within time and memory constraints. In this paper, we present two efficient algorithms for collecting the statistics of large-scale IP addresses that balance time efficiency and memory consumption. The proposed solutions take into account the sparse nature of the statistics of IP addresses while maintaining a dynamic balance among layered memory blocks. There are two layers in the first proposed method, each of which contains a limited number of memory blocks. Each memory block contains 256 elements of size 256×8 bytes for a 64-bit system. In contrast to built-in hash mapping functions, the proposed solution completely avoids expensive hash collisions while retaining the linear time complexity of hash-based solutions. Moreover, the mechanism dynamically determines the hash index length according to the range of IP addresses, and can balance the time and memory constraints. In addition, we propose an efficient parallel scheme to speed up the collection of statistics. The experimental results on several synthetic datasets show that the proposed method substantially outperforms the baselines with respect to time and memory space efficiency.

Keywords: large-scale IP addresses, memory blocks, hash table, sorting, network traffic.

1. Introduction

In recent years, the amount of network traffic has increased significantly because of the rapid development of emerging network services, such as video streaming, instant messaging, and online payment services. It is crucial to evaluate the potential impact of the features of user behavior on network traffic management. User behavior features are usually extracted from IP packets, which contain IP addresses. There is a close correspondence between informative user behavior features and the IP addresses that frequently

⋆ Corresponding author

appear in the packets. Hence, how to effectively obtain the statistics of large-scale IP address data in a timely manner, such as every few minutes, is a challenging problem in network traffic measurement. Obtaining the statistics of large-scale IP address data typically consists of two tasks: counting the number of occurrences of each IP address and sorting the results in a specific order.

Each IP address serves as a unique identifier for a device on a network. Analyzing large-scale IP address data can reveal vulnerabilities in the network traffic measurements; this enables administrators to discover and resolve weak spots before they are exploited. As a result, potential cybersecurity risks can be effectively reduced in various network applications. Moreover, analyzing such data helps in understanding traffic patterns and behaviors in network traffic measurement process. The collection and analysis of large-scale IP address data provide valuable insights that can guide more reliable and efficient network systems; this helps administrators balance loads and improve the overall performance of the network. Therefore, collecting statistics from large-scale IP address data is an essential task for the efficient, secure, and scalable management of networks.

A number of statistical algorithms for solving this problem have been studied in the last few decades [11], [15], [16], [8], [2]. A classic divide-and-conquer strategy has been proposed in which IP addresses are first divided into multiple subsets. Then, each subset is individually computed using a statistics collection method, such as a top k trie algorithm [12]. Finally, to merge and sort the results of the multiple subsets, sorting algorithms (e.g., bubble sort, insertion sort, merge sort, selection sort, or quick sort) are used [9], [6]. However, the reduction of computational cost is an intractable problem in the sorting procedure [21], [18], [1]. For example, the average complexity and worst-case complexity of the bubble, insertion, and selection sort algorithms are $O\left(n^2\right)$ [9], [24], where n represents the number of unsorted records. This indicates that the merging and sorting step of the multiple subsets occupies a large amount of memory and is extremely computationally costly to perform when collecting the statistics of millions or tens of millions of IP addresses. Therefore, statistics collection algorithms using the divide-and-conquer strategy still face several challenges caused by the rapid increase of the large-scale records, such as bounded memory and computational cost restrictions.

A hash table is an effective method for collecting the statistics of IP addresses [19]. It uses a hash function to compute a hash codes for an array of buckets with the statistical results. The hash function assigns each key to a unique bucket for each IP address. Unfortunately, the hash function can generate the same hash code for more than one IP address. With the increase in the generation of big data, millions or tens of millions of records have become ubiquitous in network traffic. Therefore, this approach could cause several hash collisions, especially for a large number of IP addresses. Although many strategies can be employed to avoid collisions, such as linear probing, quadratic probing, and double hashing, they require extra storage space and computation.

The statistics collection algorithm should be stable, effective, and efficient for large-scale records. Recent advancements in sorting techniques have concentrated on improving the efficiency and scalability of algorithms for processing large-scale data.To overcome the disadvantages of general statistics collection methods, a number of parallel techniques have been developed for large-scale records by optimizing the efficiency and complexity. For example, these algorithms have been extended to the corresponding parallel structures

for parallel hardware architectures, such as many-core and multi-core platforms [3], [27], [23],[26].

An IP address consists of a series of numbers separated by periods. However, the sorting techniques ignore the importance of the original characteristics of the unsorted IP records. Additionally, parallel statistics collection algorithms are simply parallel computing implementations of the original algorithms. The uniqueness of the IP addresses is vital for distinguishing between the different devices. An IPv4 address is typically represented in four parts. Consequently, collecting statistics from large-scale IP address data still face two significant challenges. First, the unique structure of unsorted IP records is ignored when large-scale IP address data are collected. Second, the extension of the collection task to the corresponding parallel computation paradigm deserves a further investigation into parallel hardware architectures.

In this paper, we present two efficient algorithms for collecting the statistics of large-scale IP address data. We can obtain the frequently occurring IP addresses from the statistics, which can be regarded as a pre-processing step of user behavior analysis in network traffic management. Because of the increasing volume and speed of network traffic, it has become expensive and impractical to handle all IP addresses contained the IP packets. By taking full advantage of the successive characteristics of memory addresses and the fixed range of each individual part of an IP address, we design two relationship mapping mechanisms between memory blocks and IP addresses for a four-dimensional sparse matrix. The sparse matrix stores the number of occurrences of the individual IP addresses, in which the positions of the rows and columns are employed to represent the mapping relationship between the memory blocks and IP addresses. Specifically, we construct a two-layer memory block (TLMB) to implement the first mapping mechanism for the IP addresses. In addition, we employ a single shared memory block (SSMB) for all IP addresses to implement the other mapping mechanism of the IP addresses. The mechanisms of the mapping relationship effectively remove the information about trivial user behaviors that are irrelevant for statistical analysis. The proposed methods can be extended to the corresponding parallel versions for specific hardware architectures. Extensive experiments on several synthetic datasets show the effectiveness of the proposed method.

Our contributions are summarized as follows.

1. We present two efficient methods for collecting the statistics of large-scale IP address data that use two different relationship mapping mechanisms, TLMB and SSMB, between memory blocks and IP addresses for a four-dimensional sparse matrix.
2. The computational cost of TLMB is linearly proportional to the number of IP addresses with a limited memory resource, and the memory use of SSMB remains almost unchanged with a reasonable computational cost, regardless of the number of IP addresses.
3. A parallel computation optimization scheme for multiple computers is proposed to effectively improve the computational efficiency and dramatically reduce memory use.
4. Our extensive experimental results using synthetic datasets demonstrate that our proposed method shows clear superior performance in comparison with the baselines, striking a balance between computational cost and memory use.

The remainder of this paper is organized as follows. We briefly describe some of the original sorting techniques as well as various parallel sorting algorithms in Section 2. In

Table 1. Time complexity of various sorting algorithms [6], [17], [7].

Algorithm	$Best\,(n)$	$Average\,(n)$	$Worst\,(n)$
Bubble sort	$O\,(n)$	$O\left(n^2\right)$	$O\left(n^2\right)$
Insertion sort	$O\,(n)$	$O\left(n^2\right)$	$O\left(n^2\right)$
Selection sort	$O\left(n^2\right)$	$O\left(n^2\right)$	$O\left(n^2\right)$
Merge Sort	$O\,(n \log n)$	$O\,(n \log n)$	$O\,(n \log n)$
Quick Sort	$O\,(n \log n)$	$O\,(n \log n)$	$O\left(n^2\right)$
Heap Sort	$O\,(n \log n)$	$O\,(n \log n)$	$O\,(n \log n)$

Section 3, we introduce the proposed method. The experimental results are presented in Section 4. Finally, we draw the conclusions of the study in Section 5.

2. Related Work

2.1. Classical Sort Techniques

Bubble sort is a classical sorting algorithm in which each element in a list is compared with its neighboring elements and swapped until they are in the desired order [25]. Bubble sort leads to $(n-1)$ number of passes and $\frac{n(n-1)}{2}$ number of iterations if n elements are given. Insertion sort is a simple and efficient sorting algorithm that iteratively takes one element and finds its appropriate position in the sorted list by comparing it with neighboring elements. It becomes less efficient as the number of records increases. Selection sort determined the smallest number in an unsorted list and swaps it with the first number in the sorted list. Then, it finds the next smallest element from the remaining list and swaps with the second element in the sorted list. Consequently, the number of sorted elements at the top of the list increases while the rest remain unsorted. Merge sort, which is based on the divide-and-conquer principle, repeatedly divides the array into two halves and then combines them in a sorted manner For more details of classical sort algorithms, such as quick sort and heap sort, we refer the reader to the comments in [10], [6]. It is well known that computational cost and required memory are the primary concerns in sorting algorithms [22], [14]. Table 1 shows the time complexity of the best case, average case, and worst case of several classical sorting algorithms [6], [17], [7]. These sorting algorithms can be used to find the first k IP addresses of the most frequent occurrences from the statistical results obtained for IP address data.

2.2. Accelerating Large-scale Sorting Techniques

Sorting is an essential part of modern computing. Significant efforts have recently been dedicated to accelerating large-scale sorting techniques [3], [13], [27]. For example, Alhabboub *et al.* improve the computation efficiency of the classical QuickSort algorithm by combining with parallel implementations [3]. The improved QuickSort algorithm can be applied on the sorting large-scale data, and exhibits slightly superior computational efficiency compared to classical sequential QuickSort. Jugé *et al.* proposed an adaptive ShiversSort algorithm for efficiently sorting partially sorted data, which is considered as a variant of the well-known algorithm TimSort [13]. Yang *et al.* proposed a high-performance

parallel sorting algorithm on a CPU-DSP heterogeneous processor [27]. These methods typically take into account different sorting settings, such as parallel sorting environments, data criterions, or specific CPU architectures. These large-scale sorting algorithms provide an efficient post-processing step for collecting the statistics of large-scale IP address data.

2.3. Parallel Hardware Architectures

The hardware architecture of modern processors usually consists of more than two independent central processing units (CPUs) or graphics processing units (GPUs). Parallel software platforms can be implemented using high-level programming frameworks for specific hardware architectures [5]. The Compute Unified Device Architecture (CUDA) is a parallel computing platform for general computing on GPUs. Most parallel sorting algorithms are variants of standard, well-known sorting algorithms adapted to GPU hardware architecture. For example, Cederman designed a quick sort for the GPU platform [4], and Peters proposed an adaptive bitonic sorting algorithm with a bitonic tree for GPUs [20]. The parallel computation of sorting algorithms is considered to be the most efficient way of sorting elements on parallel hardware architectures [26].

Fig. 1. Example of a memory block of size 256×8 bytes containing 256 elements

3. Proposed Method

3.1. Problem Formulation

IP flow data FD is a sequence of IP records, that is, $FD = \{(x_1, p_1), ..., (x_n, p_n)\}$ and $n \geq 1e^6$, where each pair of elements (x_i, p_i) $(i \in [1, n])$ consists of an IP address x_i and a set of corresponding user behavior attributes p_i. Given a finite set of IP addresses $X = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^{m \times n}$, the purpose of the IP address statistics task is to efficiently determine the first k IP addresses of the most frequent occurrences in X, where m the dimensionality of an individual IP address and $k \ll n$.

3.2. IP Address Statistics Task

A standard IP address is composed of four decimal numbers ranging from 0 to 255 which are separated by dot symbols. An individual IP address is logically divided into four parts by splitting it with respect to each dot symbol, and each part of an IP address has an integer value. We create a four-dimensional array for the statistics of IP addresses, where the length of each dimension in the array is 256. Each element of the array can be employed

to store the number of occurrences of the IP address according to the relationship mapping between the index of each dimension of the array and the integer value of the corresponding part of the IP address. For example, consider the individual IP address 1.2.3.4 and the four-dimensional array fd_array. The number of occurrences of this IP address is stored in $fd\_array[1][2][3][4]$. However, individual IP addresses in the host logs often make up a small proportion of all IP addresses. The array can be considered to be sparse because most of its elements are zeros. Consequently, we can carefully design a four-dimensional sparse matrix to store the number of an individual IP address by taking full advantage of the successive characteristics of array addresses and the fixed range of an individual part of an IP address.

Algorithm 1 TLMB

Input:

A finite set of IP addresses $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{m \times n}$, number $k > 1$.

Construct 256×256 memory blocks of size 128 MB for the first layer;
for $i = 1 : n$ **do**
  Assume that a, b, c and d each represent one of the integer values of the four parts of IP address x_i.
  Calculate the index of the memory block in the first layer: $p = a \times 255 \times 255 + b \times 255$.
  if the value of the j-th element is null in the p-th memory block **then**
    Create a memory block in the second layer, set all elements of the memory block to zero, and store the starting address of the memory block in the j-th element.
  else
    Obtain the starting address of the memory block m in the second layer by finding the j-th element of the p-th memory block.
  end if
  Add 1 to the d-th element of memory block m in the second layer.
end for
Construct a minimum heap of size k using each non-zero element of the memory blocks in the second layer.

Output:

Traverse the nodes of the heap to obtain the k IP addresses and their number of occurrences.

3.3. IP Usage Storage and Retrieval Strategies for the Four-dimensional Sparse Matrix

Assume that each memory address of a 64-bit system can be stored in an element 8 bytes in size, and the number of occurrences of an individual IP address is no more than 2^{64}. There is an array of size 256 elements that consists of 256×8 bytes of memory. The array is regarded as a memory block that contains a contiguous address space, as shown in Fig. 1. In other words, the addresses of all bytes of the array are sequential in the memory block. Therefore, the position of the array can be indexed by the integer value of a particular part of the IP address. We present two efficient methods to collect the statistics of large-scale IP address data, each of which contains a relationship mapping mechanism between memory blocks and IP addresses for the four-dimensional sparse matrix.

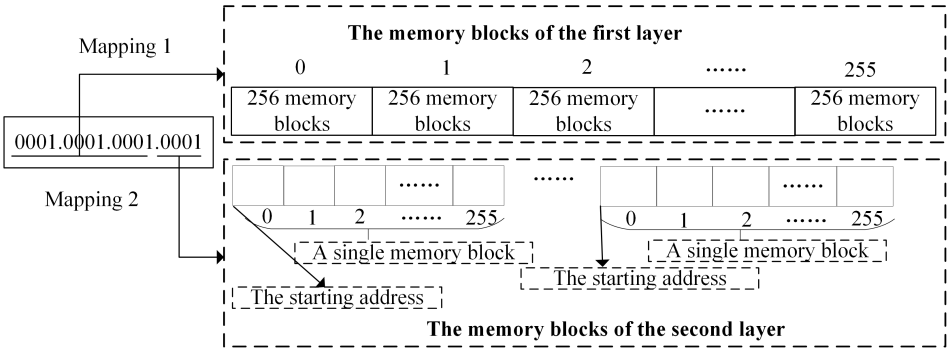


Fig. 2. An example of the mapping relationships between memory blocks and an IP address

First method: TLMB The first proposed mapping mechanism of IP addresses is TLMB. The four parts of the IP address are represented in four layers, where each layer is made up of one or more memory blocks. The first layer only contains one memory block, whereas the second layer contains 256 memory blocks. Each memory block contains 256 elements. Each element of the memory block in the first layer is employed to store the starting addresses of the corresponding 256 memory blocks in the second layer. Similarly, the third layer contains 256×256 memory blocks, the size of which is 128 MB in memory. Then, the element of each memory block in the third layer stores the starting address of the corresponding memory block in the fourth layer. This would be 32 GB in size if we adopted a pre-allocation strategy for all memory blocks in the four layers. Hence, we present an alternative pre-allocation strategy for the memory blocks. A memory block will be allocated only when the first three parts of an initial IP address have been given. In particular, pre-allocating a big memory block of size 128 MB containing 256×256 contiguous memory blocks is feasible in a modern computer. Consequently, the first two layers can be removed from this architecture if the third layer has contiguous memory blocks of 128 MB.

We formally present a storage strategy for IP addresses that consists of two layers that consist of a limited number of memory blocks. The first layer contains 256×256 memory blocks. The first three parts of the IP address can be mapped into the corresponding position of the element in a particular memory block of the first layer according to the individual values of the three parts. We allocate a memory block in the other layer for the IP address when its first three parts are initially given. Each element of a memory block in this layer stores the number of occurrences of the corresponding IP address. Figure 2 shows an example of the relationship mapping between the memory blocks of two layers and an IP addresses.

Consider the individual IP address 1.2.3.4, we have $1 \times 255 \times 255 + 2 \times 255 = 65535$, which represents the index of the memory block in the first layer. The positions of the first two dimensions of the sparse matrix can be mapped to the elements of the memory blocks included in the first layer. The third part of the IP address denotes the index of the memory block in the second layer. The positions of the final dimensionality of the sparse matrix

can be indexed by combining the starting address of the memory block in the second layer with the integer value of the four parts of the IP address.

Algorithm 2 SSMB

Input:

A finite set of IP addresses $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{m \times n}$, number $k > 1$.

Construct a memory block of size 128 MB saved by all IP addresses;
All IP addresses are logically partitioned into q subsets according to the first part of each IP address.
for $i = 1 : n$ **do**
    Assume that a, b, c and d each represent one of the integer values of the four parts of IP address x_i.
    for $j = 1 : q$ **do**
        if $j == q$ **then**
            Calculate the position of the memory block in the first layer: $p = b \times 255 \times 255 + c \times 255 + d$.
            Add 1 to the p-th element of the memory block.
        end if
    end for
    Construct a minimum heap of size k using the each non-zero elements of the memory block.
end for

Output:

Traverse of the nodes of the heap to obtain the k IP addresses and their number of occurrences.

We traverse all elements of the memory blocks of the second layer to obtain the maximum number of occurrences of elements if $k = 1$. Otherwise, we construct a minimum heap of size k. The statistical results of the first k IP addresses are saved in the heap, which is a special binary tree and implemented by an array of size k. The construction of the heap is completed by traversing all elements of the memory blocks of the second layer. The tree node in the leap contains two important attributes: an IP address and its number of occurrences. The final IP addresses and numbers of occurrences can be obtained by a traversal of the nodes of the heap. The complete procedure for determining the first k of the most frequent IP addresses from host logs is outlined in Algorithm 1.

Second method: SSMB We also designed an SSMB that stores all IP addresses and their statistics. IP addresses are logically divided into at most 256 subsets according to the value of the first part of each individual IP address. We construct a single memory block of $256 \times 256 \times 256$ elements that is $256 \times 256 \times 256 \times 8$ bytes in size, that is, 128 MB. The memory block is shared by all subsets. For each subset, the last three parts of the IP address can be mapped into the corresponding position of the element in the shared single memory block, which is always initialized at the beginning of the relationship mapping. Then, the element of this memory block stores the statistics of the IP addresses in this subset.

We further perform a round traversal of the memory block to initialize a minimum heap of size k after the relationship mapping has been completed in the first subset. Then, we continue to perform a round traversal of the memory block to adjust the heap after the

relationship mapping has been completed for the subsequent subsets. Finally, we obtain the first k most frequently occurring IP addresses in the heap. The complete procedure for finding the first k of the most frequency IP addresses from host logs is outlined in Algorithm 2.

3.4. Memory Use and Complexity Analysis

We first evaluate the memory use and computational complexity of the first proposed method, TLMB. The size of each memory block is 256×8 bytes, and there are 256×256 memory blocks in the first layer. Hence, the size of the memory blocks in the first layer is 128 MB. Assume that the number of the distinct first three parts of the IP addresses is s. The number of memory blocks in the second layer is linearly proportional to s. Moreover, the memory size of the minimum heap is $(k + 8)$ bytes, where each tree node contains two attributes, that is, an IP address and the number of occurrences. The total size of the memory of the proposed method is approximately the sum of the three parts, that is, 128 MB, s KB, and $(k + 8)$ bytes. The computational complexity of the two layers for calculating the IP address statistics is $O(n)$ in Algorithm 1, where n is the number of IP addresses. In addition, the computational complexity of constructing a minimum heap of size k is $O(k \log k)$, where k is the number of tree nodes in the heap. Consequently, the overall computational complexity of the proposed algorithm is $O(k \log k + n)$.

We next evaluate the memory use and computational complexity of the second proposed method, SSMB. The memory size of SSMB is 128 MB. Assume that the number of the distinct first parts of the IP addresses is q. The computational complexity of the mapping mechanism of the IP addresses in Algorithm 1 is $O(qn)$, where n is the number of IP addresses. Similarly, the computational complexity of constructing a minimum heap of size k is $O(k \log k)$ for each subset. Consequently, the overall computational complexity of the proposed algorithm is $O(q(k \log k + n))$.

3.5. Parallel Computation Optimization Techniques

Parallel computation mechanism of TLMB In the worst case, the distinct first three parts of the IP addresses cover all the binary combinations. Hence, the size of the memory blocks in the second layer is 32 GB. We present a parallel computation scheme on multiple computers for improving the computational efficiency and reducing memory use. Assume that there are 2^r computers available for parallel computation, where r represents a positive integer. The task of collecting IP address statistics is then divided into multiple subtasks, which are performed by 2^r computers, respectively, according to the first r bits of the first part of the IP addresses. The number of memory blocks reduces to $(256 \times 256)/2^r$ for 2^r computers. Simultaneously, the number of memory blocks will decrease to $32/2^r$ GB in the second layer. For example, the number of memory blocks in the first layer is 256×64 for each computer when $r = 2$, and the size of memory blocks in the second layer is 8 GB in the worst case. In addition, if four computers perform the task of computing IP address statistics by partitioning the first two bits of the first part of the IP addresses, then the second layers of multiple computers are merged into a complete second layer, where is employed to construct a minimum heap of size k. Finally, the parallel computation results can be obtained in a manner similar to the last step of Algorithm 1.

Table 2. Statistics of the datasets

Data sets	IP Records	Individual IP Addresses	Size	Type
1	5, 000, 000	50, 000	77.5 MB	Synthetic
2	10, 000, 000	100, 000	155 MB	Synthetic
3	50, 000, 000	500, 000	775 MB	Synthetic
4	1, 114, 633	107, 988	14.4 MB	Real
5	1, 430, 258	133, 116	18.5 MB	Real

Parallel Computation Mechanism of SSMB Assume that all IP addresses are logically divided into q subsets according to the value of the first part of an individual IP address. Further assume are q computers for parallel computation, where the statistics collection task of each subset can be performed by an individual computer. A minimum heap of size k is shared among these computers. Hence, this greatly increases the computational efficiency of the task by q times.

4. Experiments

4.1. Experimental Settings

In this section, we evaluate the performance of the proposed methods [4] on two different types of datasets, i.e., three synthetic datasets and two real-world datasets. The three synthetic datasets contain 5 million, 10 million, and 50 million randomly generated IP records. Each individual IP address contains one or more of IP records. The average number of IP records is 100 for each individual IP address. In particular, the IPv4 addresses were specifically divided into four segments in the experiments. These experimental settings ensure a comprehensive evaluation of the capacity of TLMB and SSMB to efficiently collect statistics for large-scale IP address data. Parameter k represents the number of frequently occurring IP addresses. Additionally, two real-world network traffic datasets, provided by the Center for Applied Internet Data Analysis (CAIDA) [5], are collected from various parts of the internet. These two datasets are widely used in networking and traffic analysis research. The statistics of these datasets are summarized in Table 2.

We compared the proposed method with the following baselines:

- **Hash Mapping**. Each IP record is mapped into an entry with a statistical result using a hash table [6]. Next, the statistical results are used to construct a minimum heap of size k.
- **IP Mapping**. All IP records are partitioned into q subsets according to the first part of each IP address. The statistics of the IP records in each subset are mapped into an array, whose memory is pre-allocated on a computer according to the last three parts of each IP address. The first k most frequent IP addresses are chosen from each subset. Next, a minimum heap of size k is constructed using the $q \times k$ IP addresses.

[4] https://github.com/chenjie20/IPStatistics
[5] https://www.caida.org/catalog/datasets/ipv4_prefix_probing_dataset
[6] https://github.com/activesys/libcstl

Table 3. Computational cost (s) of different methods on the three synthetic datasets

Data	k	Hash Mapping	IP Mapping	Ours (TLMB)	Ours (SSMB)
1	10	1,458.14 (2.32)	15.18 (0.07)	**2.51** (0.01)	18.43 (0.05)
	100	1,458.56 (2.76)	15.21 (0.04)	**2.52** (0.01)	18.43 (0.06)
2	10	2,927.23 (11.64)	17.41 (0.11)	**4.92** (0.01)	27.59 (0.05)
	100	2,934.65 (28.93)	17.45 (0.05)	**4.95** (0.01)	27.65 (0.06)
3	10	14,547.09 (13.95)	35.33 (0.09)	**24.22** (0.07)	101.01 (0.17)
	100	14,566.36 (26.78)	35.39 (0.05)	**24.24** (0.04)	101.20 (0.2)

Table 4. Memory use (MB) of different methods on the three synthetic datasets

Data	k	Hash Mapping	IP Mapping	Ours (TLMB)	Ours (SSMB)
1	10	**43.63** (0.12)	16,332.81 (0.07)	189.61 (0.16)	139.77 (0.13)
	100	**43.73** (0.06)	16,332.77 (0.05)	189.64 (0.07)	139.79 (0.07)
2	10	**74.8** (0.12)	16,332.75 (0.12)	239.11 (0.1)	139.74 (0.13)
	100	**74.83** (0.05)	16,332.77 (0.07)	239.06 (0.15)	139.7 (0.12)
3	10	**324.45** (0.99)	16,332.59 (0.06)	628.33 (0.07)	139.6 (0.15)
	100	**324.4** (1.02)	16,332.6 (0.16)	628.27 (0.01)	139.71 (0.16)

Memory blocks were pre-allocated for TLMB and SSMB, with sizes ranging from small-scale (e.g., 5 million) to large-scale (e.g., 50 million) to test scalability. Two metrics were employed to evaluate the sorting performance, that is, computational cost and memory use. All experiments were implemented using the C language on a Windows platform with an Intel i7-9700k CPU and 32 GB RAM.

4.2. Experiment Results

Experimental Evaluation on Synthesized Data Parameter k was set to 10 or 100, and we repeated each experiment 10 times. The average computational costs and standard deviations are reported in Table 3, and the mean memory use and standard deviations are given in Table 4. The results show that TLMB consistently outperformed all the other methods in terms of computational cost. For example, TLMB achieves computational costs of 2.51 s and 2.52 s when $k = 10$ and $k = 100$, respectively. When the number of IP addresses increases from 5 million to 50 million with $k = 10$, the gap in computational costs of TLMB and IP Mapping are 12.6 s and 11.11 s, respectively. We also observed the same advantages when $k = 100$. In addition, SSMB shows competitive results when compared with the comparison methods in terms of memory use. The memory use of SSMB is always approximately 139 MB, even when the number of IP addresses changes. In contrast, the computational cost of SSMB substantially outperforms that of Hash Mapping under different numbers of IP addresses. IP Mapping obtained the lowest computation cost for all numbers of IP addresses. However, the highest memory use results of IP Mapping are consistent with expectations.

Experimental Evaluation on Real-World Data We evaluate the proposed and competing methods on two real-world datasets. Tables 5 and 6 show the computational costs and memory usages levels of different methods, respectively. TLMB consistently incurs lower

Table 5. Computational cost (s) of different methods on the two real-world datasets

Data	k	Hash Mapping	IP Mapping	Ours (TLMB)	Ours (SSMB)
4	10	1675.3	6.73	**0.61**	7.00
	100	1673.9	6.90	**0.62**	7.09
5	10	1996.8	7.33	**0.78**	6.88
	100	1989.9	7.47	**0.80**	7.02

Table 6. Memory use (MB) of different methods on the two real-world datasets

Data	k	Hash Mapping	IP Mapping	Ours (TLMB)	Ours (SSMB)
4	10	238.9	16,392.1	249.8	**183.3**
	100	238.8	16,392.1	249.8	**183.4**
5	10	266.7	16,391.9	281.4	**183.4**
	100	266.7	16,392.3	281.4	**183.4**

computational costs than do the other methods. SSMB and IP mapping demonstrate comparable computational costs. However, SSMB significantly reduces the memory requirements relative to IP mapping. Furthermore, Hash mapping incurs a higher computational cost than the competing methods do because of its use of hash computations, making it prohibitively time-consuming in practice. As the number of IP addresses increases across the two real-world datasets, SSMB maintains relatively stable memory usage. These findings highlight the effectiveness of both TLMB and SSMB.

4.3. Ablation study

To investigate the impact of the memory blocks in the proposed TLMB and SSMB methods, we performed ablation studies on the three synthetic datasets. Specially, we examined two particular cases in the experiments. The hash table was employed to replace the second part of TLMB and SSMB, respectively. The primary goal of the ablation study is to demonstrate the importance of the memory blocks on collecting the statistics of large-scale IP address data. The invariants of TLMB and SSMB corresponding to these two cases are referred to as TLMB$_{hash}$ and SSMB$_{hash}$, respectively.

Tables 7 and 8 show the results of the ablation study regarding the computational cost and memory use. TLMB$_{hash}$ exhibits similar computational cost and memory use compared to SSMB$_{hash}$ on the first two synthetic datasets. Additionally, the computational cost of TLMB$_{hash}$ is slightly higher than that of SSMB$_{hash}$, while its memory usage is marginally lower. TLMB$_{hash}$ and SSMB$_{hash}$ achieve an acceptable computational cost. However, TLMB and SSMB achieves superior performance in computational cost and memory use compared with those of TLMB$_{hash}$ and SSMB$_{hash}$. These results further emphasize that integrating a hash table scheme into TLMB and SSMB is both time-consuming and memory-intensive. Therefore, the results of the ablation study demonstrate the effectiveness of the memory blocks in the proposed TLMB and SSMB methods.

Table 7. Ablation study concerning the computational costs (s) incurred on the three synthetic datasets

Data	k	TLMB$_{hash}$	SSMB$_{hash}$	Ours (TLMB)	Ours (SSMB)
1	10	37.95	37.16	**2.51**	18.43
	100	37.9	37.34	**2.52**	18.43
2	10	75.68	74.51	**4.92**	27.59
	100	75.19	74.05	**4.95**	27.65
3	10	384.72	374.49	**24.22**	101.01
	100	384.23	375.9	**24.24**	101.20

Table 8. Ablation study concerning the memory use (MB) required for the three synthetic datasets

Data	k	TLMB$_{hash}$	SSMB$_{hash}$	Ours (TLMB)	Ours (SSMB)
1	10	3,397.3	3397.2	189.61	**139.77**
	100	3,397.5	3,397.2	189.64	**139.79**
2	10	6,654.4	6,654.3	239.11	**139.74**
	100	6,654.4	6,654.4	239.06	**139.7**
3	10	25,665.6	27,367.6	628.33	**139.6**
	100	26,008.6	26,517.5	628.27	**139.71**

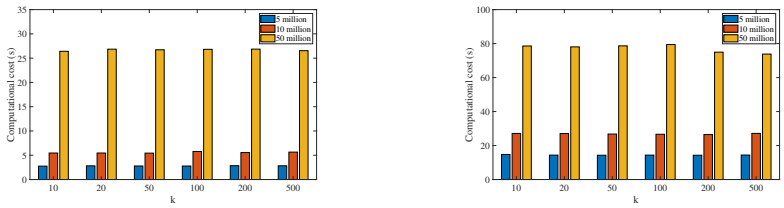
4.4. Empirical Investigation

We empirically examined the effect induced by varying the k considered in the proposed TLMB and SSMB methods. Here k was selected from the set $\{10, 20, 50, 100, 200, 500\}$. The computational cost and memory use were employed to evaluate TLMB and SSMB with different k values.

Fig. 3 shows the computational costs of TLMB and SSMB with different k values. As expected, the computational cost gradually increases as the number of IP records increases from 5 million to 50 million. Moreover, the computational costs of TLMB and SSMB remain relatively stable when k varies from 10 to 500 on each synthetic dataset. This finding demonstrates the stability of TLMB and SSMB for computational efficiency when collecting the statistics of large-scale IP address data. Fig. 4 shows the memory uses of TLMB and SSMB with different k values. We observe that TLMB requires more memory use as the number of IP records grows. In contrast, SSMB maintains relatively stable memory across varying numbers of IP records. This finding indicates that SSMB can satisfy certain memory requirements when handling varying numbers of IP records.

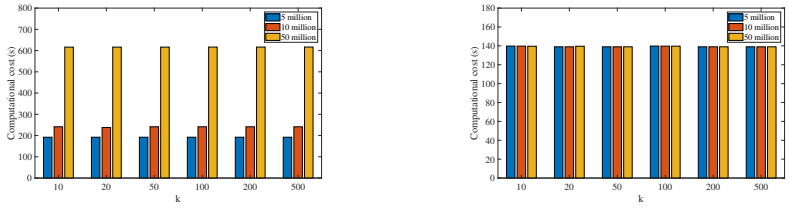
4.5. Discussion

The gap in computational cost between TLMB and Hash Mapping dramatically increases when the number of IP records increases from 5 million to 50 million. This is because TLMB avoids hash collisions when an IP address is mapped to the corresponding memory block. The computational cost of TLMB is linearly proportional to the number of IP addresses. Moreover, the memory use of SSMB remains almost unchanged regardless of the number of IP addresses. This is consistent with the theory underlying the second proposed mapping mechanism. There is a negligible effect on the computational costs of



(a) (b)

Fig. 3. The computational costs of the proposed TLMB and SSMB methods with different k values. (a) TLMB and (b) SSMB



(a) (b)

Fig. 4. The memory uses of the proposed TLMB and SSMB methods with different k values. (a) TLMB and (b) SSMB

TLMB and SSMB when k increases from 10 to 100. Moreover, the changes in the memory use of TLMB and computational cost of SSMB are tolerable in practical applications as the number of IP addresses increases. Consequently, TLMB and SSMB reach a reasonable balance between computational cost and memory use when compared with Hash Mapping and IP Mapping. This reveals that the two relationship mapping mechanisms for memory blocks and IP addresses are effective approaches for the design of the four-dimensional sparse matrix.

The memory blocks designed in TLMB and SSMB exhibit superior relationship mapping capabilities compared to those of hash mapping. The memory block takes fully advantages of the inherent property of the memory address, which is employed to corresponding to each part of an IP address. This indicates that integrating the memory block into TLMB and SSMB is both time-stable and memory-stable. In contrast, hash mapping uses a pair of key and value to store the statistics of IP address data. Unfortunately, IP addresses are often sparse in practical applications. Hash mapping requires additional memory to store the remaining two or three parts of the IP address as keys corresponding to TLMB and SSMB, respectively. This has a significant negative impact on the memory use of hash mapping. Therefore, the proposed memory block significantly enhances the capacity of TLMB and SSMB in collecting the statistics of large-scale IP address data.

5. Conclusion

The collection of the statistics of large-scale IP address data is one of the most fundamental problems in network traffic measurement. In this paper, we addressed this problem. Specifically, the two proposed methods present two different relationship mapping mechanisms between memory blocks and IP addresses to strike a balance between computational cost and memory use. They can be employed to search for frequently occurring IP addresses in practical applications. The extensive experimental results demonstrate the effectiveness of the proposed methods.

Acknowledgments. This work was supported in part by the Guangdong Province Science and Technology Innovation Strategy Special Fund under Grant PDJH2024B648, in part by the Guangdong Province Characteristic Innovation Project for Normal Universities under Grant 2023KTSCX338, and in part by the Province Ordinary Higher Education Engineering Technology Research (Development) Center under Grant 2024GCZX028.

References

1. Abdel-Hafeez, S., Gordon-Ross, A.: An efficient o(n) comparison-free sorting algorithm. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 25(6), 1930–1942 (Jun 2017)
2. Agapitos, A., Lucas., S.M.: Evolving efficient recursive sorting algorithms. In: 2006 IEEE International Conference on Evolutionary Computation. pp. 2677–2684. Vancouver, BC, Canada (Jul 2016)
3. Alhabboub, Y., Almutairi, F., Safhi, M., Alqahtani, Y., Almeedani, A., Alguwaifli, Y.: Accelerating large-scale sorting through parallel algorithms. Journal of Computer and Communications 12(1), 131–138 (Jan 2024)
4. Cederman, D., Tsigas, P.: A practical quicksort algorithm for graphics processors. In: European Symposium on Algorithms. pp. 246–258 (2008)
5. Chen, S., Qin, J., Xie, Y., Zhao, J., Heng, P.A.: A fast and flexible sorting algorithm with cuda. In: International Conference on Algorithms and Architectures for Parallel Processing. pp. 281–290. Taipei, Taiwan, China (Jun 2009)
6. Cormen, T., C.Leiserson, Rivest, R., C.Stein: Introduction to Algorithms. MIT press (2009)
7. Faujdar, N., Ghrera, S.P.: Analysis and testing of sorting algorithms on a standard dataset. In: 2015 Fifth International Conference on Communication Systems and Network Technologies. pp. 1–10. Gwalior, India (Apr 2015)
8. Fredman, M.L.: An intuitive and simple bounding argument for quicksort. Information Processing Letters 3(114), 137–139 (Mar 2014)
9. Hammad, J.: A comparative study between various sorting algorithms. International Journal of Computer Science and Network Security 15(3), 358–367 (Mar 2015)
10. Idrizi, F., Rustemi, A., Dalipi, F.: A new modified sorting algorithm: A comparison with state of the art. In: 2017 6th Mediterranean Conference on Embedded Computing. pp. 1–6. Bar, Montenegro (Jul 2017)
11. Jing, Y.N., Tu, P., Wang, X.P., Zhang, G.D.: Distributed-log-based scheme for ip traceback. In: The Fifth International Conference on Computer and Information Technology. pp. 711–715. Shanghai, China (Dec 2005)
12. Jing, Y.N., Tu, P., Wang, X.P., Zhang, G.D.: Space-efficient data structures for top-k completion. In: Proceedings of the 22nd international conference on World Wide Web. pp. 583–594. Rio de Janeiro, Brazil (May 2013)

13. Jugé, V.: Adaptive shivers sort: an alternative sorting algorithm. ACM Transactions on Algorithms 20(4), 1–55 (Aug 2024)
14. Jukna, S., Seiwert, H.: Sorting can exponentially speed up pure dynamic programming. Information Processing Letters 159, 451–469 (Apr 2020)
15. Kapur, E., Kumar, P., Gupta, S.: Proposal of a two way sorting algorithm and performance comparison with existing algorithms. International Journal of Computer Science, Engineering and Applications 2(3), 61–78 (Jun 2012)
16. Klein, S.T.: On the connection between hamming codes, heapsort and other methods. Information Processing Letters 113(17), 617–620 (May 2017)
17. Kocher, G., Agrawal, N.: Analysis and review of sorting algorithms. International Journal of Scientific Engineering and Research 2(3), 81–84 (Mar 2014)
18. Louza, F.A., Gog, S., Telles., G.P.: Optimal suffix sorting and lcp array construction for constant alphabets. Information Processing Letters 118, 30–34 (Sep 2017)
19. Múller, I., Sanders, P., Lacurie, A., Lehner, W., Fárber, F.: Cache-efficient aggregation: Hashing is sorting. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. p. 1123–1136. Melbourne, Bictoria, Australia (May 2015)
20. Peters, H., Schulz-Hildebrandt, O., Luttenberger, N.: Fast in-place, comparison-based sorting with cuda: a study with bitonic sort. Concurrency and Computation: Practice and Experience 23(7), 681–693 (Jan 2011)
21. Puglisi, S.J., Smyth, W.F., Turpin, A.: The performance of linear time suffix sorting algorithms. In: Data Compression Conference. pp. 358–367. Snowbird, UT, USA, USA (Mar 2005)
22. Rusu, I.: Sorting signed permutations by reversals using link-cut trees. Information Processing Letters 132, 44–48 (Apr 2018)
23. Satish, N., Harris, M., Garland, M.: Designing efficient sorting algorithms for manycore gpus. In: 2009 IEEE International Symposium on Parallel & Distributed Processing. pp. 1–10. Rome, Italy (May 2009)
24. Shabaz, M., Kumar, A.: Sa sorting: a novel sorting technique for large-scale data. Journal of Computer Networks and Communications pp. 1–7 (2019)
25. Shutler, P.M.E., Sim, S.W., Lim, W.Y.S.: Analysis of linear time sorting algorithms. The Computer Journal 51(4), 451–469 (Jul 2008)
26. Singh, D.P., Joshi, I., Choudhary, J.: Survey of gpu based sorting algorithms. International Journal of Parallel Programming 46(6), 1017–1034 (Apr 2018)
27. Yang, M., Zhang, P., Fang, J., Liu, W., Huang, C.: thsort: an efficient parallel sorting algorithm on multi-core dsps. CCF Transactions on High Performance Computing 20(4), 1–16 (Jan 2024)

References

1. Abdel-Hafeez, S., Gordon-Ross, A.: An efficient o(n) comparison-free sorting algorithm. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 25(6), 1930–1942 (Jun 2017)
2. Agapitos, A., Lucas., S.M.: Evolving efficient recursive sorting algorithms. In: 2006 IEEE International Conference on Evolutionary Computation. pp. 2677–2684. Vancouver, BC, Canada (Jul 2016)
3. Alhabboub, Y., Almutairi, F., Safhi, M., Alqahtani, Y., Almeedani, A., Alguwaifli, Y.: Accelerating large-scale sorting through parallel algorithms. Journal of Computer and Communications 12(1), 131–138 (Jan 2024)
4. Cederman, D., Tsigas, P.: A practical quicksort algorithm for graphics processors. In: European Symposium on Algorithms. pp. 246–258 (2008)
5. Chen, S., Qin, J., Xie, Y., Zhao, J., Heng, P.A.: A fast and flexible sorting algorithm with cuda. In: International Conference on Algorithms and Architectures for Parallel Processing. pp. 281–290. Taipei, Taiwan, China (Jun 2009)

6. Cormen, T., C.Leiserson, Rivest, R., C.Stein: Introduction to Algorithms. MIT press (2009)
7. Faujdar, N., Ghrera, S.P.: Analysis and testing of sorting algorithms on a standard dataset. In: 2015 Fifth International Conference on Communication Systems and Network Technologies. pp. 1–10. Gwalior, India (Apr 2015)
8. Fredman, M.L.: An intuitive and simple bounding argument for quicksort. Information Processing Letters 3(114), 137–139 (Mar 2014)
9. Hammad, J.: A comparative study between various sorting algorithms. International Journal of Computer Science and Network Security 15(3), 358–367 (Mar 2015)
10. Idrizi, F., Rustemi, A., Dalipi, F.: A new modified sorting algorithm: A comparison with state of the art. In: 2017 6th Mediterranean Conference on Embedded Computing. pp. 1–6. Bar, Montenegro (Jul 2017)
11. Jing, Y.N., Tu, P., Wang, X.P., Zhang, G.D.: Distributed-log-based scheme for ip traceback. In: The Fifth International Conference on Computer and Information Technology. pp. 711–715. Shanghai, China (Dec 2005)
12. Jing, Y.N., Tu, P., Wang, X.P., Zhang, G.D.: Space-efficient data structures for top-k completion. In: Proceedings of the 22nd international conference on World Wide Web. pp. 583–594. Rio de Janeiro, Brazil (May 2013)
13. Jugé, V.: Adaptive shivers sort: an alternative sorting algorithm. ACM Transactions on Algorithms 20(4), 1–55 (Aug 2024)
14. Jukna, S., Seiwert, H.: Sorting can exponentially speed up pure dynamic programming. Information Processing Letters 159, 451–469 (Apr 2020)
15. Kapur, E., Kumar, P., Gupta, S.: Proposal of a two way sorting algorithm and performance comparison with existing algorithms. International Journal of Computer Science, Engineering and Applications 2(3), 61–78 (Jun 2012)
16. Klein, S.T.: On the connection between hamming codes, heapsort and other methods. Information Processing Letters 113(17), 617–620 (May 2017)
17. Kocher, G., Agrawal, N.: Analysis and review of sorting algorithms. International Journal of Scientific Engineering and Research 2(3), 81–84 (Mar 2014)
18. Louza, F.A., Gog, S., Telles., G.P.: Optimal suffix sorting and lcp array construction for constant alphabets. Information Processing Letters 118, 30–34 (Sep 2017)
19. Múller, I., Sanders, P., Lacurie, A., Lehner, W., Fárber, F.: Cache-efficient aggregation: Hashing is sorting. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. p. 1123–1136. Melbourne, Bictoria, Australia (May 2015)
20. Peters, H., Schulz-Hildebrandt, O., Luttenberger, N.: Fast in-place, comparison-based sorting with cuda: a study with bitonic sort. Concurrency and Computation: Practice and Experience 23(7), 681–693 (Jan 2011)
21. Puglisi, S.J., Smyth, W.F., Turpin, A.: The performance of linear time suffix sorting algorithms. In: Data Compression Conference. pp. 358–367. Snowbird, UT, USA, USA (Mar 2005)
22. Rusu, I.: Sorting signed permutations by reversals using link-cut trees. Information Processing Letters 132, 44–48 (Apr 2018)
23. Satish, N., Harris, M., Garland, M.: Designing efficient sorting algorithms for manycore gpus. In: 2009 IEEE International Symposium on Parallel & Distributed Processing. pp. 1–10. Rome, Italy (May 2009)
24. Shabaz, M., Kumar, A.: Sa sorting: a novel sorting technique for large-scale data. Journal of Computer Networks and Communications pp. 1–7 (2019)
25. Shutler, P.M.E., Sim, S.W., Lim, W.Y.S.: Analysis of linear time sorting algorithms. The Computer Journal 51(4), 451–469 (Jul 2008)
26. Singh, D.P., Joshi, I., Choudhary, J.: Survey of gpu based sorting algorithms. International Journal of Parallel Programming 46(6), 1017–1034 (Apr 2018)
27. Yang, M., Zhang, P., Fang, J., Liu, W., Huang, C.: thsort: an efficient parallel sorting algorithm on multi-core dsps. CCF Transactions on High Performance Computing 20(4), 1–16 (Jan 2024)

Hui Liu received the PhD degree in Information and Communication Engineering from Hohai University, Nanjing, China in 2021. From 2003 to 2021, he served as a full-time faculty member at Jiangxi University of Science and Technology. He is currently an Associate Professor with the School of Electronic Information, Foshan Polytechnic, China. His current research interests include deep learning, computer image processing, and big data analysis.

Yi Cao received the MSc degree in Software Engineering from the School of Computer Science and Information Engineering, Anhui Normal University, in 2022. From 2022 to date, he has been a full - time teacher, assistant professor, and senior engineer at the School of Electronics and Information, Foshan Vocational and Technical College. His current research interests include blockchain technology and big data analytics.

Zehan Cai is a Class of 2022 student at the School of Electronic Information, Foshan Polytechnic, China. His research focuses on machine learning.

Jie Chen received the BSc degree in Software Engineering, MSc degree and PhD degree in Computer Science from Sichuan University, Chengdu, China, in 2005, 2008 and 2014, respectively. From 2008 to 2009, he was with Huawei Technologies Co., Ltd. as a software engineer. He is currently an Associate Professor with the College of Computer Science, Sichuan University, China. His current research interests include machine learning, big data analysis, and deep neural networks.

Hua Mao received the B.S. degree and M.S. degree in Computer Science from University of Electronic Science and Technology of China (UESTC) in 2006 and 2009, respectively. She received her Ph.D. degree in Computer Science and Engineering from Aalborg University, Denmark in 2013. She is currently a Senior Lecturer in Department of Computer and Information Sciences, Northumbria University, U.K. Her current research interests include Deep Neural Networks and Big Data.

Received: December 01, 2024; Accepted: February 28, 2025.

Computer Science and Information Systems 22(3):945–970 https://doi.org/10.2298/CSIS250131052Z

PFLIC: A Novel Personalized Federated Learning-Based Iterative Clustering

Shiwen Zhang[1,2], Shuang Chen[1,2], Wei Liang[1,2], Kuanching Li[1,2,⋆], Arcangelo Castiglione[3], and Junsong Yuan[4]

[1] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[2] Sanya Research Institute, Hunan University of Science and Technology, Sanya 572024, China
{shiwenzhang,wliang,aliric}@hnust.edu.cn
shuangchen@mail.hnust.edu.cn
[3] Department of Computer Science, University of Salerno, Fisciano, SA, Italy
arcastiglione@unisa.it
[4] University at Buffalo, State University of New York, Buffalo 14201, New York, USA
jsyuan@buffalo.edu

Abstract. Federated learning (FL) is a machine learning framework that effectively helps multiple organizations perform data usage and machine learning models while meeting the requirements of user privacy protection, data security, and government regulations. However, in practical applications, existing federated learning mechanisms face many challenges, including system inefficiency due to data heterogeneity and how to achieve fairness to incentivize clients to participate in federated training. Due to this fact, we propose PFLIC, a novel personalized federated learning based on an iterative clustering algorithm, to estimate clusters to mitigate data heterogeneity and improve the efficiency of FL. It is combined with sparse sharing to facilitate knowledge sharing within the system for personalized federated learning. To ensure fairness, a client selection strategy is proposed to choose relatively "good" clients to achieve fairer federated learning without sacrificing system efficiency. Extensive experiments demonstrate the superior performance and effectiveness of the proposed PFLIC compared to the baseline.

Keywords: Federated learning, Clustering algorithm, Client Selection, Sparse sharing.

1. Introduction

Contemporary mobile smart technologies have reached an unprecedented level of sophistication. This technological evolution enables ubiquitous smartphone/sensor applications [17], which collectively produce real-time data streams at an extraordinary scale [32, 34]. The question of how to utilize and store these data has become a hot topic across industries. In traditional deep learning, these data are collected and stored in a central location to train neural networks. However, in some special cases, the data owner is reluctant to share these data with a third party due to privacy protection. To solve this problem, a promising distributed machine learning method, Federated Learning (FL) is proposed [12]. The

⋆ Corresponding author

traditional framework of FL is illustrated in Fig. 1, which learns a global model that aggregates information from each client while protecting participants' privacy.

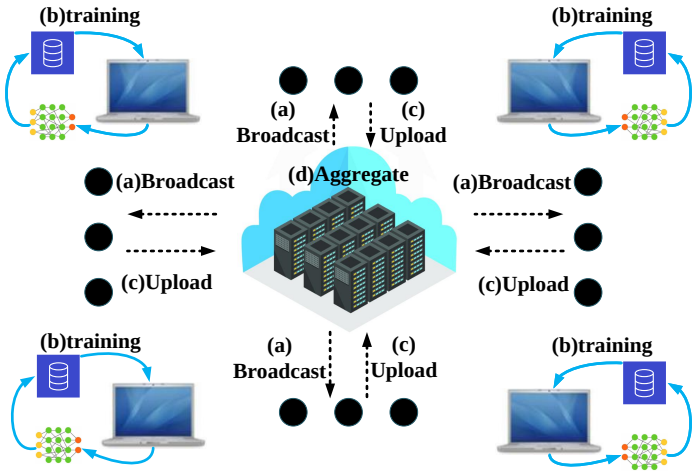


Fig. 1. A general FL framework. (a) Cloud server broadcasts the global model. (b) Clients training model. (c) Clients upload the local model. (d) Cloud server aggregation model.

Although FL can solve a distributed machine learning model without anyone seeing or touching the raw data of each client, several problems need to be addressed for efficiency, especially those regarding data heterogeneity and fairness of FL. In real applications, different clients generate data in various ways, resulting in non-independent and identically distributed (non-iid) data among clients, also known as data heterogeneity [20, 33]. Several studies have found that data heterogeneity can seriously affect the system's convergence and the model's accuracy [19, 20]. On the other hand, it is necessary to ensure FL's fairness by not favoring either party while achieving global knowledge sharing. If fairness cannot be guaranteed, some clients with relatively small contributions, *i.e.*, those less involved in FL training, will terminate their participation at any time. Therefore, how to design an efficient federated learning framework that solves data heterogeneity and protects fairness is of paramount importance.

Data heterogeneity is one of the central issues in the development of FL. To address this problem, many researchers have devoted themselves to designing a number of FL schemes [2, 19, 20, 23, 29]. To mitigate the impact of data heterogeneity, some have used Personalized Federated Learning (PFL) based on clustering algorithm [2, 23, 29] to cluster clients with similar data distributions and divide them into the same cluster to train a model dedicated to each cluster. Others have penalized local models that deviate too much from the global model to prevent local models from deviating from the global model [12, 15, 16, 31]. In addition, some have processed client data by dividing the dataset or data augmenting to synthesize regular datasets that correct for the heterogeneity of private

data for the client [5, 27]. Nevertheless, the penalty mechanism and data augmentation cannot solve the data heterogeneity problem. On the contrary, the clustering algorithm starts from data distribution, clusters clients with the same data distribution, and trains a cluster model for each cluster instead of a single global model, which solves the problem caused by data heterogeneity. However, the centralized clustering algorithm will bring the issue of high communication overhead, the iterative clustering algorithm can effectively reduce the overhead. The iterative clustering algorithms for PFL in FL systems introduce a new problem, *i.e.*, the inability of global knowledge sharing among groups with different data distributions.

How to incentivize active client participation in training and achieve equity is another central issue in the development of FL. To solve this problem, some pay more attention to the fairness of the results and expect to achieve the equilibrium fairness [9, 10, 25], *i.e.*, the model performs equilibrium among the clients and does not favor a specific data party. However, the robust data side is more likely to be selected in this process, and the final trained global model will also be biased in favor of the firm side, resulting in the data's inferior side being ignored. Other researchers focus more on the fairness of the process and want to achieve contribution fairness [18, 22], *i.e.*, allowing differences in model performance but pursuing balanced contributions from all parties. Nevertheless, this may cause dissatisfaction among high-contributing clients, leading to the problem of "free-riding" [26] and affecting the sustainable development of FL. For the sustainability of FL, we need to be unbiased towards either party, ensure fairness, and allow for the case where models perform differently. Hence, it is still challenging to mitigate the impact of data heterogeneity, following the client's most primitive features and ensuring the system's fairness without increasing the communication overhead.

Unlike previous studies, we design PFLIC, a novel personalized federated learning based on an iterative clustering algorithm in this work, to effectively address the issue of data heterogeneity and achieve the fairness of FL. We leverage the similarity among clients to implement iterative clustering and thus improve the convergence speed of such a system to address data heterogeneity. Besides, we integrate weight sharing in multi-task learning to enable PFL and facilitate inter-cluster collaboration to minimize communication overhead. To maintain the fairness of FL, we design a client selection strategy to select the clients to participate in the training process to guarantee the participation rate of clients in the training process. This work mitigates the data heterogeneity problem while maintaining FL's fairness without sacrificing accuracy. The contributions of this work are listed as follows:

- To solve the problem of data heterogeneity and system unfairness, minimizing system overhead, and enabling personalized federated learning, we propose and design PFLIC.
- To alleviate the heterogeneous data problem, we design an iterative clustering algorithm to cluster customers with high similarity by continuously determining their identities before training. To achieve PFL, we combined the weight sharing to drop to achieve knowledge learning between clusters and reduce the system's communication overhead.
- To ensure the fairness of federated learning, we design a client selection strategy to actively select "good" clients according to the established metrics. This means that

the weak clients are no longer feeble. And the model not only equalizes performance between clients but also allows for variance.
– To demonstrate the performance of PFLIC in terms of accuracy and efficiency, we conduct extensive experiments on real-world datasets. Compared to the baseline, experimental results show that the proposed method achieves promising results.

The remainder of this article is organized as follows. The related work is introduced in Section II, the system model and problem formulation are described in Section III, the construction and workflow of PFLIC are presented in Section IV. Experimental results and analysis are provided to show the superiority of PFLIC over the baseline methods in Section V. Finally, the concluding remarks and future directions are given in Section VI.

2. Related work

2.1. Data Heterogeneity in Federated Learning

A central issue in developing FL systems in recent years is heterogeneity, categorized into data heterogeneity and structural heterogeneity. To solve the problem of poor model performance due to data heterogeneity, several works [2,7,19,23] consider using clustering to address data heterogeneity. Clustering Federated Learning (CFL) is a promising approach to solving the data heterogeneity problem. Sattler *et al.* [19] proposed a methodological framework for federated client grouping learning, an algorithm in which the parameter server dynamically divides the participants based on their gradient or update information. However, the server has a high computational cost. Tu *et al.* [23] dynamically learned personalized models for different users by learning the similarity between user model weights to form a shared structure. Briggs *et al.* [2] combined hierarchical clustering with FL to separate client clusters by calculating the similarity of the client's local updates to the global model. After separation, the clusters are trained independently and in parallel on specialized models. Tu *et al.* [23] and Briggs *et al.* [2] divided the clients into clusters by calculating the distance of the local model weights, which improves the accuracy but brings about slower convergence. Ghosh *et al.* [7] divided similar client data distributions into a cluster, but their clustering results are unstable and have some impact on the model accuracy. Unlike the single clustering algorithm mentioned above, we use an iterative clustering algorithm, which reduces the high communication overhead associated with centralized clustering algorithms.

2.2. Fairness in Federated Learning

Client selection has become an emerging topic that addresses fairness in FL. It chooses which clients will participate in each round of training. If all clients are involved in each round of training, the communication cost will be high. Thus, using a client selection strategy is also an excellent way to reduce the cost of communication. A good selection strategy can improve the model accuracy, reduce the training cost, and enhance the fairness of the system [1,11,13,14,21,28]. Tang *et al.* [13], Lai *et al.* [21] and Cho *et al.* [11] improved the convergence speed of the model by implementing the client selection function. Tang *et al.* [13] argued that the clients do not contribute equally and do not contribute independently, so they use the loss correlation of clients for client selection.

Lai *et al.* [21] designed the Oort framework that allows developers to specify on their own what kind of federated learning clients can be added, combining fairness and statistical usage. Cho *et al.* [11] made a trade-off between convergence speed and solution bias and found that biasing client selection toward clients with higher local loss of clients achieves faster error convergence. Xu *et al.* [28] argued that optimizing the learning performance depends critically on how the clients are selected, but it only considers the data heterogeneity. Li *et al.* [14] selected clients to achieve fairer network performance. Li *et al.* [14] and AbdulRahman *et al.* [1] selected clients based on different strategies to improve the global accuracy of the model and the speed of convergence. However, they may destroy the clustering structure.

2.3. Personalized Federated Learning

The strategies for implementing personalized federated learning can be divided into global model personalization and learning customized models. The former intends to enhance the performance of a global model federally trained on heterogeneous data. Wu *et al.* [27] augmented local datasets using a self-encoder, which enhances the usability of the local dataset to represent the overall data distribution, but with the possibility of privacy leakage. Wang *et al.* [24] used deep learning for training to select the participating clients to mitigate the effect of non-iid data. This approach samples from a more homogeneous data distribution, improving model generalization performance even though it may incur higher computational costs. The latter is intended to provide personalized solutions by modifying the FL model aggregation process to build customized models. Bui *et al.* [3] considered personalized feature representations for each client by using users as private model parameters but are limited in supporting a high degree of model design personalization. Annavaram *et al.* [8] proposed population knowledge transfer through a bidirectional distillation approach using alternating minimization to train local and global models to support personalized model architectures for each client, but this can lead to inferior training of student models if there are too many differences between the teacher model and the student model. With the help of sparse sharing techniques, we allow knowledge learning between different clusters to achieve personalized federated learning while guaranteeing the accuracy of models within this cluster.

3. Overview of PFLIC

This section presents the system model, the problem formulation, and the description of the design goal, while Table 1 summarizes notations commonly used throughout this work.

3.1. System Model

PFLIC is a novel personalized federated learning scheme that can solve the problem of non-iid data, improve the model's accuracy and convergence speed of the system, and reduce communication costs. Specifically, we propose clustering before training to solve the excessive difference in data distribution of each user in the FL system. The process is iterated during training to prevent incorrect identity estimation at the first clustering. On this

Table 1. Summary of Notations

Notation	Explanation
$\hat{\theta_t^j}$	Model of client j in round t
θ_t^i	Cluster model i in round t
N	Number of clients
k	Number of clusters
c_i	One cluster i in the set of all clusters
P_t	Clients selects for training at round t
D	Total quantity of data
D_j	Data of client j
id_j	The identity of client j
$\mathcal{L}_j$	Loss function of client j
C_j	The utility value of client j
S_j	The total number of training rounds for a client
a_j	Accuracy of client j
η	Learning rate
T	Training rounds
TS	Threshold for client participation in training
V	Model accuracy distribution variance
A	Global model mean test accuracy

basis, we propose to combine sparse sharing to reduce communication costs and improve convergence. Finally, we design a client selection strategy to actively select clients participating in training to achieve fairer federated learning. Fig. 2 shows the overall system model of PFLIC, which can be categorized into two cases:

The first case is when the clustering results are not stable. Firstly, the cloud server broadcasts k cluster models to the client; then the client uses the received models and local data to estimate the cluster identity, uses the local data to update the models, and uploads the trained models to the cloud server; secondly, the server determines whether the clustering results are stable or not, and aggregates the received models within each cluster.

The second scenario is that the clustering results are already stable. Firstly, the cloud server broadcasts the weight subsets and shared layers of the models of the clusters of the class to the clients; then the clients use local data to update the models and upload the trained models to the cloud server; finally, the cloud server performs client selection, as well as aggregation of the received models within each cluster respectively.

More precisely, the whole process of the program involves two interactions between the server and the client. The first iteration is used for clustering and the second iteration is used for active client selection and model updating.

3.2. Problem Formulation

One center cloud server and N clients exist in personalized federated learning settings. The cloud server and clients can communicate using a predefined communication protocol. Clients have different data, denoted as $\{D_1, D_2,...,D_N\}$. In this work, we assume

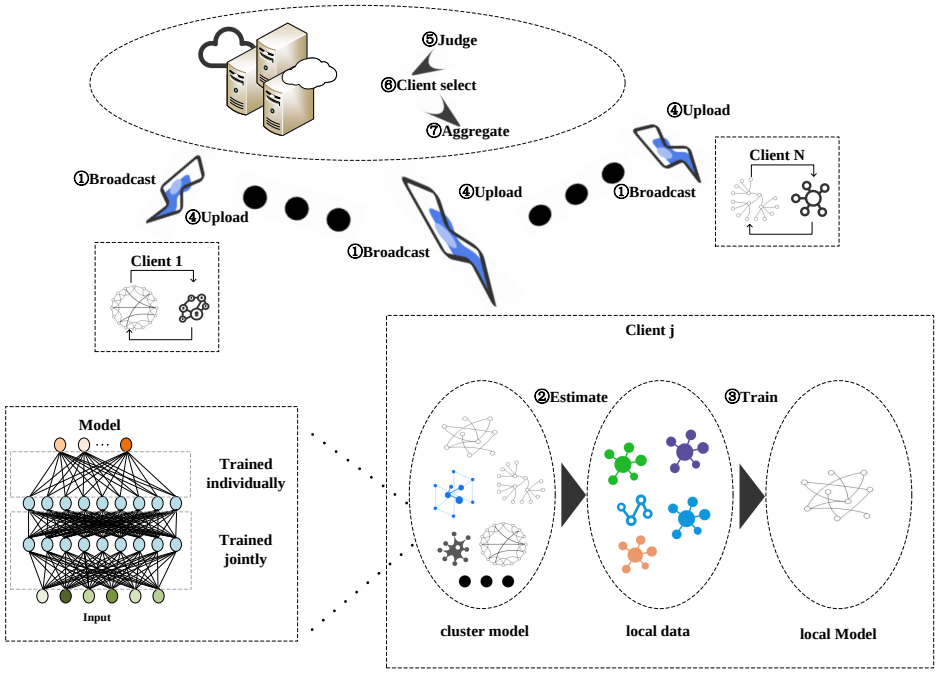


Fig. 2. System model of PFLIC

that there are potential clustering relationships between clients' data, which can be divided into k clusters, denoted as $\{ c_1, c_2,...,c_k \}$, whereas the goal is to learn k good cluster models θ^{i*}:

$$\theta^{i*} = argmin_{j\in[k]}\mathcal{L}_j(\theta_i), i \in [k]\,. \tag{1}$$

For each cluster by combining information from all cluster classes without data exchange, where $\mathcal{L}()$ is a loss function.

3.3. Design Goal

The scheme not only effectively solves the data heterogeneity problem and achieves personalized federated learning but also ensures the fairness of FL. Specifically, the FL scheme we designed needs to satisfy the following design goals:

Accuracy: In the absence of other contingencies, the scheme cannot sacrifice the accuracy of the global model. Compared with the baseline, the accuracy of our proposed scheme should be consistent or better.

Efficiency: Due to the limited computational resources of the edge devices, it cannot generate too much extra computational overhead and communication overhead. Compared with the baseline, we propose that the scheme should not add too much workload to the participants.

Fairness: Since the clients participating in the training are self-interested and differ from each other in terms of computational communication resources, data, and others. The sustainability of federated learning needs to maximize client incentives, distribute rewards appropriately, and promote motivation among federated participants. In other words, we need to ensure a certain level of fairness in the system.

4. Design of PFLIC

4.1. Workflow of PFLIC

When the clustering results are not stabilised, the first stage in Fig. 3 is performed. The cloud server distributes k cluster models (line 4). After receiving the cluster model, the server estimates the identity. After the server estimates the cluster identity, it trains using that cluster model and local private data (lines $7-9$). After training, the client uploads the cluster identity and the updated model to the cloud server (line 10). The cloud server aggregates the cluster model separately (lines $26-28$).

When the clustering results are stable, the second phase in Fig. 3 is performed. The cloud server distributes to each participating client the weight subset of the cluster model to which the client belongs and the shared layer (line 18). The client is trained using this model and local data (line 21). After training, the client uploads the trained model parameters to the cloud server (line 16). The cloud server calculates and ranks the utility values of the client and uses the ranked values for client selection(Line 17). The cloud server aggregates the clustered models respectively and uses these aggregated models to generate the shared layer (Lines $27-28$).

4.2. Client Side

FL system is trained jointly on clients having different datasets, where each client dataset has different samples and different kinds of features. Direct model aggregation for models trained with non-iid user datasets affects the model's overall performance, slowing convergence. Thus, we adopt an essential assumption that there are potential clustering relationships between the data of individual clients involved in training. Our goal is to utilize the similarity (gradient) between the client data samples to cluster the clients with higher similarity for training to improve the model's convergence speed and model accuracy for this system.

Since one-shot clustering is prone to chance errors, and once a wrong clustering estimate is generated, it cannot be corrected in any subsequent training phase, it will impact the whole FL system training. Therefore, iterative clustering is used in this paper. Before the clustering results are stabilized, the cloud server performs a clustering analysis before aggregating the models. While training during training, the clustering results are dynamically adjusted according to the model parameters $\widetilde{\theta}_t$. Fig. 4 compares primary and iterative clustering.

Algorithm 2 gives pseudo-code for iterative clustering. The global model obtained in this step depends on the clustering and client selection results. In the t-th training round where the clustering results are not stabilized, all clients receive models $\theta_t^i (i \in [k])$

Algorithm 1 PFLIC

Input: initialize parameters $\theta_0^i (i \in [k])$, number of all clients N, learning rate η, number of clusters K

for all $t = 0, 1, ..., T$ **do**
  SERVER SIDE:
  if the clustering results are not stabilized **then**
    Broadcast k models $\theta_t^i (i \in [k])$
    CLIENT SIDE:
    for all each client $i \in [k]$ **do**
      Compute $\hat{i} = argmin_{i \in [k]} \mathcal{L}_j(\theta_t^i, D_j)$
      Estimated $id_j = \{id_{i,j}\}_{i=1}^k, id_{i,j} = 1\{i = \hat{i}\}$
      Compute $\hat{\theta_t^j} = \theta_t^i - \eta \nabla \mathcal{L}_j(\theta_t^i, D_j)$
      Send $\hat{\theta_t^j}$,id_j to server
    end for
    SERVER SIDE:
    Cluster $\{\hat{\theta_t^j}\}_{i=1}^N$ into $c_1, c_2, ...c_k$
  else
    SERVER SIDE:
    Clients selection using Algorithm 3
    $P_t = participatingclients$
    Broadcast one shared layer and k subsets of different versions of weights
    CLIENT SIDE:
    for all each client $i \in [k]$ **do**
      Compute $\hat{\theta_t^j} = \theta_t^i - \eta \nabla \mathcal{L}_j(\theta_t^i, D_j)$
      Send $\hat{\theta_t^j}$ to server
    end for
  end if
  SERVER SIDE:
  for all each cluster$(c_1, c_2, ...c_k)$ in parallel **do**
    $\theta_{t+1}^i = \theta_t^i + \sum_{j=1}^N \frac{D_j}{D} \hat{\theta_t^j}, i \in [k]$
    Cloud server generates shared layers using cluster model
  end for
end for

Algorithm 2

Input: number of clients N, loss function $\mathcal{L}$, number of clusters k, clients P_t participating in training at round t

Server broadcast $\theta_t^i, i \in [k]$
Clients($j \in P_t$) for identity estimation and training
Clients($j \in P_t$) send model $\hat{\theta_t^j}$ and id_j to the CS
for all each each cluster$(c_1, c_2, ..., c_k)$ in parallel do **do**
  $\theta_{t+1}^i = \theta_t^i + \sum_{j=1}^N \frac{D_j}{D} \hat{\theta_t^j}, i \in [k]$
end for

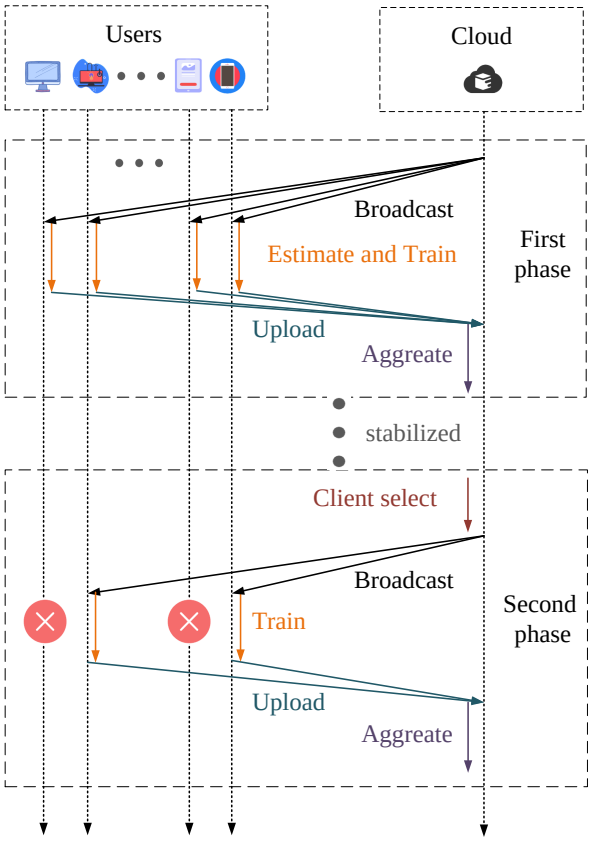


Fig. 3. Workflow of PFLIC

broadcast from the cloud server and use these models and its local empirical loss function $\mathcal{L}_j()$ to find the model parameter that minimizes loss by $\hat{i}$:

$$\hat{i} = argmin_{i\in[k]}\mathcal{L}_j(\theta_t^i, D_j), j \in [N]\,. \tag{2}$$

The identity id_j:

$$id_j = \{id_{i,j}\}_{i=1}^{k}\,. \tag{3}$$

$id_{i,j} = 1\{i = \hat{i}\}$ of this client is determined by using $\hat{i}$ to estimate which cluster this client is in after clustering. Then use $\theta_t^{\hat{i}}$ to perform stochastic gradient descent training by using local data to compute the model parameter $\hat{\theta_t^j}$ of $\mathcal{L}_j$. Therefore, the client sends the model parameter result $\hat{\theta_t^j}$ and the clustering identity id_j to the cloud server. When the clustering results are stabilized, all clients involved in the training receive the model $\theta_t^{\hat{i}}$ broadcast from a cloud server, train it by local stochastic gradient descent, and update the model. When a predetermined number of local training sessions is reached, the client uploads the parameters to a cloud server.

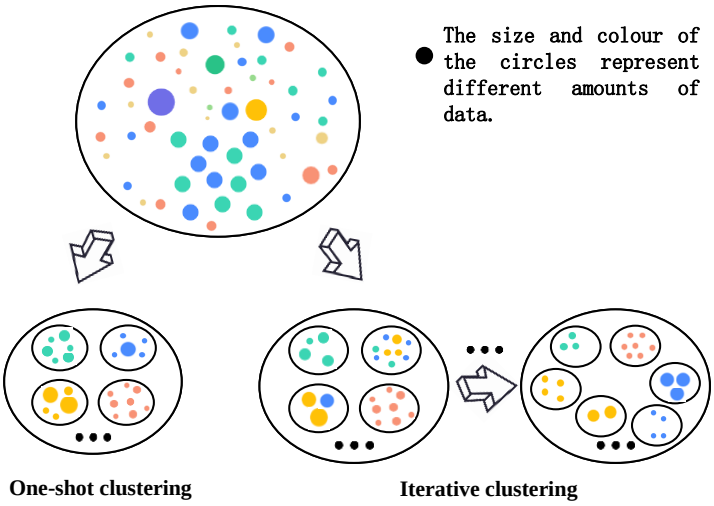


Fig. 4. One-shot clustering vs. Iterative clustering. *Left*: When one-shot clustering is performed, the client's participation in the training does not represent the client's overall data distribution, resulting in an incompletely accurate estimate of the client's identity. *Right*: When iterative clustering is performed, it is trained several times before clustering is performed.

4.3. Server Side: Broadcast

Compared to centralized machine learning, where computational costs dominate and communication costs are negligible, communication costs in FL are much higher than computational costs. Based on previous experience, the communication cost is directly related to the parameters transmitted among participants. To solve the problem of high communication costs in FL, we draw on the sparse sharing in sparse sharing proposed by SUN *et al.* [30] to allow the sharing of some parameters among different clusters to achieve PF and reduce the communication cost. Fig. 5 utilizes two representations of the sparse sharing mechanism to illustrate the concept of sparse sharing.

To reduce the communication cost, we combine sparse sharing by replacing k models broadcast by the cloud server with one shared layer and k subsets of different versions of weights, which reduces the transmitted model parameters and lowers the communication cost. Specifically, At the first FL system training round, the cloud server initializes k models $\theta_0^i (i \in [k])$. After that, these models are sent to all clients $j (j \in [N])$. In the t-th round of FL training where the clustering results are not stabilized, the cloud server sends k subsets of the models $\theta_t^i (i \in [k])$ and a shared layer to all clients. When the clustering results are stabilized, the cloud server sends a weighted subset of the model $\theta_t^i (i \in [k])$ with corresponding clusters and a shared layer to all clients selected to participate in the training.

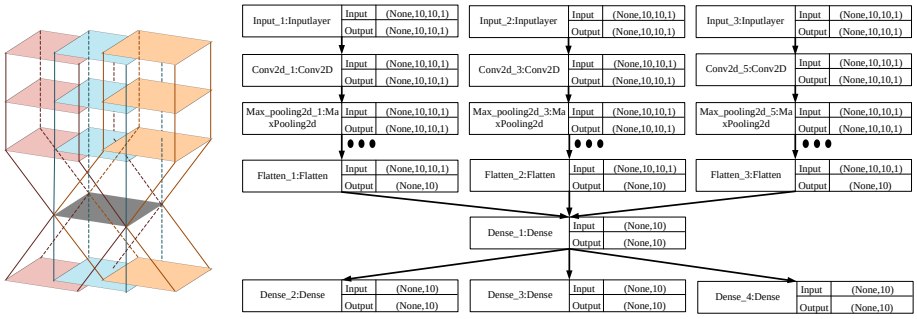


Fig. 5. Two representations of the same layered sharing mechanism case (Three models share a single layer). *Left*: Expressed using a convolutional neural network graph structure. *Right*: Expressed using a function called 'building block' form

We use sparse sharing, which not only reduces the number of transmitted parameters and lowers the communication cost; it also allows the sharing of task parameters between different clusters, breaks down the barriers between different clusters, facilitates the sharing of knowledge between clusters, and improves the convergence speed and accuracy of the system.

4.4. Server Side: Client Select

After receiving the model $\theta_t^i (i \in k)$, the client estimates its clustering identity through training using its experience loss function $\mathcal{L}()$. After receiving the uploaded identity estimation, the cloud server first determines whether the clustering result is stable or not. If it is not stable, the clustering continues. If it has been stable, no further clustering operation is performed for subsequent training, and client selection begins.

The previous client selection scheme is generally random [20], which leads to some clients with unique data distributions being challenging to select; there is little variability in the clients selected by extraction; some clients are extracted frequently, etc., which reduces the representativeness of the client population and makes the convergence of the global model more unstable. This limitation may affect the clustering results. To solve this problem, this paper proposes a client selection strategy, which selects relatively "good" clients to achieve fairer federated learning. Algorithm 3 outlines pseudo-code for this part.

To solve the problem raised above and improve the fairness of FL, this strategy selects clients with higher losses and considers their participation rounds simultaneously. Specifically, after the client uploads the parameters to a cloud server, the server calculates the value $\mathcal{C}$ of the client based on the received loss. Then, the server prioritizes all clients based on this value $\mathcal{C}$. Assuming that the total number of training rounds for a client in round t is S_j, we define this value $\mathcal{C}$:

$$\mathcal{C} = \frac{\sum_t^{S_j} \mathcal{L}_{j_t}}{S_j} . \tag{4}$$

Algorithm 3 Client Select

Input: number of clients N, loss function $\mathcal{L}$, number of clusters k, clients P_t participating in training at round t, participation rounds t_j for the t-th client, value of utility measures C_t^j, threshold TS for client participation in training

1: **while** $t_j < TS$ **do**
2: Add 1 to the number of rounds t_j for client j participating in the training
3: Clients P_t involved in training utilize local data for training
4: Compute the value of utility measures of the client j: $C_j = \sum_t^{S_j} \mathcal{L}_{j_t}$
5: Sorting the client's value of utility measures
6: Select the top n clients with the largest value C in each cluster for the next round of training
7: **end while**

After sorting, clients with larger value $\mathcal{C}$ will be selected to participate in training. According to the above equation, it can be concluded that clients with larger loss values have a greater chance to participate in training. This aspect makes the model accuracy distribution variance smaller, the client model accuracy distribution more balanced, and the FL system more fair. $\mathcal{V}$ is defined as

$$\mathcal{V} = \frac{\sum_{j=1}^{N}(a_j - \mathcal{A})^2}{N}. \tag{5}$$

Where $\mathcal{A} = \frac{\sum_{j=1}^{N} a_j}{N}$ is the global model average test accuracy, a_j is the accuracy of each participant, and N is the total number of clients. To ensure training efficiency, a participation threshold mechanism is implemented, limiting the maximum number of client engagements per training round. When a client's training rounds exceed this threshold, the client will no longer participate in training, and it would increase the participation rate of other clients.

The client selection strategy proposed in this section actively selects clients that participate in training, which reduces the variance of the model accuracy distribution and improves the accuracy of both the client and the global model; setting a threshold prevents clients from endlessly participating in training and enhances the participation rate of other clients.

4.5. Efficiency Analysis

The convergence of iterative clustering in FL has been demonstrated in previous studies [7]. Furthermore, Cho *et al.* [4] showed that a biased client selection strategy does not affect the convergence properties of FL. Therefore, a biased client selection framework also does not change the convergence property of CFL. Thus, we focus on evaluating the efficiency of our proposed PFLIC and baseline algorithms (FedAvg and CFL). The definitions used for our analysis are provided next.

Definition 1: Efficiency(E) is defined as the sum of the computational efficiency(E_{Cal}) and the communication efficiency(E_{Com}). Therefore, the efficiency can be written as follows: $E = E_{Cal} + E_{Com}$.

Definition 2: Computational efficiency(E_{Cal}) is defined as the total computational cost required to train the model to achieve the desired test accuracy threshold. Assuming

that the expected test accuracy threshold is Acc, the corresponding number of training rounds spent by the algorithm is denoted $round$. Additionally, the computational cost of one iteration of the algorithm is Cal. Therefore, the computational efficiency E_{Cal} can be written as follows: $E_{Cal} = Cal * round$.

Definition 3: Communication efficiency(E_{Com}) is defined as the total communication cost required to train the model to achieve the expected test accuracy threshold. Assuming that the expected test accuracy threshold is Acc, the corresponding number of training rounds spent by the algorithm is denoted as round. Furthermore, the communication cost of one iteration of the algorithm is Com. Therefore, the communication efficiency E_{Com} can be written as follows: $E_{Cal} = Cal * round$.

Based on the number of clients N, the number of clusters k, the participation rate ρ, the number of model parameters per participant P, and the number of training rounds $round$, we present the results of the computational efficiency for different algorithms in Table 2.

Table 2. Results on computational efficiency between different algorithms

Scheme	E_{Cal}
FL	$E_{Cal}^{FL} = N * Cal * round$
CFL	$E_{Cal}^{CFL} = N * Cal * round + N * k * log(N)$
PFLIC	$E_{Cal}^{PFLIC} = N * (Cal + k) * t + N * \rho * Cal * (round - t)$

Computational efficiency: Each participant in FL performs local training, so the computational complexity of each round is the number of participants N multiplied by the training cost Cal of each participant. Therefore, the computational efficiency E_{Com} of FL can be written as follows:

$$E_{Cal}^{FL} = N * Cal\,. \tag{6}$$

Clustering federated learning requires clustering operations in addition to the complexity of local training. Assuming that the algorithm complexity used for clustering is $O(N * log(N))$, then the computational efficiency E_{Cal} of CFL can be written as follows:

$$E_{Cal}^{CFL} = N * Cal + N * k * log(N)\,. \tag{7}$$

The computational overhead of PFLIC is divided into pre-stabilization and post-stabilization computational overheads. Before stabilization, each participant in PFLIC performs local training and identity estimation, then the computational complexity is ${E_{Cal}^{PFLIC}}_{pre} = N * (Cal + k)$. After stabilization, PFLIC performs client selection, and the selected participant performs local training, then the computational complexity is ${E_{Cal}^{PFLIC}}_{post} = N * \rho * Cal$.

Communication efficiency: Federation learning requires each participant to send model parameters to the central server after each iteration round, so the communication complexity of FL can be written as follows:

$$E_{Com}^{FL} = N * P\,. \tag{8}$$

Clustering federation learning after clustering, only the clustering center communicates with the central server, so the communication complexity of CFL can be written as follows:

$$E_{Com}^{CFL} = K * P\,. \tag{9}$$

The communication overhead of PFLIC is divided into pre-stabilization and post-stabilization communication overhead. Before stabilization, each participant needs to send model parameters to the central server, so the communication complexity is $E_{Cal}^{PFLIC}{}_{pre} = N * P$. After stabilization, only the selected participants need to send model parameters to the central server, so the communication complexity is $E_{Com}^{PFLIC}{}_{post} = N * \rho * P$, we present the results of the communication efficiency for different algorithms in Table 3.

Table 3. Results on communication efficiency between different algorithms

Scheme	E_{Com}
FL	$E_{Com}^{FL} = N * P * round$
CFL	$E_{Com}^{CFL} = K * P * round$
PFLIC	$E_{Cal}^{PFLIC} = N * P * t + N * \rho * P * (round - t)$

5. Experimental Results

5.1. Experiment Settings

Models and Datasets: We conducted experiments on two real datasets, MNIST and CIFAR-10). To adhere to the assumption of a potential clustering relationship among cross-client data, we refer to Ghosh *et al.* [7] for rotating data on the MNSIT dataset. In MNIST and CIFAR-10 experiments, We used two Fully Connected Neural Networks (FCNN) and one Convolutional Neural Network (CNN) model that includes two convolutional layers followed by two fully connected layers. In the first FCNN model, we created two fully connected layers and chose to share the last fully connected layer. In the second FCNN model, we created three fully connected layers and chose to share the last fully connected layer. In the CNN model, we share one convolutional layer and utilize it to demonstrate the general applicability of our proposed algorithm.

Benchmarks: We compare the performance evaluation of our proposed algorithm with three well-known federated learning algorithms. The first comparison scheme is the Fedavg algorithm [20] with improved communication overhead, which reduces the communication overhead of the system by reducing the number of communication rounds in the federated learning process compared to the previous algorithms. We also compared it with a one-shot clustering algorithm [6]. [6] provides two different clustering algorithms based on sample size(one-shot-1) and model similarity(one-shot-2), which do not require any additional operations on the client side and can be seamlessly integrated into standard FL implementations. The fourth benchmark [7] is an iterative federated clustering algorithm that alternately estimates the clustering identities of users and optimizes the model parameters for user clustering via gradient descent.

Performance Metrics: We will focus on the loss value, the accuracy, the convergence rate, and the number of communication rounds. The effectiveness of the scheme is verified by the first three metrics, and the overhead of the scheme is illustrated by the last metric.

Experiment Parameters: In all experiments, we default the learning rate θ is 0.01. We default the number of local updates per epoch to $H = 10$. In the MNIST and CIFAR-10 datasets, we randomly distributed the data evenly across all clients. To avoid accidents, we used the average results of multiple independent experiments in all experiments.

5.2. Effects of the proposed scheme

Effect of Clients Number: To determine the impact of the amount of data owned by the user on our scheme, after fixing the number of clusters, we use the same dataset and set a different number of clients, then the data samples assigned to each client is also changed. The number of clients varies, and the amount of data the clients have also varies; thus, we determine whether we can affect the whole system's performance.

We cannot arbitrarily set the number of clients due to the effect of the sample size of the dataset itself. Thus, in the MNIST dataset, we set the number of clients m in the training set to 48, 96, and 192 and the corresponding number of clients in the test set to 8, 16, and 32. In the Cifar-10 dataset, we set the number of clients m in the training set to 50, 100, and 200 and the corresponding number of clients in the test set to 10, 20, and 40. In Fig. 6 (a), (b) and (e), (f), we compare the accuracy and loss values of our scheme for three different numbers of users. In Fig. 6 (c), (d) and (g), (h), we compare the standard deviation of the accuracy and the standard deviation of the loss values of our scheme for three different numbers of users. As shown in Fig. 6, the performance of m = 48, 96, and 192 is very close. By the time the run reaches 40 rounds, it has roughly stabilized with an accuracy of 97% and a loss below 0.1.

Intuitively, the size of the data volume owned by the client slightly affects the performance of our scheme, which was demonstrated experimentally when the proposed scheme converges and stabilizes to a sure accuracy after a certain number of training rounds.

Accuracy and Convergence: We designed two sets of experiments in which the performance metrics compared were accuracy and convergence speed. One set was used to compare the overall performance of PFLIC with the other three baseline methods, and one set used empirical loss to compare model robustness. We compare the model performance of each scheme for a certain number of communication rounds in Fig. 7 and 8, respectively.

In the MNIST dataset, it is observed that the iterative clustering idea plays a vital role in improving the performance of the federated learning system, both for training CNN models and FCNN models. In terms of accuracy, when training the CNN model, as shown in (a) in Fig. 7, our scheme and IFCA reach 99.40% after stabilization, while the one-shot clustering scheme and the traditional federated learning scheme are below 80%. When training the FCNN model, as shown in Fig. 7 (b) and (c), our scheme PFLIC is also more accurate than the clustered federated learning and traditional federated learning schemes. Regarding convergence speed, PFLIC and IFCA schemes with iterative clustering ideas are far better than the single and traditional federated learning schemes. Our scheme has converged and stabilized in the fiftieth round, while the conventional federated learning and one-shot clustering schemes are still fluctuating. Due to the data distribution imbal-

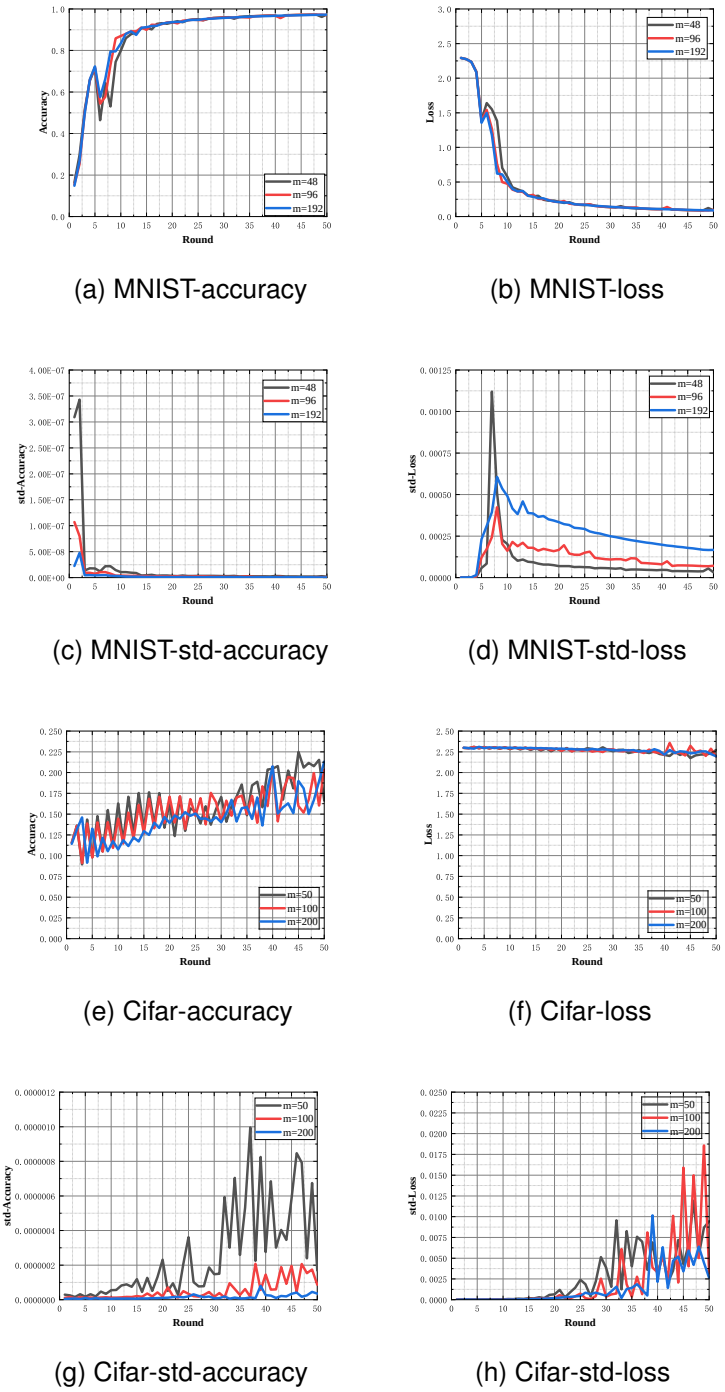
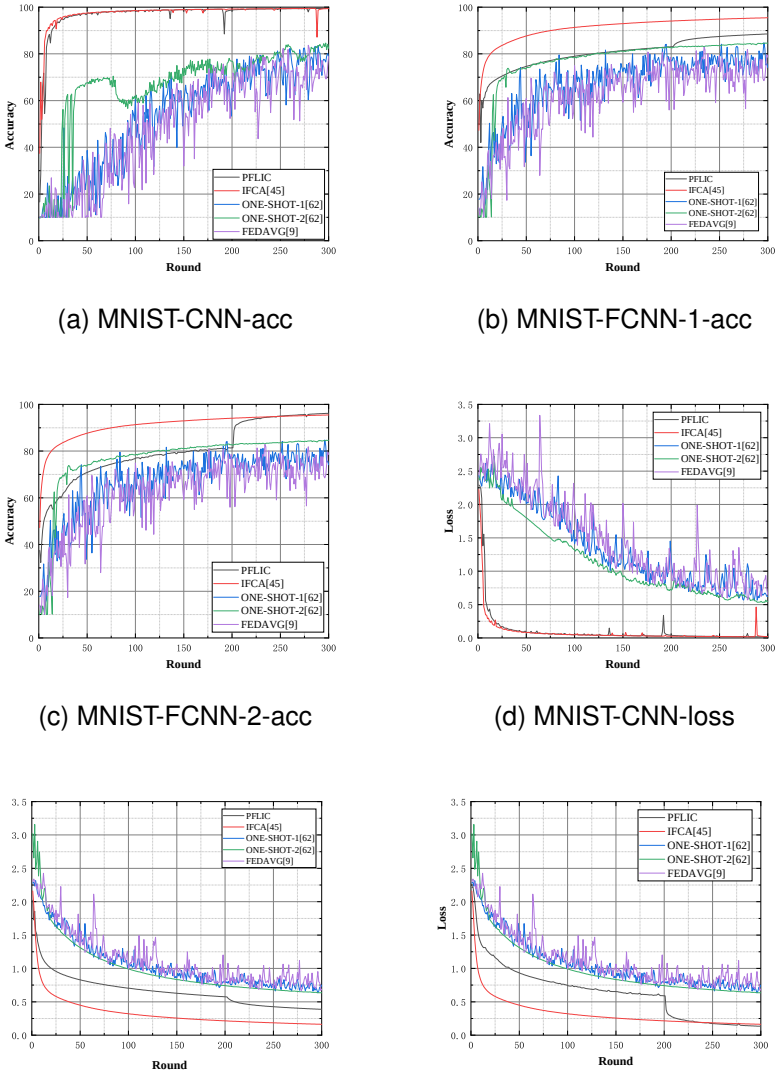


Fig. 6. Comparison of PFLIC schemes under different number of clients (m = 48, 96, 192). (a), (b), (c), and (d) are compared in the MNIST dataset. (e), (f), (g) and (h) are compared in the Cifar-10 dataset. Specifically, (a) and (e) compare the accuracy of our scheme for three different numbers of clients under different datasets, respectively. (b) and (f) Compare the loss of our scheme for three different numbers of clients under different datasets, respectively. (c) and (g) Compare the standard deviation of the accuracy of our scheme for three different numbers of clients under different datasets, respectively. (d) and (h) Compare the standard deviation of the loss of our scheme for three different numbers of clients under different datasets, respectively

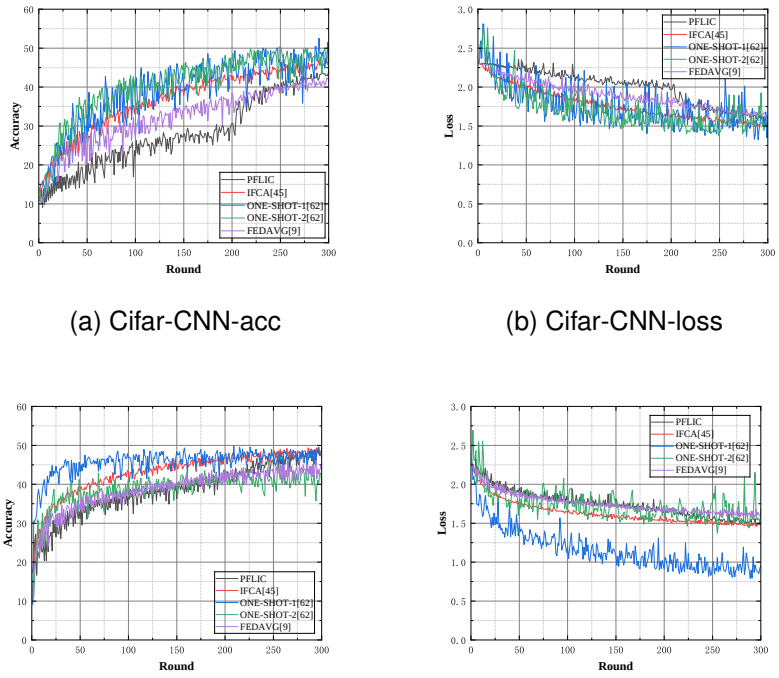


(a) MNIST-CNN-acc (b) MNIST-FCNN-1-acc

(c) MNIST-FCNN-2-acc (d) MNIST-CNN-loss

(e) MNIST-FCNN-1-loss (f) MNIST-FCNN-2-loss

Fig. 7. Comparison of our scheme PFLIC with FL, one-shot CFL, and IFCA regarding scheme performance. (a), (b), (c), (d), (e), and (f) compare the three schemes using CNN models and FCNN models under the MNIST dataset, respectively. Specifically, (a) illustrates the accuracy of training CNN models. (b) Illustrates the accuracy of training the first class of FCNN models (one layer shared and one layer trained separately). (c) Illustrates the accuracy of training the second class of FCNN models (one-layer shared and two layers trained separately). (d) Illustrates the loss of training CNN models. (e) Illustrates the loss of training the first class of FCNN models (one layer shared and one layer trained separately). (f) Illustrates the loss of training the second class of FCNN models (one layer shared and two layers trained separately)



(a) Cifar-CNN-acc (b) Cifar-CNN-loss

(c) Cifar-FCNN-acc (d) Cifar-FCNN-loss

Fig. 8. Comparison of our scheme PFLIC with FL, one-time CFL, and IFCA regarding scheme performance. (a), (b), (c), and (d) compare the three schemes using CNN models and FCNN models under the Cifar-10 dataset, respectively. Specifically, (a) illustrates the accuracy of training CNN models. (b) Illustrates the loss of training the CNN models. (c) Illustrates the accuracy of training FCNN models. (d) Illustrates the loss of training the FCNN models

ance in the client, the model briefly fluctuates after it stabilizes. Our scheme PFLIC shows fluctuation in around 190 rounds, while IFCA shows it in around 280 rounds.

In the Cifar-10 dataset, our scheme is less prominent in accuracy and convergence performance than before because it does not abide by the assumption that there is a potential cluster relationship between clients to set up the data in advance as the MNIST dataset does. When training the two models, it can be seen from Fig. 8 that after multiple rounds of training, our scheme still has some improvement in accuracy over the traditional federated learning scheme.

When constructing the model, we share some layers to promote knowledge sharing within clusters. This theoretically abandons the model's accuracy and the system's convergence speed to a certain extent, and the model's performance is slightly inferior to that of the IFCA scheme in specific implementations. However, in some specifics, our scheme PFLIC is pretty close to IFCA.

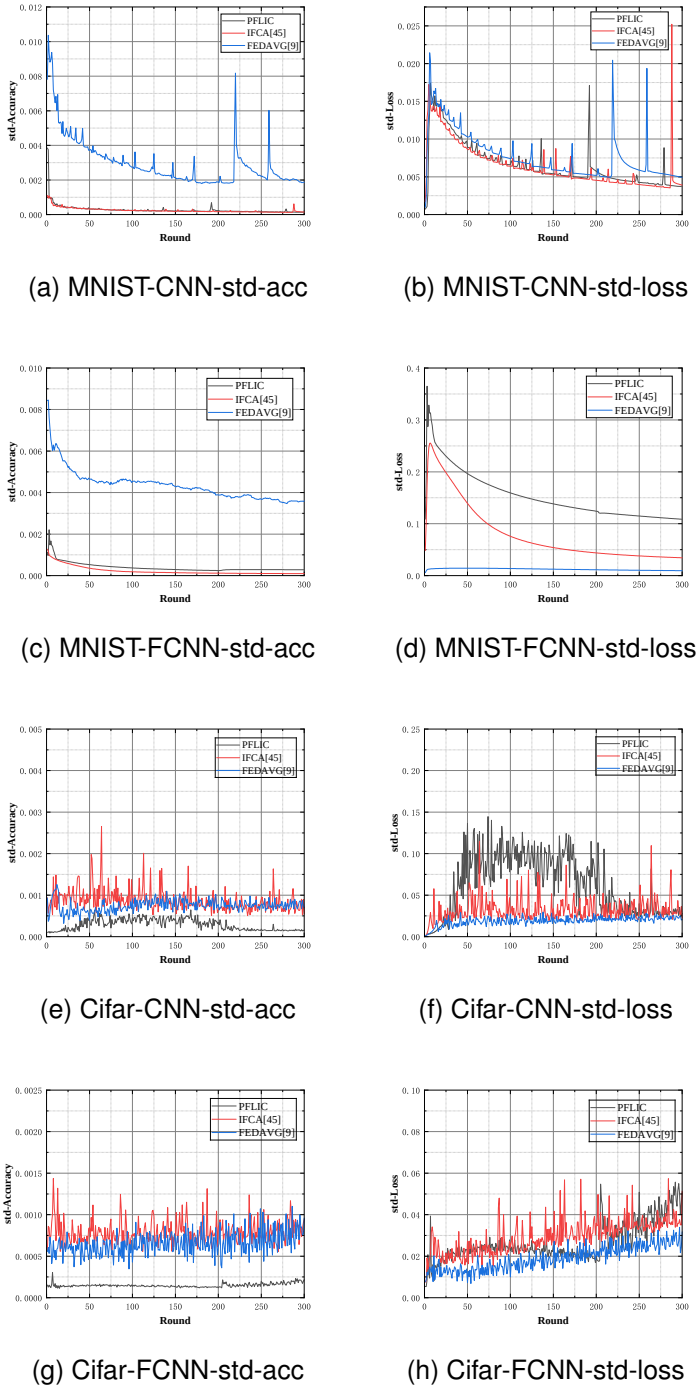
Overhead: We divide the overhead into communication and computation overhead. In this work, the communication overhead refers to the total amount of data and the number of communication rounds required to be transmitted for federated learning to reach a predefined performance metric (e.g., a specific accuracy value).

Regarding the number of communication rounds, we list the comparison results of the number of communication rounds required for different schemes to train the model to reach a specific accuracy for the first time under different datasets in Table 4. It can be seen that the two schemes, PFLIC and IFCA, which possess the idea of iterative clustering, outperform the other three schemes in the experimental results in the MNIST dataset that follows the assumptions. In the Cifar-10 dataset, which does not follow the assumptions, the clustered federated learning scheme (ont-shot-1 and one-shot-2) slightly outperforms the other three schemes. However, our scheme does not lag behind the traditional federated learning scheme either.

Table 4. Comparison of the number of communication rounds required to train a model to achieve a specific accuracy for the first time under different datasets for different schemes

	MNIST									**Cifar-10**					
	CNN			FCNN-1			FCNN-2			CNN			FCNN		
	30%	50%	80%	30%	50%	80%	30%	50%	80%	20%	30%	40%	20%	30%	40%
FL [20]	46	92	279	10	31	176	10	34	176	15	71	245	4	10	123
one-shot-1 [6]	26	92	259	10	18	121	5	26	132	13	41	69	2	3	12
one-shot-2 [6]	24	27	244	5	12	132	10	15	121	16	22	81	4	6	41
IFCA [7]	1	2	8	1	2	13	1	2	13	13	55	143	3	13	58
PFLIC	3	3	10	1	2	119	1	8	157	41	186	233	4	24	130

Regarding the total amount of data transmitted, traditional federated learning only needs to transmit one global model during each communication. Cluster federated learning must only transmit one identity-compliant cluster model to the client during each communication. The IFCA scheme transmits k cluster models during each round of communication. Our scheme PFLIC replaces the k models broadcast by the cloud server with a shared layer and k subsets of different version weights after the clustering is stabilized.



(a) MNIST-CNN-std-acc (b) MNIST-CNN-std-loss

(c) MNIST-FCNN-std-acc (d) MNIST-FCNN-std-loss

(e) Cifar-CNN-std-acc (f) Cifar-CNN-std-loss

(g) Cifar-FCNN-std-acc (h) Cifar-FCNN-std-loss

Fig. 9. Comparison of our scheme PFLIC with FL and IFCA regarding scheme performance. Specifically, (a) and (e) illustrate the standard deviation of the accuracy of training CNN models under different datasets. (b) and (f) Illustrate the standard deviation of the loss of training CNN models under different datasets. (c) and (g) Illustrate the standard deviation of the accuracy of training the first class of FCNN models (one layer shared and one layer trained separately) under different datasets. (d) and (h) Illustrate the standard deviation of the loss of training the first class of FCNN models (one layer shared and one layer trained separately) under different datasets

After the clustering is stabilized, our scheme only needs to transmit a specific model to the client. In FL systems, frequent data transmission inevitably brings about a rise in communication overhead. Compared to the communication overhead, the computational overhead is relatively small and will not be discussed in this paper.

Facts and theories show that although our scheme adds some overheads for knowledge sharing within the system compared to other schemes, the impact is irrelevant to the overall overheads.

Fairness: Since the clients involved in training are self-interested and differ from each other in terms of computational communication resources and data, among others, how to maximize client incentives, rationally distribute rewards, and promote motivation among federated participants is essential for sustainable federated learning. We utilize the standard deviation of accuracy and loss values to illustrate the fairness of our scheme.

Our scheme, PFLIC, selects clients with higher losses and simultaneously considers their number of participating rounds. Fig. 9 shows the results of the compared schemes trained with different models and different datasets. We compare the IFCA scheme, which also has the idea of iterative clustering, with the traditional federated learning scheme. The MNIST dataset that follows the hypothesis shows that the standard deviation of the model accuracy distribution of the two schemes with the iterative clustering idea is lower than the traditional federated learning scheme, indicating a more balanced distribution of client-side model accuracies. It can be seen on the Cifar-10 dataset that our scheme works better than the other two schemes that randomly select clients to participate in training.

Our scheme, PFLIC, actively selects clients to participate in training, which reduces the model accuracy distribution variance and makes the whole system more fair.

5.3. Summary of Experiment Results

From the above experimental results, the iterative clustering idea can significantly accelerate the convergence speed of the system and improve the accuracy of the model to some extent. It significantly outperforms the baseline scheme in both MNIST and Cifar-10 datasets, verifying the versatility of its algorithm design. PFLIC improves the performance and fairness of both CNN and FCNN structures, but its effect varies depending on the model characteristics (CNN improvement is more significant). However, it will increase the overhead of the whole federated learning system when there is no potential clustering relationship between clients.

The client selection strategy can to some extent can improve the fairness of the system, motivate the clients to participate in the training, and promote the enthusiasm of the federated learning participants. The knowledge sharing idea improves the convergence speed and accuracy of the system to a certain extent while allowing different clusters to share task parameters, breaking down the barriers between different clusters, and promoting knowledge sharing between clusters.

Through the synergy of client selection, robust aggregation and communication optimisation, the three core problems of data heterogeneity, communication bottleneck and fairness imbalance in federated learning are solved. All in all, our scheme solves the problem that a single global model cannot adapt to clients with different data distributions, and achieves Personalized federated learning, while ensuring the fairness of the federated learning system.

6. Concluding Remarks and Future Work

In this work, we design and implement a PFLIC scheme to accomplish a novel personalized federated learning. The proposed scheme designed an iterative clustering algorithm that utilizes the similarity among clients to solve the problem of data heterogeneity. It eliminates the chance of single clustering and reduces the computational overhead of single clustering. Subsequently, we combined sparse sharing to facilitate knowledge sharing among clusters and enable personalized federated learning. Moreover, we designed a client selection strategy to ensure the fairness of federated learning. By selecting the clients to participate in the training to ensure the participation rate of all clients in the training process, we prevent some clients from participating in the training process too frequently while others do not have the opportunity to receive training. Experimental results depict that the proposed algorithm performs better than the baseline.

In future directions, we will continue to explore more and better metrics for client selection and ways to deal with clients who drop out of the network due to unstable network conditions.

Acknowledgments. This work is supported by the Natural Science Foundation of Fujian Province under Grant 2022J05106, and the Natural Science Foundation of Hunan Province under Grant 2025JJ50399.

References

1. Abdulrahman, S., Tout, H., Mourad, A., Talhi, C.: Fedmccs: Multicriteria client selection model for optimal iot federated learning. IEEE Internet of Things Journal 8(6), 4723–4735 (2021)
2. Briggs, C., Fan, Z., Andras, P.: Federated learning with hierarchical clustering of local updates to improve training on non-iid data. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–9 (2020)
3. Bui, D., Malik, K., Goetz, J., Liu, H., Moon, S., Kumar, A., Shin, K.G.: Federated user representation learning (2019), https://arxiv.org/abs/1909.12535
4. Cho, Y.J., Wang, J., Joshi, G.: Client selection in federated learning: Convergence analysis and power-of-choice selection strategies (2020), https://arxiv.org/abs/2010.01243
5. Duan, M., Liu, D., Chen, X., Liu, R., Tan, Y., Liang, L.: Self-balancing federated learning with global imbalanced data in mobile systems. IEEE Transactions on Parallel and Distributed Systems 32(1), 59–71 (2021)
6. Fraboni, Y., Vidal, R., Kameni, L., Lorenzi, M.: Clustered sampling: Low-variance and improved representativity for clients selection in federated learning. In: Meila, M., Zhang, T. (eds.) Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 139, pp. 3407–3416. PMLR (18–24 Jul 2021), https://proceedings.mlr.press/v139/fraboni21a.html
7. Ghosh, A., Chung, J., Yin, D., Ramchandran, K.: An efficient framework for clustered federated learning. IEEE Transactions on Information Theory 68(12), 8076–8091 (2022)
8. He, C., Annavaram, M., Avestimehr, S.: Group knowledge transfer: Federated learning of large cnns at the edge. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 14068–14080. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/a1d4c20b182ad7137ab3606f0e3fc8a4-Paper.pdf
9. Hu, Z., Shaloudegi, K., Zhang, G., Yu, Y.: Federated learning meets multi-objective optimization. IEEE Transactions on Network Science and Engineering 9(4), 2039–2051 (2022)

10. Huang, T., Lin, W., Wu, W., He, L., Li, K., Zomaya, A.Y.: An efficiency-boosting client selection scheme for federated learning with fairness guarantee. IEEE Transactions on Parallel and Distributed Systems 32(7), 1552–1564 (2021)
11. Jee Cho, Y., Wang, J., Joshi, G.: Towards understanding biased client selection in federated learning. In: Camps-Valls, G., Ruiz, F.J.R., Valera, I. (eds.) Proceedings of The 25th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 151, pp. 10351–10375. PMLR (28–30 Mar 2022), https://proceedings.mlr.press/v151/jee-cho22a.html
12. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S.J., Stich, S.U., Suresh, A.T.: Scaffold: stochastic controlled averaging for federated learning. In: Proceedings of the 37th International Conference on Machine Learning. ICML'20, JMLR.org (2020)
13. Lai, F., Zhu, X., Madhyastha, H.V., Chowdhury, M.: Oort: Efficient federated learning via guided participant selection. In: 15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21). pp. 19–35. USENIX Association (Jul 2021), https://www.usenix.org/conference/osdi21/presentation/lai
14. Li, C., Zeng, X., Zhang, M., Cao, Z.: Pyramidfl: a fine-grained client selection framework for efficient federated learning. In: Proceedings of the 28th Annual International Conference on Mobile Computing And Networking. p. 158–171. MobiCom '22, Association for Computing Machinery, New York, NY, USA (2022), https://doi.org/10.1145/3495243.3517017
15. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10708–10717 (2021)
16. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. In: Dhillon, I., Papailiopoulos, D., Sze, V. (eds.) Proceedings of Machine Learning and Systems. vol. 2, pp. 429–450 (2020), https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Paper.pdf
17. Liao, Y., Shen, X., Rao, H.: Analytic sensor rules for optimal distributed decision given k-out-of-l fusion rule under monte carlo approximation. IEEE Transactions on Automatic Control 65(12), 5488–5495 (2020)
18. Lyu, L., Yu, J., Nandakumar, K., Li, Y., Ma, X., Jin, J., Yu, H., Ng, K.S.: Towards fair and privacy-preserving federated deep models. IEEE Transactions on Parallel and Distributed Systems 31(11), 2524–2541 (2020)
19. Sattler, F., Müller, K.R., Samek, W.: Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE Transactions on Neural Networks and Learning Systems 32(8), 3710–3722 (2021)
20. Sattler, F., Wiedemann, S., Müller, K.R., Samek, W.: Robust and communication-efficient federated learning from non-i.i.d. data. IEEE Transactions on Neural Networks and Learning Systems 31(9), 3400–3413 (2020)
21. Tang, M., Ning, X., Wang, Y., Sun, J., Wang, Y., Li, H., Chen, Y.: Fedcor: Correlation-based active client selection strategy for heterogeneous federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10102–10111 (June 2022)
22. Thi Le, T.H., Tran, N.H., Tun, Y.K., Nguyen, M.N.H., Pandey, S.R., Han, Z., Hong, C.S.: An incentive mechanism for federated learning in wireless cellular networks: An auction approach. IEEE Transactions on Wireless Communications 20(8), 4874–4887 (2021)
23. Tu, L., Ouyang, X., Zhou, J., He, Y., Xing, G.: Feddl: Federated learning via dynamic layer sharing for human activity recognition. In: Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems. p. 15–28. SenSys '21, Association for Computing Machinery, New York, NY, USA (2021), https://doi.org/10.1145/3485730.3485946

24. Wang, H., Kaplan, Z., Niu, D., Li, B.: Optimizing federated learning on non-iid data with reinforcement learning. In: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. pp. 1698–1707 (2020)
25. Wang, H., Qu, Z., Guo, S., Gao, X., Li, R., Ye, B.: Intermittent pulling with local compensation for communication-efficient distributed learning. IEEE Transactions on Emerging Topics in Computing 10(2), 779–791 (2022)
26. Wang, J., Chang, X., Mišić, J., Mišić, V.B., Wang, Y.: Pass: A parameter audit-based secure and fair federated learning scheme against free-rider attack. IEEE Internet of Things Journal 11(1), 1374–1384 (Jan 2024), http://dx.doi.org/10.1109/JIOT.2023.3288936
27. Wu, Q., Chen, X., Zhou, Z., Zhang, J.: Fedhome: Cloud-edge based personalized federated learning for in-home health monitoring. IEEE Transactions on Mobile Computing 21(8), 2818–2832 (2022)
28. Xu, J., Wang, H.: Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective. IEEE Transactions on Wireless Communications 20(2), 1188–1200 (2021)
29. Xue, Z., Wang, H.: Effective density-based clustering algorithms for incomplete data. Big Data Mining and Analytics 4(3), 183–194 (2021)
30. Yang, M., Wang, X., Zhu, H., Wang, H., Qian, H.: Federated learning with class imbalance reduction. In: 2021 29th European Signal Processing Conference (EUSIPCO). pp. 2174–2178 (2021)
31. Yao, X., Sun, L.: Continual local training for better initialization of federated models. In: 2020 IEEE International Conference on Image Processing (ICIP). pp. 1736–1740 (2020)
32. Yu, S., Chen, X., Zhou, Z., Gong, X., Wu, D.: When deep reinforcement learning meets federated learning: Intelligent multitimescale resource management for multiaccess edge computing in 5g ultradense network. IEEE Internet of Things Journal 8(4), 2238–2251 (2021)
33. Zhang, J., Cheng, X., Wang, C., Wang, Y., Shi, Z., Jin, J., Song, A., Zhao, W., Wen, L., Zhang, T.: Fedada: Fast-convergent adaptive federated learning in heterogeneous mobile edge computing environment. World Wide Web 25(5), 1971–1998 (Sep 2022), https://doi.org/10.1007/s11280-021-00989-x
34. Zhang, S., He, J., Liang, W., Li, K.: Mmds: A secure and verifiable multimedia data search scheme for cloud-assisted edge computing. Future Gener. Comput. Syst. 151(C), 32–44 (Feb 2024), https://doi.org/10.1016/j.future.2023.09.023

Shiwen Zhang received his Ph.D. degree from the College of Computer Science and Electronic Engineering, Hunan University, in 2016. He is currently an associate professor at the School of Computer Science and Engineering, Hunan University of Science and Technology. He is a member of IEEE and CCF. His research interests include identity authentication, security and privacy issues in Wireless Body Area Networks (WBAN), cloud computing, privacy protection, and information security.
Email: shiwenzhang@hnust.edu.cn.

Shuang Chen received a B.S. degree from Hunan University of Science and Technology (HNUST) in 2022 and am pursuing an M.S. degree from the School of Computer Science and Engineering at Hunan University of Science and Technology. Her research interests include clustering algorithms in federated learning, edge computing, and information security.

Wei Liang received a Ph.D. degree in computer science and technology from Hunan University in 2013. He was a Post-Doctoral Scholar at Lehigh University from 2014 to

2016. He is currently a Professor and the Dean of the School of Computer Science and Engineering at Hunan University of Science and Technology, China. He has authored or co-authored more than 140 journal/conference papers. His research interests include federated learning, edge computing, and blockchain.

Kuanching Li is a Professor at the School of Computer Science and Engineering, Hunan University of Science and Technology. Dr. Li has co-authored over 150 conference and journal papers, holds several patents, and serves as an associate and guest editor for various scientific journals. He has also held chair positions at several prestigious international conferences. His research interests include cloud and edge computing, big data, and blockchain technologies. Dr. Li is a Fellow of the Institution of Engineering and Technology (IET).

Arcangelo Castiglione is an associate professor at the Department of Computer Science, University of Salerno, Italy. He received a Ph.D. degree in Computer Science from the University of Salerno, Italy. He is an Associate Editor for several high-ranked journals, involved in several renowed international conference organizational roles, a member of the IEEE TC on Secure and Dependable Measurement, and a founding member of the IEEE TEMS TC on Blockchain and Distributed Ledger Technologies. His research mainly focuses on cryptography, network security, data protection, digital watermarking, and automotive security.

Junsong Yuan is a Professor and Director of the Visual Computing Lab at the Department of Computer Science and Engineering (CSE), State University of New York at Buffalo, USA. Before joining SUNY Buffalo, he was an Associate Professor (2015-2018) and a Nanyang Assistant Professor (2009-2015) at Nanyang Technological University (NTU), Singapore. He obtained his Ph.D. from Northwestern University in 2009, M.Eng. from the National University of Singapore in 2005, and B.Eng. from Huazhong University of Science Technology (HUST) in 2002. He received the Chancellor's Award for Excellence in Scholarship and Creative Activities from SUNY, Nanyang Assistant Professorship from NTU, and Outstanding EECS Ph.D. Thesis award from Northwestern University, and Best Paper Award from IEEE Trans. on Multimedia. He serves as Senior Area Editor of Journal of Visual Communication and Image Representation (JVCI), Associate Editor of IEEE Trans. on Pattern Analysis and Machine Intelligence (T-PAMI), IEEE Trans. on Image Processing (T-IP), IEEE Trans. on Circuits and Systems for Video Technology (T-CSVT), and Machine Vision and Applications (MVA). He also serves as General/Program Co-chair of ICME and Area Chair for CVPR, ICCV, ECCV, ACM MM, etc. He was elected as a faculty senator at both SUNY Buffalo and NTU. He is a Fellow of IEEE and IAPR.

Received: January 31, 2024; Accepted: April 24, 2025.

Computer Science and Information Systems 22(3):971–989 https://doi.org/10.2298/CSIS250215053W

A spatio-temporal Graph Neural Network for EEG Emotion Recognition Based on Regional and Global Brain

Xiaoliang Wang[1,2], Chuncao Li[1,2], Yuzhen Liu[1,2],
Wei Liang[1,2], Kuanching Li[1,2,⋆], and Aneta Poniszewska-Maranda[3]

[1] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[2] Sanya Research Institute, Hunan University of Science and Technology, Sanya 572024, China
{fengwxl, wliang, aliric}@hnust.edu.cn
lichuncao2008@163.com
yzhenliu@126.com
[3] Institute of Information Technology, Lodz University of Technology, Lodz, Poland
aneta.poniszewska-maranda@p.lodz.pl

Abstract. Effective emotion recognition based on electroencephalography (EEG) is crucial for the development of Brain-Computer Interface (BCI). Neuroscientific studies highlight the importance of localized brain activity analysis for understanding emotional states. However, existing deep learning methods often fail to extract spatio-temporal features of EEG signals adequately. Accordingly, we propose a novel spatio-temporal graph neural network, MSL-TGNN, by integrating local and global brain information. A multi-scale temporal learner is designed to extract EEG temporal dependencies. And a brain region learning block and an extended global graph attention network are introduced to explore the spatial features. Specifically, the brain region learning block aggregates local channel information, whereas the extended global graph attention network can effectively capture nonlinear dependencies among regions to extract global brain information. We conducted subject-dependent and subject-independent experiments on the DEAP dataset, and the results indicate that our proposed model outperforms compared to state-of-the-art methods.

Keywords: Bidirectional gated recurrent unit, EEG emotion recognition, graph attention network, multi-dimensional attention, deep learning.

1. Introduction

Emotions reflect an individual's current psychological and physiological states, influencing various aspects of our daily lives. Accurate and efficient emotion recognition is crucial for advancing the development of BCI. Both physiological and non-physiological signals can convey a person's emotional state. Non-physiological signals include facial expressions [27], language [24], body posture [26], among others. Physiological signals

⋆ Corresponding author

encompass electroencephalography (EEG), electrooculography (EOG), electrocardiography (ECG), and the like. Compared to non-physiological signals' more easily disguised nature, physiological signals authentically represent a person's emotions. Additionally, due to its non-invasive, convenient, and cost-effective nature, EEG is widely employed in emotion recognition.

One of the challenges in EEG emotion recognition is designing a more efficient method with good adaptability and generalization for automatically extracting relevant information from EEG signals. Traditional EEG emotion recognition often relies heavily on manual feature extraction. The most commonly used features are frequency domain features. The brainwave signals are initially decomposed into five frequency bands through Fourier Transform [32, 15]. Features are then extracted from each frequency band. The Power Spectral Density (PSD) [8], the Differential Entropy (DE) [44, 20], and the Differential Asymmetry (DASM) [30] are examples of frequently used EEG features. Zheng *et al.* [42] extended traditional DE features to dynamic DE features, achieving higher accuracy in EEG emotion classification. Gao *et al.* [13] fused both frequency domain and time domain features for EEG emotion recognition, showing that feature fusion can effectively enhance recognition accuracy. However, these traditionally manually extracted features often fail to extract the temporal and spatial information from EEG signals fully. Additionally, frequency domain features are typically based on static analysis of the entire EEG signal, failing to capture the dynamic changes in the signal over time.

Deep learning has developed rapidly and has been widely used in various fields [45]. It has also been extensively utilized in EEG emotion recognition. Fourati *et al.* [12] proposed a model based on Echo State Network (ESN) and utilized filtered signals as network inputs without employing any feature extraction methods. Li *et al.* [22] utilized the Bidirectional Long Short-Term Memory (BiLSTM) to extract spatio-temporal features, introducing a collaborative working mechanism between a classifier and discriminator to help reduce the disparities in emotion recognition across domains. Tao *et al.* [33] employed a Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) combined with attention mechanisms to address EEG emotion recognition problems. Ding *et al.* [10] used multiple one-dimensional convolutions and Graph Neural Network (GNN) to explore the spatio-temporal features of EEG signals. Gu *et al.* [15] utilized Generative Adversarial Networks (GAN) in conjunction with GNN and long short-term memory (LSTM) to explore EEG emotion classification. Zhang *et al.* [40] explored deep-level information of graph-structured data by stacking multiple graph convolution layers.

Nevertheless, several problems still require research to increase the accuracy of EEG emotion classification, since EEG signals have poor spatial resolution [39], as many methods struggle to extract spatial information from different brain regions adequately. The brain is a highly organized and differentiated organ, composed of various regions responsible for different functions. Complex connections and interactions exist between these regions, and studies indicate that the strength of interaction between brain regions attenuates with increasing physical distance [29]. Understanding the local activities of the brain is crucial in neuroscience and clinical fields for exploring cognitive functions, neurological disorders, and the impact of brain injuries. In EEG emotion recognition, another challenge is how to simultaneously focus on the brain's overall structure and information from individual local regions.

To address the challenges mentioned above in EEG emotion recognition, we propose Multi-scale spatio-temporal Graph Neural Network (MSL-TGNN). Inspired by Ding *et al.* [11], we introduce a multi-scale temporal learner in MSL-TGNN, employing three parallel Bidirectional Gated Recurrent Units (BiGRU) to capture different frequency representations in EEG signals. EEG data typically contain signals of multiple frequencies, and these frequencies may play crucial roles at different time points. The superposition of different hidden layer states in BiGRU can comprehensively handle information from different frequency dimensions, aiding the model in achieving a multi-scale representation of the signals. Additionally, by introducing a brain region learning block to aggregate local channel information, the model better understands the roles played by various brain regions in EEG emotion classification. Simultaneously, we integrate the multi-dimensional "t2t" self-attention mechanism [31] into the Graph Attention Network (GAT) to capture intra-node dependencies and inter-node connectivity. GAT is effective in learning interactions between nodes, and the multi-dimensional "t2t" self-attention mechanism can learn relationships between multi-dimensional features within nodes. The extended global graph attention network comprehends relationships between nodes at a higher level and captures global patterns more effectively. The main contributions of this work include the following aspects:

- To propose a novel end-to-end deep learning framework to overcome the limitations of manually extracting features. The proposed model can automatically learn spatio-temporal features from EEG signals, demonstrating better adaptability and generalization across different subjects. It can comprehensively capture the complexity of EEG signals.
- By leveraging learned local weights, we perform a weighted fusion of information from each brain region. This enables the model to understand better each brain region's unique roles in EEG emotion recognition. Furthermore, introducing the extended global graph attention network strengthens the model's capacity to capture nonlinear dependencies between nodes and global information, as integrating local and global modules ensures that the model adequately acquires spatial information from the brain.
- Extensive subject-dependent and subject-independent experiments conducted on the DEAP dataset have demonstrated that the proposed MSL-TGNN method can significantly improve performance. Additionally, ablation studies illustrate the effectiveness of each module in the proposed method.

The remainder of this work is organized as follows. The background of the proposed model is provided briefly in Section 2, a comprehensive description of the proposed MSL-TGNN and its application in EEG emotion recognition is provided in Section 3, the experimental details and the discussions of results are presented in Section 4, and finally, the concluding remarks and future directions are depicted in Section 5.

2. Related Work

This section briefly introduces methods that serve as the foundation for the proposed model.

2.1. BiGRU

EEG data comprise multi-channel information that varies over time, and RNNs are good at learning long-term dependencies. In recent years, RNN has found widespread application in EEG data. However, RNN suffers from issues such as exploding and vanishing gradients. To overcome the shortcomings of traditional RNN, Hochreiter *et al.* [16] proposed LSTM. To simplify the structure of LSTM for improved training efficiency, Chung *et al.* [6] introduced modifications based on LSTM and presented GRU. Unidirectional GRU performs a state transition from past to future. In specific tasks, particularly those requiring simultaneous consideration of past and future information, unidirectional models may fail to utilize all available contextual information fully. However, EEG data typically contain intricate spatio-temporal information, and the signal feature may depend on both past and future time points. BiGRU can capture past and future information in sequence data by considering forward and backward hidden states simultaneously [17]. So BiGRU assists in handling the time dependencies in EEG data, facilitating a more effective capture of dynamic information at different time points. Abgeena *et al.* [1] proposed a CNN-BiGRU model, demonstrating its effectiveness in emotion classification based on EEG signals. However, it still fails to extract spatial information from the brain entirely.

2.2. GAT

GNNs are widely used in various fields [3, 23, 9, 18], and GAT is a special type of GNN. Compared to Graph Convolutional Networks (GCN), GAT possesses unique advantages. Different from GCN that employs uniform neighbor aggregation, thus disregarding the heterogeneity among nodes, GAT introduces an attention mechanism that assigns different weights to neighboring nodes for each node, allowing for a more flexible capture of the graph structure [35]. Because of this, GAT handles complicated graph structures very well and provides a more accurate representation of the relationships between nodes. Zhao *et al.* [41] offered an epilepsy detection method based on GAT and highlighted the potential advantages of GAT in handling multi-channel biological signals. Huang et al. used GAT to extract features from EEG signals and perform emotion recognition [19]. However, little study has been done on applying GAT to EEG emotion recognition.

2.3. Self-attention and Multi-dimensional Attention

The self-attention mechanism has been widely used in various natural language processing (NLP) tasks [34]. It allows models to establish weight relationships between positions in a sequence, capturing global contextual information, so the models can be suited for sequences of different lengths. The multi-dimensional attention mechanism is an extension of the attention mechanism at the feature level [31], aimed at more comprehensively capturing relationships across multiple dimensions in input data. Compared to traditional self-attention, multi-dimensional attention introduces independent attention to different feature dimensions. Each feature dimension has its weight allocation in multi-dimensional attention, allowing the model to attend to critical information in different dimensions flexibly. This mechanism is well-suited for handling multimodal data or data with multiple dimensions, such as multi-channel EEG data. The introduction of multi-dimensional attention enhances the proposed model's perception of the multi-dimensional relationships, thereby exhibiting excellent performance in handling multi-channel data.

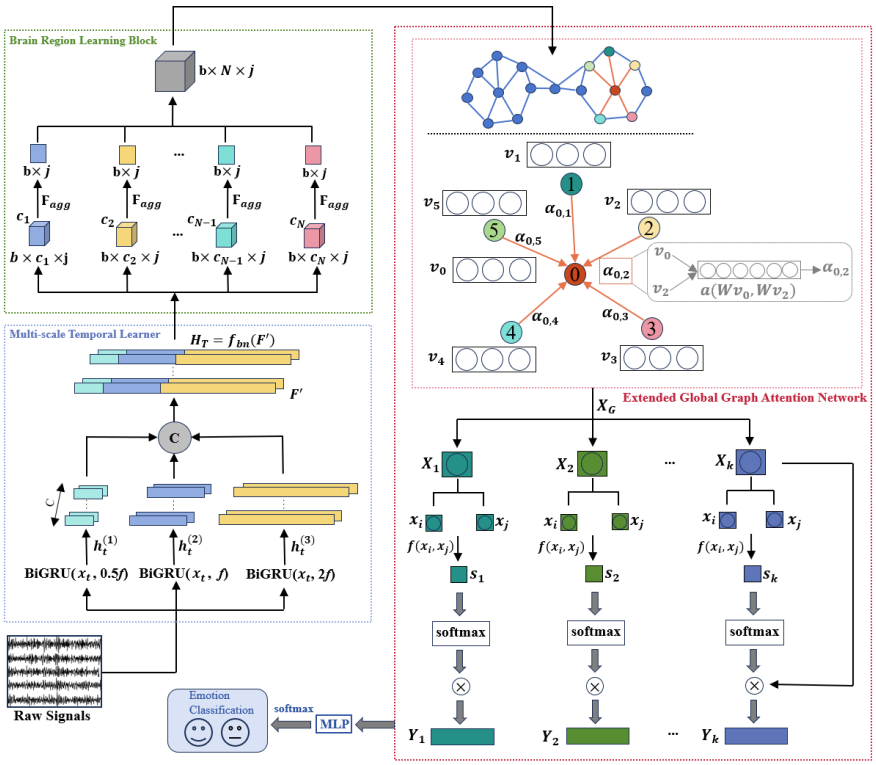


Fig. 1. The structure of MSL-TGNN. MSL-TGNN consists of the multi-scale temporal learner and the spatial feature learner, which further comprises the brain region learning block and the extended graph attention network. The multi-scale temporal learner illustrates three parallel BiGRUs learning information from different frequency dimensions of multi-channel EEG data. The brain region learning block demonstrates the information aggregation process of four brain regions. The extended global graph attention network displays the weight distribution between nodes and the weight distribution of internal multiple feature dimensions at the node level

3. Method

MSL-TGNN consists of two major modules: the multi-scale temporal learner and the spatial feature learner. The multi-scale temporal learner automatically extracts information from different frequency dimensions more comprehensively by overlaying the bidirectional hidden states of multiple dimensions, replacing manual feature extraction. The spatial feature learner includes the brain region learning block and the extended global graph attention network. The brain region learning block captures the neural activity of brain regions, and the aggregated local channel information serves as input to the extended global graph attention network to learn complex relationships among different brain regions. Fig. 1 shows the structure of the proposed MSL-TGNN.

3.1. Multi-scale Temporal Learner

The multi-scale temporal learner learns information from different frequency dimensions of EEG data by configuring parallel hidden state sizes. To comprehensively capture the temporal dynamics in the input sequence, we set the size of the hidden state in different proportions according to the sampling rate f. The ratio coefficient is denoted as $\lambda_i \in R$, where i represents the layer number of BiGRU, $i = [1, 2, 3]$. The hidden state size $h^{(i)}$ for the i-th layer can be defined as

$$h^{(i)} = \lambda_i \times f, \lambda_i \in [0.5, 1, 2] \tag{1}$$

The values of λ_i were determined through extensive experimentation. We systematically evaluated a wide range of values for λ_i, and the optimal setting was selected based on comprehensive experimental results that balanced performance and stability. Given the baseline-corrected EEG data $X_t \in R^{c \times l}$, where t is the time step of the input sequence, c is the number of channels, and l is the sample length along the time dimension. We apply three parallel multi-scale BiGRUs to learn dynamic frequency representations, and the update of the hidden state in a unidirectional GRU can be represented as

$$h_t^{(i)} = z_t^{(i)} \odot \widetilde{h_t^{(i)}} + \left(1 - z_t^{(i)}\right) \odot h_{t-1}^{(i)} \tag{2}$$

where $z_t^{(i)}$ is the output of the update gate, $\widetilde{h_t^{(i)}}$ is the output of the memory cell after activation function and $\odot$ represents the dot product operation.

The output sequence of the forward GRU in the i-th layer can be expressed as

$$\overrightarrow{h_t^{(i)}} = \overrightarrow{GRU}\left(\overrightarrow{h_{t-1}^{(i)}}, X_t\right) \tag{3}$$

And the output sequence of the backward GRU in the i-th layer can be expressed as

$$\overleftarrow{h_t^{(i)}} = \overleftarrow{GRU}\left(\overleftarrow{h_{t+1}^{(i)}}, X_t\right) \tag{4}$$

We concatenate the outputs of all parallel BiGRUs along the feature dimension. Therefore, the final output of the multi-scale temporal learner is represented as

$$H_T = f_{bn}\left(\Gamma\left(\overrightarrow{h_t^{(1)}}, \overleftarrow{h_t^{(1)}}, \ldots, \overrightarrow{h_t^{(i)}}, \overleftarrow{h_t^{(i)}}\right)\right) \tag{5}$$

where $\Gamma(\cdot)$ represents the concatenation operation along the feature dimension and f_{bn} denotes batch normalization operation.

3.2. Spatial Feature Learner

Brain region learning block. It has been demonstrated that different brain functional regions contribute variably to emotion processing [25]. For instance, the anteriofrontal and frontal regions are closely associated with cognitive and emotional regulation, while the temporal and parietooccipital regions play critical roles in processing sensory and

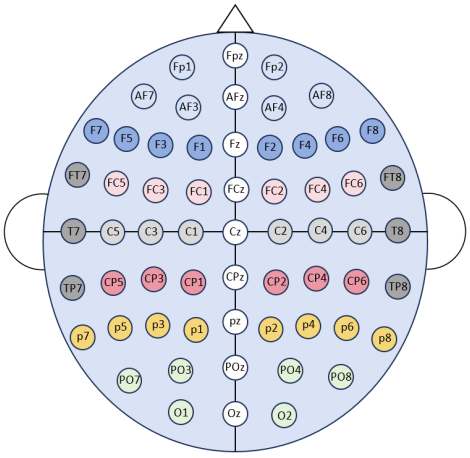


Fig. 2. Schematic diagram of EEG electrode positions. Electrodes with the same color in the same hemisphere represent a brain region. The 62 electrodes are divided into 17 regions

affective information. Dividing electrodes based on these functional regions allows for the focus on localized neural activities, thereby enabling the identification of regions that significantly contribute to emotion. By assigning higher attention weights to these key areas, the overall accuracy of emotion recognition can be enhanced. We input the output of the multi-scale temporal learner into the brain region learning block to aggregate the local information within brain regions. The placement of EEG electrodes on the scalp follows the 10-20 system [28]. Following the definition in [14], the 62 electrodes were split into 17 regions, as shown in Fig. 2. Taking the right hemisphere as an example: anteriofrontal (AF: Fp2, AF4, AF8), frontal (F: F2, F4, F6, F8), temporal (T: FT8, T8, TP8), frontocentral (FC: FC2, FC4, FC6), central (C: C2, C4, C6), centroparietal (CP: CP2, CP4, CP6), parietal (P: P2, P4, P6, P8), and parietooccipital (PO: PO4, PO8, O2).

In the process of channel information aggregation, we first introduce the local weight matrix $W_{c\times j}$, where c denotes the number of channels, and j represents the number of features for every channel. Through initialization and learning processes, this matrix can adaptively adjust the weight for each channel. By applying different weights to each channel, we can more finely capture local features of brain regions. Then, we use the aggregation function $F_{agg}(\cdot)$ to aggregate the channel information output by the multi-scale temporal learner. After extensive experimentation with aggregation operations such as average, sum, and variance, we chose the average operation. Therefore, the output H_{area} of the brain region learning block can be represented as

$$H_{area} = F_{agg}\left(W_{c\times j} \odot H_T\right) \tag{6}$$

Extended global graph attention network. As seen in the right half of Fig. 1, the extended global graph attention network aims to learn the correlations between different brain regions. We integrate the multi-dimensional " t2t " self-attention mechanism into the

GAT to focus on the multi-dimensional features within nodes. Each subject's raw data are represented as a new structured time series and serve as input to the extended global graph attention network after being processed by the multi-scale temporal learner and brain region learning block. Initially, each subject's data is treated as a cyclic-free graph. Then, a correlation matrix is generated based on the neural activity relationships between brain regions, representing the adjacency matrix of the corresponding graph. After obtaining the associations between nodes through GAT, we further enhance the expression capability of nodes' internal features using the multi-dimensional "t2t" self-attention mechanism.

Inspired by Zhao *et al.* [41] and Wang *et al.* [36], we consider the information of each subject's brain as a graph. The information aggregated from each region is regarded as a node in the graph, and the associations between each pair of regions are considered edges. Pearson correlation matrix is employed to calculate spatial correlations.

The input of GAT consists of a set of node features $V = \{v_1, v_2, \ldots, v_N\}, v_N \in R^F$, and the corresponding adjacency matrix. Here, N denotes the number of brain regions or nodes, and F denotes the feature dimensions of each node. GAT initially performs a linear transformation on each node by multiplying it with a weight matrix. Subsequently, the attention coefficients between each pair of nodes are calculated as

$$e_{ij} = a\left(Wv_i, Wv_j\right) \tag{7}$$

where $W \in R^{F \times F'}$, with F' representing the feature dimensions of the output nodes, i and j denote any two nodes and $a(\cdot)$ represents a feedforward neural network that concatenates the resulting vectors to accomplish feature mapping. Next, we compute the attention coefficients of node i to all other nodes and employ softmax to normalize the attention weights, obtaining the ultimate attention coefficients α_{ij}. The calculation formula is defined as

$$\alpha_{ij} = \frac{\exp\left(LeakyReLU\left(\overrightarrow{a}^T\left[Wv_i \,||\, Wv_j\right]\right)\right)}{\sum_{n \in N_i} \exp\left(LeakyReLU\left(\overrightarrow{a}^T\left[Wv_i \,||\, Wv_n\right]\right)\right)} \tag{8}$$

where n represents an arbitrary node, $||$ is the concatenation operator. Finally, during the convolution process, we employ the multi-head attention mechanism. After being processed by the GAT layer, the features of node i can be represented as

$$v_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i}\alpha_{ij}^k W^k v_j\right) \tag{9}$$

where k denotes an arbitrary attention head, and i represents an arbitrary node.

The attention mechanism is applied in a shared manner to all edges in the graph. Unlike GCN, where each neighboring node equally influences the representation of the target node, the proposed model assigns different attention weights to adjacent nodes, permitting the model to flexibly consider different relationships between nodes when updating node representations, aiding in capturing local information within the graph. The final output of the GAT layer is denoted as X_G.

To simultaneously focus on the multiple dimensions of features within nodes, we introduce a multi-dimensional " t2t " self-attention mechanism to assign weights to different

features of nodes. In contrast to traditional self-attention mechanisms that primarily focus on relationships between different nodes within a sequence, the multi-dimensional "t2t" self-attention mechanism can comprehensively capture information in the input sequence by considering the multi-dimensional features of each node and calculating attention scores across multiple feature dimensions. The input is split into multiple attention heads to compute attention scores, which are applied to value vectors. The outputs of each attention head are then concatenated to form the final output.

Let X_k represent the k-th sample in X_G, and s_k denote the inherent correlation between different feature dimensions x_i and x_j of X_k. The multi-dimensional "t2t" self-attention mechanism adds biases both inside and outside the activation function. Let W and b represent the weight and bias of the σ function, respectively. Thus, s_k can be expressed as

$$s_k = f\left(x_i, x_j\right) = W^T \sigma\left(W_1 x_i + W_2 x_j + b_1\right) + b \tag{10}$$

For each x_i, we calculate a probability matrix $P = \{p_1, p_2, \ldots, p_r\}$. The calculation output for x_i is defined as

$$Y_i = \sum_{j=1}^{n} p_j \odot x_j \tag{11}$$

The output of the multi-dimensional "t2t" self-attention mechanism for all samples X_G is denoted as $Y = [Y_1, Y_2, \ldots, Y_k]$.

4. Experimental Results

In this section, we first briefly introduce the dataset and the pre-processing steps. Then, we describe the experimental settings and model parameters. Subsequently, we present the results of subject-dependent and subject-independent experiments of MSL-TGNN and engage in relevant discussions.

4.1. DEAP Dataset

The DEAP dataset [21] was collected from 32 volunteers (16 males, 16 females) with ages ranging from 19 to 37 years, and 26.9 as average age. Each participant underwent 40 trials, where they watched emotionally evocative music videos lasting one minute each to induce corresponding emotional states. Simultaneously, EEG data from 32 channels and peripheral physiological signals from 8 channels of each participant were collected. After each trial, participants rated their arousal, valence, dominance, and liking for each video using a continuous 9-point scale. In this investigation, we utilized EEG data from 32 channels.

4.2. Data Preprocessing

Due to our model being an end-to-end framework, for the DEAP dataset, we only performed baseline correction on the pre-processed data provided by the authors. Baseline correction is applied to reduce errors caused by scalp potential variations, equipment drift, or other environmental interferences, aiming to obtain EEG data with a high signal-to-noise ratio [5]. The original EEG data were downsampled to 128 Hz. We selected the

stable-state EEG data before the stimulus as the baseline, corresponding to the first 3 seconds of each trial in the DEAP dataset. The average baseline value is then calculated for each channel, resulting in a baseline level per channel. Subsequently, the baseline value of each corresponding channel is subtracted from every time point of the EEG signal in that channel. This process shifts the signal as a whole to a zero baseline level. Finally, the 3-second baseline data were removed. For the label processing, we projected the continuous 9-point scale onto high and low classes for each dimension by thresholding the valence and arousal at 5. We segmented the data from each trial into non-overlapping segments of 3 seconds, further splitting each segment into three 1-second data segments.

4.3. Experiment Settings

We conducted subject-dependent and subject-independent experiments on the DEAP dataset to evaluate MSL-TGNN. In the subject-dependent experiments, after the pre-processing stage, each data sample from a subject is represented as $X_i \in R^{3\times32\times128}$, where $i = [1, 2, \ldots, 800]$. All samples from different trials were shuffled for every subject. In the subject-independent experiments, we combined all subjects' samples and shuffled them. The data samples can be represented as $X_j^{'} \in R^{3\times32\times128}$, where $j = [1, 2, \ldots, 25600]$. The experiments employed 10-fold cross-validation to evaluate the model's performance, and the average performance was taken as the final experimental results. Our model was implemented with the PyTorch framework and trained on an NVIDIA GeForce RTX 3080 Ti GPU. The GAT layer was set to 1, and the number of nodes was set to 17. We used the Adam optimizer with a learning rate 10^{-4} to update the model parameters, minimizing the cross-entropy loss function. During the training process, dropout operations randomly discarded input neurons with probability 0.5, and batch normalization was applied for each mini-batch, addressing the vanishing gradient problem, accelerating the training process, and improving model generalization.

4.4. Comparative Studies

In this subsection, we present the results of subject-dependent and subject-independent experiments to validate the effectiveness of the proposed method, followed by a brief analysis.

Subject-dependent experiments. We compared our method with five recent deep learning methods and one traditional machine learning method, including GAT [41], STFFNN [37], GCNN [32], DCNN+ConvLSTM [2], STS-Transformer [43], and Decision Tree (DT) [38]. In [41], GAT was used for epilepsy detection based on EEG data. STFFNN captures electrode dependencies using power topography maps, employs CNN for spatial feature learning, utilizes feedforward networks for temporal feature learning, and integrates spatial-temporal features using BiLSTM. GCNN is a traditional graph convolutional neural network, and we use DE features as the input to GCNN. In [2], the authors used Deep Convolutional Neural Network (DCNN) and ConvLSTM to extract features separately and then concatenated the features with attention mechanism-weighted fusion.

STS-Transformer relies on the transformer and attention mechanisms for feature extraction and weight allocation. We either directly cite their results from the literature or reproduce them based on the code they have released to guarantee an effective comparison with our method. In Table 1, we list all the features that each method used in detail.

Table 1 shows that MSL-TGNN achieves the highest accuracy of 93.09% (valence) and 93.74% (arousal). Additionally, DT outperforms GAT significantly on both valence and arousal classification tasks. We speculate that this may stem from DT utilizing DE features. Specifically, manually extracted DE features provide more stable and interpretable representations of emotional states, which are particularly beneficial for emotion classification tasks as they capture important patterns of the EEG signals. However, manual feature extraction requires significant human effort. Although GAT has advantages in modeling spatial dependencies between channels, the high dimensionality and complexity of EEG data make it difficult to effectively extract features without feature engineering techniques. However, the demand for end-to-end models is to perform EEG emotion recognition directly on raw data without manual feature extraction, which places higher demands on the model's feature extraction capabilities. Our method employs raw data as the model input. In the performance comparisons between different models presented in this paper, unless explicitly stated for a specific label dimension, the reported results are based on the average accuracy across the two label categories. Compared with GAT and DT, MSL-TGNN achieves an average accuracy improvement of approximately 20.38 and 16.35 percentage points, respectively. Although GCNN takes manually extracted DE features as input, our approach still outperforms GCNN by approximately 5.5 percentage points, indicating the effectiveness of our improvements to the GCNN. For STFFNN, DCNN+ConvLSTM, and STS-Transformer, we used the same features mentioned in the original papers as inputs. Our method outperforms these approaches by approximately 7.6, 5.65 and 5 percentage points, respectively. This significant performance improvement indicates that our method has a clear advantage in feature extraction.

Table 1. Comparison of Input Features and Performance of Different Methods on the DEAP Dataset in the Subject-dependent Experiments

Methods	Features	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
GAT(2021)	Raw signals	72.05	73.29	74.03	73.2
DT(2018)	Differential entropy	75.95	-	78.18	-
STFFNN(2022)	PSD+temporal statistics	85.42	84.33	86.16	85.5
DCNN+ConvLSTM(2021)	Raw signals	87.84	-	87.69	-
GCNN(2018)	Differential entropy	88.24	-	87.72	-
STS-Transformer(2023)	Raw signals	89.86	-	86.83	-
MSL-TGNN	Raw signals	**93.09**	**93.39**	**93.74**	**93.94**

Subject-independent experiments. In the subject-independent experiments, we compared our approach with four advanced deep learning methods: GAT, CapsNet [4], STFFNN, and STS-Transformer. In [4], the frequency domain, frequency band characteristics, and

the spatial characteristics of the EEG signals are fused and input into CapsNet for emotion classification. Table 2 displays the results of the comparison. Even when using raw data as the model input, our method performs well compared to other models. Notably, our approach's accuracy (84.14%) is comparable to STS-Transformer (84.75%), significantly outperforming GAT and CapsNet in valence classification tasks. Additionally, our method surpasses STS-Transformer and the other three methods, achieving an accuracy of 83.99% in arousal classification tasks, demonstrating the effectiveness of our method in capturing emotional states. Furthermore, our method achieves better classification accuracy despite STFFNN taking pre-processed EEG features as model input. Our method achieved better F1 scores in both valence and arousal tasks, which further confirms the robustness of our method. The experiment results indicate that by effectively capturing the spatio-temporal features of EEG data, the model can better understand the similarities and differences of EEG signals among different subjects.

Table 2. Comparison of Input Features and Performance of Different Methods on the DEAP Dataset in the Subject-independent Experiments

Methods	Features	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
GAT(2021)	Raw signals	62.88	72.19	64.62	74.27
CapsNet(2019)	Band power feature matrix	66.73	-	68.28	-
STFFNN(2022)	PSD+temporal statistics	80.17	79.97	81.28	81.09
STS-Transformer(2023)	Raw signals	**84.75**	-	82.16	-
MSL-TGNN	Raw signals	84.14	**85.83**	**83.99**	**86.05**

4.5. Ablation Study

To validate the performance of each module in our proposed method, we designed three models, namely L-TGNN, MS-TGNN, and MSL-GAT. In the first ablation experiment, aiming to emphasize the contribution of the multi-scale temporal learner, we replaced it with a single BiGRU, resulting in L-TGNN. In the second ablation experiment, we removed it from the proposed model to verify the importance of the brain region learning block, resulting in MS-TGNN. Finally, in the third ablation experiment, we replaced the original model's extended global graph attention network with a regular GAT layer to verify its effect, resulting in MSL-GAT. To control variables, we maintained consistency with the original model in data pre-processing methods. Also, we utilized the average performance from 10-fold cross-validation as the final results for all ablation experiments.

As shown in Table 3, in the subject-dependent experiments, the accuracies of MSL-GAT and L-TGNN decreased by approximately 3.55 and 3.44 percentage points, respectively, compared to MSL-TGNN. In contrast, in the subject-independent experiments, the decreases were more pronounced, reaching approximately 7.38 and 5.09 percentage points, respectively. This indicates that these models are more likely to adapt to individual-specific patterns in subject-dependent experiments, resulting in relatively minor performance differences between models. This also suggests that the extended global graph attention network contributes more to the model than the multi-scale temporal learner.

Table 3. Ablation Study on the DEAP Dataset

Experiment Schemes	Methods	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
Subject-dependent	MSL-GAT	89.07	89.58	90.67	91.01
	L-TGNN	89.34	89.95	90.61	90.97
	MS-TGNN	90.1	90.62	91.39	91.53
	MSL-TGNN	**93.09**	**93.39**	**93.74**	**93.94**
Subject-independent	MSL-GAT	76.33	78.69	77.04	80.07
	L-TGNN	78.69	80.76	79.27	81.91
	MS-TGNN	81.70	83.41	81.32	83.36
	MSL-TGNN	**84.14**	**85.83**	**83.99**	**86.05**

MS-TGNN exhibited a minor decrease in accuracy compared to MSL-GAT and L-TGNN, indicating a relatively more minor contribution from the brain region learning block than the significant contributions of the multi-scale temporal learner and extended global graph attention network. In the ablation study of subject-dependent, each of our methods was experimented on all subjects to validate the contributions of various modules in our model. Fig. 3 and Fig. 4 display the classification accuracy and standard deviation for each subject on each label task. The figures show that MSL-TGNN achieves higher accuracy on both labels than L-TGNN, MS-TGNN, and MSL-GAT, with more minor standard deviations. This indicates that each module in our model plays a unique role, resulting in better generalization and adaptability of the final model. We can observe that there is variability in the accuracy of different individuals. Since our model takes raw EEG data as input without additional feature extraction, differences in age, gender, and physical conditions are prominently reflected in our experimental results. In addition, due to individual differences, the classification accuracy of the subject-dependent experiments is higher than that of the subject-independent experiments.

4.6. Discussion

As shown in Fig. 5, MSL-TGNN achieves recognition accuracies of 94.35% for high valence and 91.52% for low valence, and 95.45% for high arousal and 91.41% for low arousal in the subject-dependent experiments. This suggests that positive emotions are more easily recognized. There are differences in the activation patterns of the brain between positive and negative emotions. Positive emotions typically involve multiple brain regions, with more pronounced interactions between these regions [7]. This makes the EEG signal features of positive emotions more enriched and easier to recognize. The confusion matrices of the subject-independent experiments also show similar results. The experiment results also show that MSL-TGNN obtains a high F1 score while achieving high accuracy. This implies that the model correctly identifies positive instances and effectively captures negative instances. In other words, for the EEG data of all subjects, the model can robustly capture features of both classes, demonstrating good robustness and generalization of the model.

In this paper, we demonstrate the effectiveness of our proposed model through extensive experiments. In comparative experiments, we contrast MSL-TGNN with six deep learning methods and one traditional machine learning method, encompassing a series of models widely applied in the processing of biosignals. While extracting effective features is crucial for EEG emotion recognition, our model achieves outstanding results even

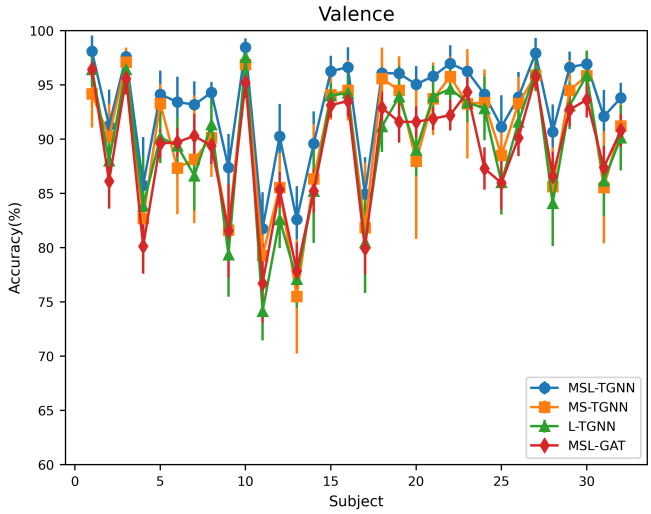


Fig. 3. Average accuracies on each subject of ablation experiments on valence classification tasks in the subject-dependent experiments

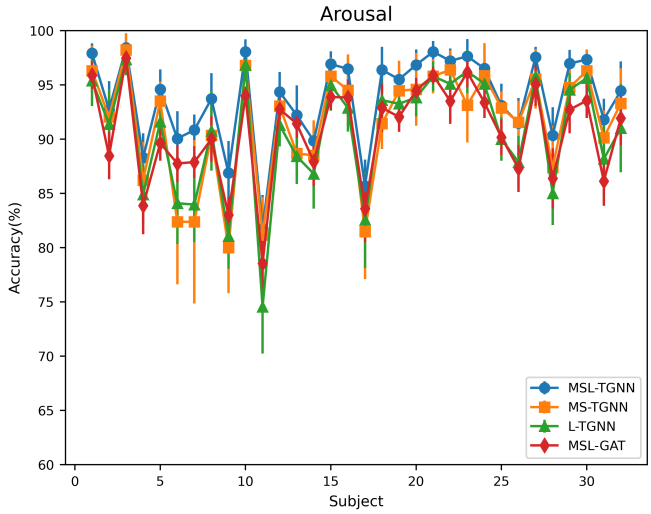


Fig. 4. Average accuracies on each subject of ablation experiments on arousal classification tasks in the subject-dependent experiments

with raw EEG data. Through ablation experiments, we conduct a thorough analysis of the performance of each module. MSL-TGNN exhibits significantly higher accuracy on valence and arousal when compared to other ablation models, confirming the unique contributions of the multi-scale temporal learner, brain region learning block, and extended global graph attention network in the entire model. GCNs adopt uniform weight assign-

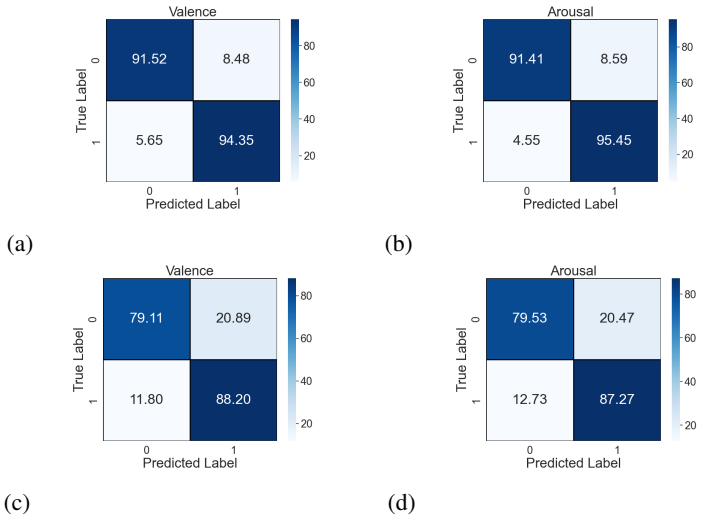


Fig. 5. Confusion matrices of MSL-TGNN (a) Subject-dependent experiments of valence (b) Subject-dependent experiments of arousal (c) Subject-independent experiments of valence (d) Subject-independent experiments of arousal

ment, neglecting functional differences across brain regions. Although GAT dynamically adjusts inter-node relationships through attention coefficients, it fails to consider the local neural activity within individual brain regions and overlooks the correlations among the multi-dimensional features within each node. In contrast, the proposed MSL-TGNN model captures the multi-scale representations of EEG signals across different frequency dimensions. It dynamically assigns weights to different functional brain regions based on their contributions to emotion, while simultaneously capturing the multi-dimensional feature representations within each region and capturing global brain patterns. This achieves a hierarchical learning framework that integrates local feature optimization with global relationship mining. This could have positive implications for future research in emotion recognition and other fields of bio-signal processing.

5. Concluding Remarks and Future Work

In this paper, we propose a novel graph neural network model that employs the multi-scale temporal learner to process different frequency dimensions in the raw EEG signals concurrently, the brain region learning block to apply different weights based on the function of each brain region, and the extended global graph attention network to capture the global patterns of brain activity. Our model can effectively capture global and local information, achieving a balanced perspective on global brain activity and detailed attention to specific regions. Moreover, MSL-TGNN is an end-to-end model capable of achieving robust recognition performance on raw EEG data. Extensive subject-dependent and subject-independent experimental results demonstrate the competitiveness of our method compared to state-of-the-art methods. Whereas our method has been proven effective in

EEG emotion recognition, there may be significant differences in emotional activity patterns among individuals. Therefore, exploring how to reduce this variability remains to be investigated. As future research directions, we will focus on enhancing the model's generalization capability, especially in cross-dataset scenarios, which may require further exploration of data augmentation, domain adaptation, and transfer learning techniques.

Acknowledgments. This work was supported by Scientific Research Fund of National Natural Science Foundation of China under Grant [62372168]; Hunan Provincial Natural Science Foundation of China under Grant [2023JJ30266]; Research Project on teaching reform in Hunan province under Grant [HNJG-2022-0791]; Hunan University of Science and Technology under Grant [2022-44-8]; and National Social Science Funds of China under Grant [19BZX044].

References

1. Abgeena, A., Garg, S.: A novel convolution bi-directional gated recurrent unit neural network for emotion recognition in multichannel electroencephalogram signals. Technology and Health Care 31(4), 1215–1234 (2023)
2. An, Y., Xu, N., Qu, Z.: Leveraging spatial-temporal convolutional features for eeg-based emotion recognition. Biomedical Signal Processing and Control 69, 102743 (2021)
3. Cai, J., Liang, W., Li, X., Li, K., Gui, Z., Khan, M.K.: Gtxchain: A secure iot smart blockchain architecture based on graph neural network. IEEE Internet of Things Journal 10(24), 21502–21514 (2023)
4. Chao, H., Dong, L., Liu, Y., Lu, B.: Emotion recognition from multiband eeg signals using capsnet. Sensors 19(9), 2212 (2019)
5. Cheng, J., Chen, M., Li, C., Liu, Y., Song, R., Liu, A., Chen, X.: Emotion recognition from multi-channel eeg via deep forest. IEEE Journal of Biomedical and Health Informatics 25(2), 453–464 (2020)
6. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
7. Cui, H., Liu, A., Zhang, X., Chen, X., Wang, K., Chen, X.: Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. Knowledge-Based Systems 205, 106243 (2020)
8. Demuru, M., La Cava, S.M., Pani, S.M., Fraschini, M.: A comparison between power spectral density and network metrics: an eeg study. Biomedical Signal Processing and Control 57, 101760 (2020)
9. Diao, C., Zhang, D., Liang, W., Li, K.C., Hong, Y., Gaudiot, J.L.: A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. IEEE Transactions on Intelligent Transportation Systems 24(1), 904–914 (2023)
10. Ding, Y., Robinson, N., Tong, C., Zeng, Q., Guan, C.: Lggnet: Learning from local-global-graph representations for brain–computer interface. IEEE Transactions on Neural Networks and Learning Systems 35(7), 9773–9786 (2024)
11. Ding, Y., Robinson, N., Zeng, Q., Chen, D., Wai, A.A.P., Lee, T.S., Guan, C.: Tsception: a deep learning framework for emotion detection using eeg. In: 2020 international joint conference on neural networks (IJCNN). pp. 1–7. IEEE (2020)
12. Fourati, R., Ammar, B., Sanchez-Medina, J., Alimi, A.M.: Unsupervised learning in reservoir computing for eeg-based emotion recognition. IEEE Transactions on Affective Computing 13(2), 972–984 (2020)
13. Gao, Q., Yang, Y., Kang, Q., Tian, Z., Song, Y.: Eeg-based emotion recognition with feature fusion networks. International journal of machine learning and cybernetics 13(2), 421–429 (2022)

14. Grabner, R.H., De Smedt, B.: Oscillatory eeg correlates of arithmetic strategies: A training study. Frontiers in psychology 3, 428 (2012)
15. Gu, Y., Zhong, X., Qu, C., Liu, C., Chen, B.: A domain generative graph network for eeg-based emotion recognition. IEEE Journal of Biomedical and Health Informatics 27(5), 2377–2386 (2023)
16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)
17. Hu, N., Zhang, D., Xie, K., Liang, W., Diao, C., Li, K.C.: Multi-range bidirectional mask graph convolution based gru networks for traffic prediction. Journal of Systems Architecture 133, 102775 (2022)
18. Hu, N., Zhang, D., Xie, K., Liang, W., Li, K., Zomaya, A.: Multi-graph fusion based graph convolutional networks for traffic prediction. Computer Communications 210, 194–204 (2023)
19. Huang, Y., Liu, T., Wu, Q.: Lg-gat: Local-global graph attention network for eeg emotion recognition. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 4355–4362. IEEE (2024)
20. Jiang, Y., Chen, N., Jin, J.: Detecting the locus of auditory attention based on the spectro-spatial-temporal analysis of eeg. Journal of Neural Engineering 19(5), 056035 (2022)
21. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing 3(1), 18–31 (2011)
22. Li, Y., Zheng, W., Wang, L., Zong, Y., Cui, Z.: From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition. IEEE Transactions on Affective Computing 13(2), 568–578 (2019)
23. Liang, W., Li, Y., Xie, K., Zhang, D., Li, K.C., Souri, A., Li, K.: Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems 24(8), 8431–8442 (2023)
24. Lindquist, K.A.: Language and emotion: Introduction to the special issue. Affective science 2(2), 91–98 (2021)
25. Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F.: The brain basis of emotion: a meta-analytic review. Behavioral and brain sciences 35(3), 121–143 (2012)
26. Mahfoudi, M.A., Meyer, A., Gaudin, T., Buendia, A., Bouakaz, S.: Emotion expression in human body posture and movement: A survey on intelligible motion factors, quantification and validation. IEEE Transactions on Affective Computing 14(4), 2697–2721 (2023)
27. Ngai, W.K., Xie, H., Zou, D., Chou, K.L.: Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. Information Fusion 77, 107–117 (2022)
28. Oostenveld, R., Praamstra, P.: The five percent electrode system for high-resolution eeg and erp measurements. Clinical neurophysiology 112(4), 713–719 (2001)
29. Perinelli, A., Tabarelli, D., Miniussi, C., Ricci, L.: Dependence of connectivity on geometric distance in brain networks. Scientific Reports 9(1), 13412 (2019)
30. Raheel, A., Majid, M., Anwar, S.M.: A study on the effects of traditional and olfaction enhanced multimedia on pleasantness classification based on brain activity analysis. Computers in biology and medicine 114, 103469 (2019)
31. Shen, T., Zhou, T., Long, G., Jiang, J., Pan, S., Zhang, C.: Disan: Directional self-attention network for rnn/cnn-free language understanding. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
32. Song, T., Zheng, W., Song, P., Cui, Z.: Eeg emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing 11(3), 532–541 (2018)
33. Tao, W., Li, C., Song, R., Cheng, J., Liu, Y., Wan, F., Chen, X.: Eeg-based emotion recognition via channel-wise attention and self attention. IEEE Transactions on Affective Computing 14(1), 382–393 (2023)

34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017)
35. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. stat 1050(20), 10–48550 (2017)
36. Wang, Y., Shi, Y., He, Z., Chen, Z., Zhou, Y.: Combining temporal and spatial attention for seizure prediction. Health Information Science and Systems 11(1), 38 (2023)
37. Wang, Z., Wang, Y., Zhang, J., Hu, C., Yin, Z., Song, Y.: Spatial-temporal feature fusion neural network for eeg-based emotion recognition. IEEE Transactions on Instrumentation and Measurement 71, 1–12 (2022)
38. Yang, Y., Wu, Q., Fu, Y., Chen, X.: Continuous convolutional neural network with 3d input for eeg-based emotion recognition. In: Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VII 25. pp. 433–443. Springer (2018)
39. Yao, N., Su, H., Li, D., Nan, J., Xia, Y., Meng, Y., Han, C., Zhu, F.: Eeg spatial projection and an improved 3d cnn with channel spatiotemporal joint attention mechanism for emotion recognition. Signal, Image and Video Processing 18(12), 9347–9362 (2024)
40. Zhang, T., Wang, X., Xu, X., Chen, C.P.: Gcb-net: Graph convolutional broad network and its application in emotion recognition. IEEE Transactions on Affective Computing 13(1), 379–388 (2019)
41. Zhao, Y., Zhang, G., Dong, C., Yuan, Q., Xu, F., Zheng, Y.: Graph attention network with focal loss for seizure detection on electroencephalography signals. International journal of neural systems 31(07), 2150027 (2021)
42. Zheng, F., Hu, B., Zheng, X., Ji, C., Bian, J., Yu, X.: Dynamic differential entropy and brain connectivity features based eeg emotion recognition. International Journal of Intelligent Systems 37(12), 12511–12533 (2022)
43. Zheng, W., Pan, B.: A spatiotemporal symmetrical transformer structure for eeg emotion recognition. Biomedical Signal Processing and Control 87, 105487 (2024)
44. Zheng, W.L., Zhu, J.Y., Peng, Y., Lu, B.L.: Eeg-based emotion classification using deep belief networks. In: 2014 IEEE international conference on multimedia and expo (ICME). pp. 1–6. IEEE (2014)
45. Zhou, S., Li, K., Chen, Y., Yang, C., Liang, W., Zomaya, A.Y.: Trustbcfl: Mitigating data bias in iot through blockchain-enabled federated learning. IEEE Internet of Things Journal 11(15), 25648–25662 (2024)

Xiaoliang Wang is a professor of information technology and chair of the Department of Internet of Things Engineering, Hunan University of Science and Technology. He leads a team of researchers and students in Information Security and the Internet of Things, such as VANET security and Anonymous Authentication in Ad Hoc Networks. He received a B.E. in computer engineering from Xiangtan University, China, and a M.S. in computer science from the joint education of Xiangtan University and the Institute of Computing Technology of the Chinese Academy of Sciences, China. He received his Ph.D. from Hunan University. He has worked at Xiangtan University and the Nanjing Government of China and has also worked as a postdoctoral researcher at the University of Alabama.

Chuncao Li is currently pursuing a Master's degree in Software Engineering at Hunan University of Science and Technology. She received her Bachelor's degree in Computer Science and Technology in 2022. Her research interests include Emotional Brain-Computer Interfaces (EBCI) and Computer Vision (CV).

Yuzhen Liu is a Lecturer at the School of Computer Science and Engineering, Hunan University of Science and Technology. He received a Ph.D. degree in computational mathematics from Xiangtan University, Xiangtan, Hunan, China, in 2012. He has authored or co-authored about 20 journal/conference papers. His research interests include network security protection and information security.

Wei Liang is currently a Professor and the Dean of the School of Computer Science and Engineering at Hunan University of Science and Technology, China. He received a Ph.D. degree in computer science and technology from Hunan University in 2013. He was a Post-Doctoral Scholar at Lehigh University from 2014 to 2016. He has authored or co-authored more than 140 journal and conference papers. His research interests include identity authentication in WBAN and security management in wireless sensor networks (WSN).

Kuanching Li is a Professor at the School of Computer Science and Engineering, Hunan University of Science and Technology. Dr. Li has co-authored over 150 conference and journal papers, holds several patents, and serves as an associate and guest editor for various scientific journals. He has also held chair positions at several prestigious international conferences. His research interests include cloud and edge computing, big data, and blockchain technologies. Dr. Li is a Fellow of the Institution of Engineering and Technology (IET).

Aneta Poniszewska-Maranda is currently an University Professor. She has been with the Institute of Information Technology, Lodz University of Technology, Poland, since 1998. She is the Head of the Software Engineering and Security of Information Systems Research Group. She has published more than 180 research papers in journals, conference proceedings, and books. Her research interests include software engineering, IS security, multi-agent systems, cloud computing, the Internet of Things, blockchain, data analysis, machine learning, LLMs, and distributed systems. She is a member of ACM, AIS, and INSTICC. She is a reviewer of more than 40 research international journals, a member of the editorial board and reviewer boards of research journals, and the chair, the vice-chair, and a PC member of many international research conferences from all over the world. She was an editor of journal special issues.

Received: February 15, 2025; Accepted: April 13, 2025.

Computer Science and Information Systems 22(3):991–1010 https://doi.org/10.2298/CSIS250312054X

Boundary-Aware Semantic Segmentation of Remote Sensing Images via Segformer and Snake Convolution

Xia Yanting, Zhang Lin, Guo Ting, Jin Qi

Geely University of China, China
zhlin002@163.com

Abstract. Semantic segmentation of remote sensing images remains challenging due to complex object structures and varying scales. This paper proposes a novel hybrid segmentation model that combines Segformer for global context extraction with Dynamic Snake Convolution to better capture fine-grained, boundary-aware features. An auxiliary semantic branch is introduced to improve feature alignment across scales. Experiments on three benchmark datasets—LoveDA, Potsdam, and Vaihingen—demonstrate that the proposed approach achieves consistent improvements in mIoU over baseline models, particularly in segmenting irregular and linear structures. This framework offers a promising solution for high-resolution land cover mapping and urban scene understanding.

Keywords: Segformer, dynamic snake convolution, remote sensing, Semantic Segmentation, Deep learning.

1. Introduction

Semantic segmentation of remote sensing images faces unique challenges due to the complex interplay of large-scale variations, mixed textures, and ambiguous boundaries. High-resolution imagery often contains slender structures (e.g., roads, rivers) and irregular objects (e.g., fragmented buildings) that are poorly captured by conventional convolutional neural network (CNN) or vision transformers (ViT). Existing methods struggle to balance global context modeling with local geometric adaptability, leading to boundary blurring and misclassification of fine-grained features. Semantic segmentation of remote sensing images is crucial for applications such as land cover mapping, environmental monitoring, and urban planning. However, the high resolution and complexity of these images[37,19,36], combined with the difficulty of annotating large-scale datasets, pose significant challenges for existing segmentation models. Accurately capturing intricate structures, such as roads, rivers[33], and urban boundaries[32], remains an unresolved problem in deep learning-based segmentation. These challenges necessitate more adaptive architectures.

With the development of deep learning, CNN-based[15] methods have become widely used for semantic segmentation. Fully Convolutional Networks (FCN) [11] replaced fully connected layers with convolutional layers, achieving end-to-end segmentation. U-Net [18,23] introduced a symmetric encoder-decoder structure with skip connections, enhancing multi-scale feature extraction, while SegNet [26] utilized unpooling layers to improve spatial detail recovery. Later, PSPNet [8] incorporated spatial pyramid pooling for better

global context understanding, and DeepLab [14] leveraged Atrous Spatial Pyramid Pooling (ASPP)[16] to integrate multi-scale features. Despite these advancements, these methods still face challenges in modeling complex interactions between pixels, often resulting in information loss. Traditional CNNs, constrained by local receptive fields, struggle to capture long-range dependencies and segment slender, irregular structures commonly found in remote sensing images. These limitations highlight the need for more adaptive and efficient segmentation approaches.

With the emergence of Transformers, semantic segmentation has further advanced. Liu et al. [35] proposed the Swin Transformer, which adopts a hierarchical structure and local attention mechanism, enabling the model to achieve higher computational efficiency while maintaining high accuracy. Zheng et al. [24] applied Transformers to semantic segmentation tasks, serializing images and feeding them into Transformers to use self-attention mechanisms at each layer for global information, thereby improving segmentation accuracy. Zheng et al. [25] proposed SETR (Semantic Segmentation Transformer Network), inspired by the ViT model, which enhances the semantic representation and generalization capabilities by introducing pixel alignment mechanisms and multi-scale attention fusion methods. However, the high complexity of Transformers results in relatively slower training speeds.

Single CNN or ViT [5] models struggle to balance local and global feature representation effectively. To address this, Zhang et al. [6] proposed a lightweight dual-branch neural network to solve intra-class heterogeneity and inter-class homogeneity problems. Jiang et al. [22] designed cross-residual feature blocks and improved skip connections to achieve dual-branch multi-scale channel cross-fusion. He et al. [30] embedded Transformers into U-Net, constructing spatial interaction modules and feature compression modules to mitigate the loss of detailed features. Wang et al. [29] proposed an algorithm based on an enhanced diffusion model. By incorporating scalable jump-connection layers into the denoising probability diffusion model, the approach effectively handles multi-scale features in campus environments, achieving superior accuracy in image semantic segmentation for autonomous driving across diverse settings. Weng et al. [13] designed the Sgformer network, incorporating multi-level feature attention to integrate the spatial details of CNNs and the contextual semantics of ViTs. Geng et al. [10] proposed DPFANet, which constructs edge optimization blocks to constrain edge features and effectively model images from local to global features. Despite advancements in CNN- and Transformer-based models, existing approaches struggle to effectively capture the elongated and irregular structures common in remote sensing images, such as roads, rivers, and building outlines. These limitations arise due to inadequate local feature representation and inefficient shape adaptation in conventional convolutional layers. While hybrid CNN-Transformer architectures improve multi-scale feature fusion, they lack specialized mechanisms to handle elongated or tortuous structures. For instance, deformable convolutions adaptively adjust receptive fields but may diverge from target boundaries in linear features. Similarly, attention mechanisms enhance global dependencies but neglect local geometric priors. To address these gaps, we propose a model that integrates Segformer's hierarchical attention with Dynamic Snake Convolution, which explicitly constrains deformable offsets to follow linear structures.

- Utilizing dynamic snake convolution to adaptively focus on slender, tortuous local structures and complex, variable global shapes, accurately capturing the tubular features in remote sensing data.
- Incorporating an auxiliary semantic branch to extract contextual information from images, ensuring the extraction of rich semantic features while maintaining inference efficiency.
- Conducting experimental analyses on publicly available datasets LoveDA and Potsdam. The experimental results demonstrate that the proposed model achieves superior segmentation performance, with mIoU reaching 52.49% on the LoveDA dataset, 79.71% on the Potsdam dataset and 76.70% on the Vaihingen dataset , representing improvements of 4.18%, 0.74% and 2.09%, respectively, over the baseline model U-Net.

The remainder of this paper is structured as follows. Section 2 presents the proposed methodology, detailing the integration of Segformer and Dynamic Snake Convolution. Section 3 describes the datasets and experimental setup, while Section 4 discusses the results and comparative analysis with baseline models. Finally, Conclusion concludes the study with key findings and future research directions.

2. Related Work

2.1. Semantic Segmentation

Semantic segmentation, a fundamental task in computer vision, aims to assign pixel-level labels to images. Early approaches relied on handcrafted features and traditional machine learning methods, but the advent of deep learning revolutionized the field. Long et al. [11] proposed Fully Convolutional Networks (FCN), replacing dense layers with convolutional layers to enable end-to-end segmentation. Building on this, U-Net [18] introduced an encoder-decoder architecture with skip connections, enhancing feature fusion across scales. Subsequent works, such as SegNet [26], improved spatial resolution recovery through unpooling layers, while PSPNet [8] and DeepLab [14] leveraged pyramid pooling and atrous convolutions to capture multi-scale context.

Despite these advancements, CNN-based methods struggled with long-range dependencies and irregular structures due to their localized receptive fields. The emergence of Vision Transformers (ViTs) addressed this limitation by modeling global interactions via self-attention. Zheng et al. [25] proposed SETR, which treats segmentation as a sequence-to-sequence problem using pure Transformers. Swin Transformer [35] further enhanced efficiency by introducing hierarchical shifted windows, balancing local and global feature extraction. Hybrid architectures, such as Swin-UNet [22], combined Transformers with U-Net to preserve spatial details while capturing global context. However, challenges persist in segmenting slender, tortuous structures (e.g., roads, rivers) in remote sensing imagery, necessitating specialized geometric modeling techniques like Dynamic Snake Convolution [21].

Recent advancements in semantic segmentation have also focused on improving the efficiency and scalability of models. Lightweight architectures, such as those proposed by Zhang et al. [6], aim to reduce computational complexity while maintaining high accuracy. These models often employ techniques like depthwise separable convolutions and

channel attention to optimize performance. Additionally, self-supervised learning methods have gained traction, reducing the dependency on large annotated datasets by leveraging unlabeled data for pre-training [7].

Another significant development is the integration of multi-task learning, where models are trained to perform multiple related tasks simultaneously, such as segmentation and object detection. This approach has been shown to improve generalization and robustness, particularly in complex scenes with diverse objects and backgrounds [28]. Furthermore, the use of generative adversarial networks (GANs) for data augmentation has proven effective in enhancing model performance, especially in scenarios with limited labeled data [20].

2.2. Attention Mechanism

Attention mechanisms have become pivotal in enhancing segmentation models by dynamically focusing on salient regions. Early efforts integrated channel-wise attention, as seen in Squeeze-and-Excitation Networks (SENet) [27], to recalibrate feature responses. Later, Non-local Networks [16] introduced self-attention to model long-range dependencies, improving contextual understanding. Transformers [35,24,25] further popularized attention by replacing convolutional operations with multi-head self-attention layers, enabling global feature interactions.

In semantic segmentation, attention mechanisms are often applied hierarchically. For instance, DeepLabv3+ [14] combined atrous spatial pyramid pooling (ASPP) with attention to refine multi-scale features. Similarly, Swin Transformer [35] employed shifted window-based attention to reduce computational complexity while maintaining global modeling capabilities. Recent works, such as DPFANet [13], integrated edge-aware attention to enhance boundary detection in remote sensing images. These mechanisms address challenges like intra-class heterogeneity and inter-class homogeneity, particularly in complex scenes. Dynamic Snake Convolution [21], with its iterative attention to linear structures, exemplifies how geometric-prior-guided attention can improve segmentation of tubular features in remote sensing data.

Attention mechanisms have also been extended to incorporate spatial and temporal dimensions, particularly in video segmentation tasks. Spatial attention focuses on relevant regions within a single frame, while temporal attention captures dependencies across multiple frames. This dual attention approach has been shown to improve the segmentation of dynamic scenes, such as those encountered in video surveillance and autonomous driving [31].

Moreover, the integration of attention mechanisms with graph neural networks (GNNs) has opened new avenues for semantic segmentation. GNNs model relationships between pixels or regions as a graph, allowing for more flexible and context-aware feature extraction. When combined with attention mechanisms, GNNs can effectively capture both local and global dependencies, leading to improved segmentation accuracy in complex scenes [3].

2.3. Remote Sensing Image Segmentation

Remote sensing image segmentation presents unique challenges due to the high resolution, complex structures, and diverse land cover types. Traditional methods often rely on

handcrafted features and machine learning algorithms, which struggle to capture the intricate details and variability in remote sensing data. With the advent of deep learning, convolutional neural networks (CNNs) have become the dominant approach, offering significant improvements in accuracy and robustness.

One of the key challenges in remote sensing image segmentation is the effective handling of multi-scale features. High-resolution images often contain objects of varying sizes, from small buildings to large agricultural fields. Multi-scale feature extraction techniques, such as pyramid pooling and atrous convolutions, have been widely adopted to address this issue. Additionally, the integration of attention mechanisms has proven effective in focusing on relevant regions and improving the segmentation of complex structures.

Another critical aspect is the ability to model long-range dependencies, which is essential for accurately segmenting large and irregular objects like rivers and roads. Vision Transformers (ViTs) have emerged as a powerful tool for capturing global context, leveraging self-attention mechanisms to model interactions between distant pixels. Hybrid models that combine CNNs and Transformers have shown promise in balancing local detail extraction with global context understanding.

Despite these advancements, segmenting slender and irregular structures remains a significant challenge. Traditional convolutional layers, with their fixed receptive fields, often fail to capture the geometric intricacies of such structures. Dynamic Snake Convolution offers a novel solution by adaptively focusing on linear and curved features, enhancing the segmentation of roads, rivers, and other elongated objects in remote sensing imagery.

3. Method

This paper proposes an efficient semantic segmentation method for remote sensing images by integrating Segformer with snake convolution. As illustrated in Figure 1, the proposed framework comprises three components: (1) a Segformer branch for multi-scale global context extraction, (2) a Dynamic Snake Convolution branch for boundary-aware feature refinement, and (3) an auxiliary semantic alignment module to harmonize cross-scale features. SegFormer is a simple, efficient, and powerful semantic segmentation framework. By combining a Transformer with a lightweight multi-layer perceptron decoder, SegFormer is capable of extracting high-resolution coarse features and low-resolution fine features, aggregating multi-scale information across different layers. Through a combination of local and global attention, it generates strong feature representations and extracts effective contextual information. The Dynamic Snake Convolution [10] enhances geometric structure perception by adaptively focusing on small and curved local features of tubular structures, thereby specifically improving the perception of such structures. To address the challenges posed by complex and variable global shapes, a multi-view feature fusion strategy is employed. The proposed method uses Dynamic Snake Convolution as the main branch structure, while SegFormer serves as an auxiliary branch for training. By employing a semantic alignment model to integrate the additional branch, the network extracts rich semantic information, achieving precise and efficient segmentation simultaneously.

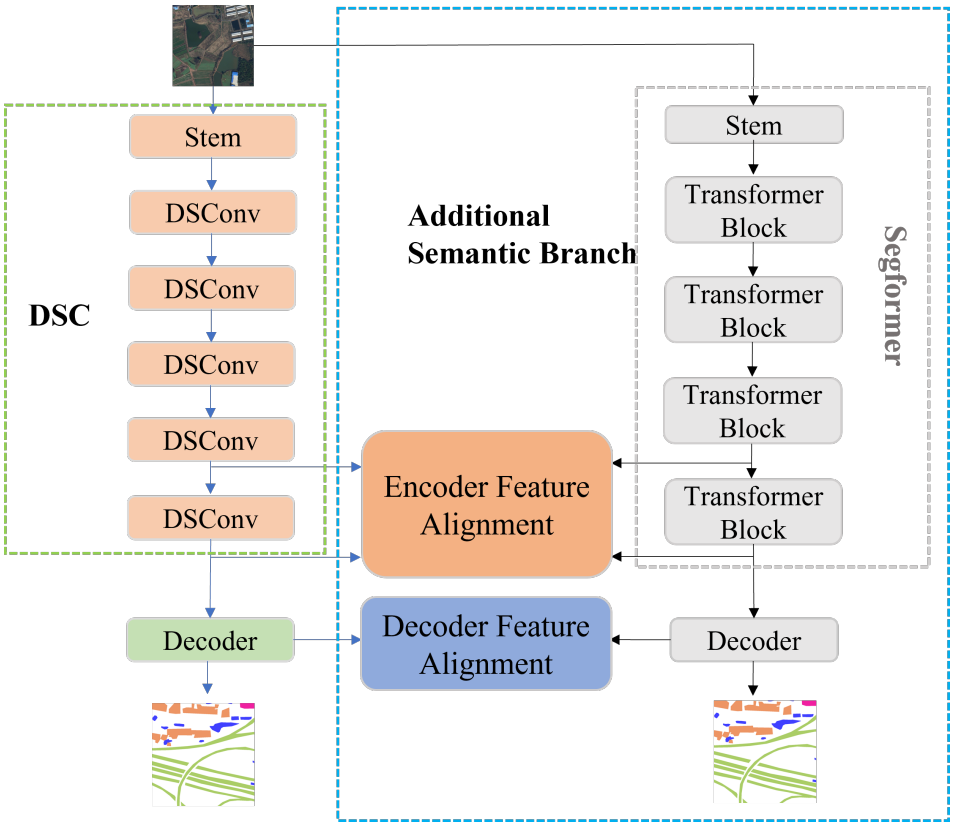


Fig. 1. The Image Segmentation Framework

3.1. Dynamic Snake Convolution

Given the standard 2D convolution coordinates K, with the center coordinate as $K_i = (x_i, y_i)$, a 3×3 convolution kernel K can be represented as:

$$K = \{(x-1, y-1), (x-1, y), \ldots, (x+1, y+1)\} \tag{1}$$

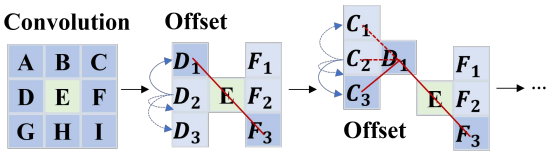


Fig. 2. Learning deformation

To provide the convolution kernel with greater flexibility and enable it to focus on the complex geometric features of targets, deformable offsets Δ[9] are introduced. However, if the model is given complete freedom to Learning deformable offsets (Figure 2), the receptive field often deviates from the target, particularly when handling slender tubular structures. Therefore, an iterative strategy is adopted (Figure 3), which sequentially selects the next position of the target to be processed for observation. This ensures continuity of focus and prevents the receptive field from spreading too far due to large deformable offsets.

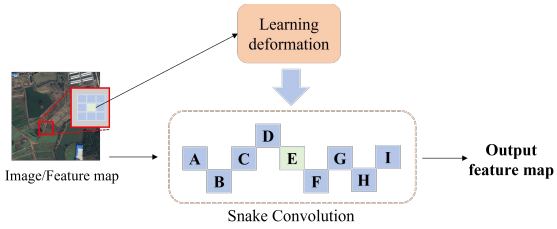


Fig. 3. Dynamic Snake Convolution

In dynamic snake convolution, the standard convolution kernel is linearized along both the x-axis and y-axis. Considering a convolution kernel of size 9, take the x-axis direction as an example. The specific position of each grid in K is represented as: $K_{i\pm c} = (x_{i\pm c}, y_{i\pm c})$,where $c = 0, 1, 2, 3, 4$ indicates the horizontal distance from the center grid. The selection of each grid position $K_{i\pm c}$ in the convolution kernel is a cumulative process. Starting from the center position K_i the position of grids farther from the center depends on the position of the preceding grid:

$K_{i+1}$is determined by adding an offset $\Delta = \{\delta \mid \delta \in [-1, 1]\}$ relative to K_i .Therefore, the offsets are accumulated $\sum$, ensuring that the kernel conforms to a linear structural pattern. The changes along the x-axis direction are illustrated in Figure 3 as:

$$K_{i\pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = \left(x_i + c, y_i + \sum_i^{i+c} \Delta y\right) \\ (x_{i-c}, y_{i-c}) = \left(x_i - c, y_i + \sum_{i-c}^{i} \Delta y\right) \end{cases} \tag{2}$$

The changes along the y-axis direction are:

$$K_{j\pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = \left(x_j + \sum_j^{j+c} \Delta x, y_j + c\right) \\ (x_{j-c}, y_{j-c}) = \left(x_i + \sum_{j-c}^{j} \Delta x, y_j - c\right) \end{cases} \tag{3}$$

Since the offset Δ is typically a decimal, while coordinates are usually in integer form, bilinear interpolation is adopted, represented as:

$$K = \sum_{K'} B\left(K', K\right) \cdot K' \tag{4}$$

Here, K represents the decimal position in Equations 2 and 3, K' enumerates all integer spatial positions, and B is the bilinear interpolation kernel, which can be decomposed into two one-dimensional kernels, as:

$$B\left(K, K'\right) = b\left(K_x, K'_x\right) \cdot b\left(K_y, K'_y\right) \tag{5}$$

Algorithm 1 Dynamic Snake Convolution

Input: Feature map F, initial kernel center $K_i = (x_i, y_i)$, kernel size $S = 9$.
Process:

for each offset step c along x/y-axis **do**
  a. Learn deformable offsets $\Delta x, \Delta y$ via a lightweight network.
  b. Accumulate offsets iteratively:
    x-direction: $x_{i+c} = x_i + c, \quad y_{i+c} = y_i + \sum \Delta y.$
    y-direction: $x_{j+c} = x_j + \sum \Delta x, \quad y_{j+c} = y_j + c.$
  c. Compute interpolated features F_{interp} using bilinear sampling.
end for
Aggregate features from all offset positions.
Output: Refined feature map F_{out}.

As shown in Algorithm 1, the Dynamic Snake Convolution algorithm is presented.The dynamic snake convolution kernel is designed to better adapt to slender tubular structures based on dynamic configurations, enabling enhanced perception of key features.

3.2. Integrating Method

As shown in Figure 3, we propose a simple yet effective alignment module for feature learning during training. It can be divided into encoder feature alignment and decoder feature alignment.

- Encoder Feature Alignment

Backbone feature alignment begins by downsampling or upsampling the features of the Transformer and CNN branches for alignment. To avoid direct feature alignment disrupting the supervision of the CNN by the ground truth during training, feature projection is employed. Specifically, the CNN features are projected to the dimension of the Transformer features. This projection unifies the number of channels and prevents direct feature alignment. Finally, semantic alignment loss is applied to the projected features to align the semantic representations [21].

- Decoder Feature Alignment

Features from stages 2 and 4 are selected for alignment. Considering the significant differences in the decoding space between the Transformer network and the backbone network, directly aligning decoding features and output logits only leads to limited improvement. Therefore, we adopt a shared decoder head alignment approach. Specifically, the features from stages 2 and 4 of the single-branch CNN are fed into a point convolution to expand their dimensions. The high-dimensional features are then passed through the Transformer decoder. The new output features and logits of the Transformer decoder are used to compute alignment loss with the original outputs of the Transformer decoder.

3.3. The Alignment Loss

To better align semantic information, an alignment loss [31] focusing on semantic information rather than spatial information is required. In this implementation, we use MGD Loss (channel-wise distillation loss) [34] as the alignment loss, which demonstrates better performance compared to other loss functions. MGD Loss consists of two components: a global distribution alignment term and a boundary constraint term.

$$\ell_{\text{align}} = \left\| E_{x\sim P[\phi(x)]} - E_{y\sim Q[\phi(v)]} \right\|^2 \tag{6}$$

Global Distribution Alignment Term:The alignment goal is achieved by measuring the difference between the feature distributions of the two modalities. In this paper, Maximum Mean Discrepancy (MMD) [3] is used as the global distribution alignment term, which is expressed as:

$$l_{\text{margin}} = \max\left(0, m + d\left(f_a(a), f_b\left(b^-\right)\right) - d\left(f_a(a), f_b\left(b^+\right)\right)\right) \tag{7}$$

Here, \(m \) represents the margin value. $f_a(a)$ and $f_b(b^-)$ are the features of modalities a and b, respectively. b^+ denotes positive samples, b^- denotes negative samples, and $d(\bullet)$ is the distance metric (e.g., Euclidean distance).

The total MGD Loss conbines the global alignment loss and margin-based loss:

$$\ell_{MGD} = \lambda \ell_{\text{align}} + (1-\lambda)\ell_{\text{margin}} \tag{8}$$

λ is a hyperparameter that balances the trade-off between alignment and discrimination. By jointly optimizing these components, MGD Loss achieves alignment of multimodal data distributions and ensures semantic consistency.

4. Experiments and Results

4.1. Dataset

To verify the proposed remote sensing image segmentation method, we used three public HR remote sensing image [4] datasets, LoveDA [12] , Potsdam and Vaihingen. The LoveDA dataset is designed to facilitate research on event detection tasks in remote sensing imagery, such as natural disaster monitoring, urban planning, etc. The LoveDA dataset provides high-resolution airborne remote sensing imagery covering a wide range of scenes

and environments. Each image is labeled with rich event categories, including natural disasters, traffic accidents, buildings, etc. In addition, a large number of remote sensing images of real scenes are also included, so that the model can be better generalized in real environments. The LoveDA dataset consists of 5987 high-resolution non-interlaced optical remote sensing images of *Nanjing*, *Changzhou*, and *Wuhan* with $166,768$ labeled objects, and the size of each pair of images is 1024×1024 pixels with a pixel separation rate of 0.3 meters, and all of the information was obtained from the Google Earth platform. Earth's platform. The dataset contains seven categories, including background, buildings, roads, water bodies, debris, forests, and agriculture, covering rural and urban areas, respectively. In these datasets, there are 2713 urban landscapes and 3274 rural landscapes. Potsdam Remote Sensing Dataset is a high-resolution airborne remote sensing image dataset provided by the University of Potsdam, Germany, which is designed to support research on remote sensing image processing tasks such as feature classification, target detection and semantic segmentation. Potsdam [2] contains 28 images of the same size, with a spatial resolution of 5 cm for the top image and DSM, and the size of each pair of images is 6000×6000 pixels. The dataset contains 38 images,we split them into 26 training, 4 validation, and 8 test images.The dataset contains 6 categories of impervious surfaces, buildings, low vegetation, trees, cars and background. Each image is labeled with precise feature classes and bounding boxes including buildings, roads, trees, etc. In addition, the Potsdam dataset provides multispectral imagery in multiple bands (e.g., red, green, blue, and near-infrared), as well as high-resolution panchromatic imagery, providing researchers with a rich data resource.The Vaihingen dataset is another widely used benchmark for high-resolution remote sensing image segmentation, provided by the International Society for Photogrammetry and Remote Sensing (ISPRS). This dataset consists of 33 aerial orthoimagery tiles covering urban and suburban areas of Vaihingen, Germany. Each image has a spatial resolution of 9 cm and a size of 2000×2000 pixels, captured in three spectral bands (near-infrared, red, and green). The dataset includes six semantic categories: impervious surfaces, buildings, low vegetation, trees, cars, and clutter/background. Additionally, it provides digital surface models (DSM) to enhance 3D feature analysis. The Vaihingen dataset is particularly challenging due to its fine-grained details, dense object distribution, and complex urban layouts, making it suitable for evaluating models' capability in handling intricate scenes. Following common practices, we utilize 16 images for training and 17 for testing, ensuring compatibility with existing research benchmarks[17].

4.2. Experimental Parameter Settings

In this experiment, the software configurations are *Ubuntu18.04 LTS* operating systems, *Python3.8* development language, and *Pytorch* Deep learning framework; the hardware configurations are one *NVIDIA RTX 3090 GPU*. Besides, and training hyperparameters are summarized in Table1.

4.3. Evaluation Metrics

To evaluate the performance of the algorithm, we employ Intersection over Union (IoU)[1], mean Intersection over Union (mIoU), F1-Score, and Overall Accuracy (OA) as evaluation metrics. IoU is defined as the ratio of the intersection to the union of the algorithm's predicted segmentation and the ground truth segmentation. mIoU is the average

Table 1. Experimental Parameter Settings

Parameters	Value
Batch Size	4
Initial Learning Rate	0.0001
Optimizer	AdamW
Iterations	500

IoU across all segmentation classes. The F1-Score is the harmonic mean of precision and recall. OA is the ratio of the number of correctly classified pixels to the total number of pixels. The specific expressions are as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

$$mIoU = \frac{1}{C}\sum_{C}^{i=1} IoU \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{13}$$

$$OA = \frac{\sum_{C}^{i=1} TP_i}{\sum_{C}^{i=1} (TP_i + FP_i + FN_i)} \tag{14}$$

where C denotes the number of segmentation classes. True Positive (TP) represents the number of pixels that are actually positive and predicted as positive. FalsePositive(FP) represents the number of pixels that are actually negative but predicted as positive. True Negative (TN) represents the number of pixels that are actually negative and predicted as negative. False Negative(FN) represents the number of pixels that are actually positive but predicted as negative.

4.4. Performance Analysis

To validate the effectiveness of the proposed algorithm, we conducted extensive experiments on three widely used remote sensing datasets, LoveDA and Potsdam, as well as the challenging Vaihingen benchmark. We compared our method with several state-of-the-art segmentation models, including FCN, U-Net, U-Net++, FPN, PSPNet, and DeepLabV3. The evaluation metrics used were IoU, mIoU, F1-Score, and Overall OA. The results are presented in Tables2, Tables3 and Tables4, and the segmentation outputs are visualized in Figures4, Figures5 and Figures6. On the LoveDA dataset, the proposed method

achieved significant improvements across all categories, with an mIoU of 52.49%, representing a 4.18% improvement over the best-performing baseline model, PSPNet. Notably, the IoU for roads and agriculture reached 57.33% (+3.90% over DeepLabV3) and 66.41% (+7.73% over U-Net++), respectively, demonstrating the efficacy of Dynamic Snake Convolution in capturing slender and irregular structures.On the Potsdam dataset, our method achieved an mIoU of 79.71%, a 0.74% improvement over DeepLabV3. Significant gains were observed for buildings (IoU: 94.01%, +0.69% over U-Net++) and trees (IoU: 80.67%, +1.28% over PSPNet), validating its ability to handle complex urban layouts and dense vegetation.On the Vaihingen Dataset,On this fine-grained urban benchmark, the proposed method achieved an mIoU of 76.70%, surpassing all baseline models. The IoU for buildings reached 93.01%, outperforming U-Net (92.62%) and PSPNet (92.78%). Notably, the model excelled in segmenting "low vegetation" (IoU: 79.21%, +0.22% over U-Net-AFS) and "car" (IoU: 81.08%, +1.49% over DeepLabV3), highlighting its robustness in distinguishing small, dense objects from cluttered backgrounds. While the IoU for "tree" (87.60%) slightly trailed PSPNet (88.79%), the overall mIoU improvement underscores the balanced performance of our approach. The integration of Segformer's multi-scale contextual modeling and Dynamic Snake Convolution's adaptive geometric perception enables precise segmentation of both large-scale structures (e.g., buildings) and fine-grained urban features (e.g., vehicles), even in highly complex scenes. The consistent superiority across all datasets stems from the synergistic design: Segformer captures global contextual semantics through hierarchical attention, while Dynamic Snake Convolution enhances local feature extraction for linear and irregular structures. The auxiliary semantic branch further aligns multi-scale features, mitigating misclassifications caused by intra-class heterogeneity and inter-class similarity.To comprehensively evaluate the model's practicality, we further compare computational complexity and inference speed across baseline methods. As shown in Tables5, the proposed method achieves a favorable balance between accuracy and efficiency.

Table 2. Comparison of Segmentation Results of Five Algorithms on the LoveDA Dataset

Methods	Backbone	IoU(%)							mIoU(%)
		background	building	road	water	barren	forest	agriculture	
FCN	VGG16	42.60	49.51	48.05	73.09	11.84	43.49	58.30	46.69
Unet	ResNet50	42.97	50.88	52.02	74.36	10.40	44.21	58.53	47.62
Unet++	ResNet50	43.06	52.74	52.78	73.08	10.33	43.05	59.87	47.84
FPN	ResNet50	42.85	52.58	52.82	74.51	11.42	44.42	58.80	48.20
PSPNet	ResNet50	42.93	51.53	53.43	74.67	11.21	44.62	58.68	48.15
DeepLabV3	ResNet50	44.40	52.13	53.52	76.50	9.73	44.07	57.85	48.31
Ours	ResNet50	**48.80**	**57.25**	**57.33**	**77.01**	**14.59**	**46.01**	**66.41**	**51.49**

4.5. Ablation Experiments

To validate the contributions of key components in the proposed framework, we conduct ablation studies on three benchmark datasets: LoveDA, Potsdam, and Vaihingen. As shown in Table6,the baseline (Segformer only) achieves mIoU scores of 49.92%, 76.24%,

Table 3. Comparison of Segmentation Results of Seven Algorithms on the PotsDam Dataset

Methods	Backbone	IoU(%)						mIoU(%)
		imp_sur	building	low vegetation	tree	car	clutter	
FCN	VGG16	85.79	93.14	77.05	77.12	90.55	40.44	77.34
Unet	ResNet50	86.82	93.62	77.21	79.07	90.89	40.32	77.99
Unet++	ResNet50	87.02	93.81	77.43	79.28	91.03	40.75	78.22
FPN	ResNet50	87.09	93.70	77.51	79.67	**92.61**	41.71	78.72
PSPNet	ResNet50	87.41	92.89	76.85	79.95	91.95	42.51	78.59
DeepLabV3	ResNet50	87.01	93.32	77.64	79.39	92.12	44.32	78.97
Ours	ResNet50	**88.07**	**94.01**	**78.25**	**80.67**	91.73	**45.51**	**79.71**

Table 4. Comparison of Segmentation Results of Seven Algorithms on the Vaihingen Dataset

Methods	Backbone	IoU(%)					mIoU(%)
		imp_sur	building	low vegetation	tree	car	
Unet	ResNet50	88.51	92.62	77.97	87.49	78.81	74.56
Unet-AFS	ResNet50	**88.93**	92.71	78.99	87.82	79.59	75.32
PSPNet	ResNet50	88.80	92.78	77.92	**88.79**	78.91	74.51
DeepLabV3	ResNet50	88.38	92.75	78.45	87.69	76.80	74.61
Ours	ResNet50	88.31	**93.01**	**79.21**	87.60	**81.08**	**76.70**

and 73.05%, respectively. Adding a CNN branch improves results slightly (e.g., 75.21% on Vaihingen), but integrating DSC delivers the largest gains, reaching 51.49%, 79.71%, and 76.70%—outperforming the baseline by up to +3.65%.

4.6. Segmentation Results

To provide a qualitative assessment, we compared the segmentation results of our algorithm with baseline models on LoveDA, Potsdam, and Vaihingen datasets, as visualized in Figures4, Figures5 and Figures6. Figure 4 shows the segmentation results for three sample images from the LoveDA dataset. The proposed method demonstrates superior performance in complex scenarios, particularly in areas with intricate structures such as roads and agricultural fields. Our method accurately segments the road network, even in areas where the roads are partially obscured by vegetation or shadows. In contrast, FCN and U-Net struggle to maintain continuity in the road segments, leading to fragmented outputs.The proposed method effectively distinguishes between agricultural fields and other land cover types, producing clean and well-defined boundaries. PSPNet and DeepLabV3, on the other hand, tend to misclassify parts of the agricultural fields as background or other categories.Our method accurately segments buildings, even in densely built-up areas. U-Net++ and FPN, while performing well in most areas, occasionally misclassify small buildings or fail to capture the exact boundaries.Figure 5 shows the segmentation results for two sample images from the Potsdam dataset. The proposed method demonstrates clear advantages in urban environments, particularly in areas with complex building layouts and dense vegetation.Our method accurately segments buildings, even in areas with overlapping structures and complex shapes. FCN and U-Net struggle to

Table 5. Comparison of number of Parameters and number of FLOPs

Model	Million Parameters	Million FLOPs
FCN	134	210
U-Net	6.4	15.41
PSPNet	32.81	79.01
DeepLabV3	35.7	83.96
Ours	27.06	48.01

Table 6. Ablation Study of Key Components (mIoU on LoveDA/Potsdam/Vaihingen)

Configuration	LoveDA(%)	Potsdam(%)	Vaihingen(%)
Segformer only	49.92	76.24	73.05
Segformer + CNN	50.07	77.20	75.21
Segformer + DSC	**51.49**	**79.71**	**76.70**

maintain the integrity of building boundaries, leading to incomplete or fragmented segments.The proposed method effectively distinguishes between trees and low vegetation, producing clean and well-defined boundaries. PSPNet and DeepLabV3, while performing well in most areas, occasionally misclassify parts of the tree canopy as low vegetation or background.While the proposed method shows a slight decrease in IoU for the car category, it still produces accurate segmentation results, particularly in areas with high car density. FPN, which performs well in this category, occasionally misclassifies cars as background or other objects. Figure 6 shows the segmentation results for three sample images from the Vaihingen dataset.In complex urban scenes, our method demonstrates exceptional performance. For example, vehicles in high-density parking lots are segmented cleanly (IoU: 81.08%), with minimal confusion with background clutter. Buildings retain sharp outlines despite intricate architectural details, outperforming U-Net-AFS and DeepLabV3 in preserving structural integrity . While PSPNet achieves marginally higher IoU for trees (88.79%), our method avoids over-segmentation errors in dense canopies, producing coherent boundaries. Additionally, the model effectively distinguishes ”low vegetation” from impervious surfaces, a critical challenge in urban planning. The qualitative results further validate the effectiveness of the proposed method in handling complex and diverse remote sensing scenes. The integration of Segformer and Dynamic Snake Convolution allows the model to capture both global context and local geometric details, leading to more accurate and consistent segmentation results. The auxiliary semantic branch ensures that the model maintains high accuracy across diverse land cover types, even in challenging scenarios with significant scale variations and unclear boundaries.In summary, the proposed method demonstrates superior performance in both quantitative and qualitative evaluations, making it a promising approach for precise, large-scale segmentation of remote sensing images. The improvements in mIoU and IoU scores, combined with the visual quality of the segmentation results, highlight the model's ability to handle complex structures and diverse land cover types, facilitating advancements in land cover mapping, environmental monitoring, and urban planning. Despite the model's strong performance, certain limitations are evident in the segmentation outputs. For ex-

ample, in Figure 4(c)-(h), fragmented road segments occur when roads are partially occluded by vegetation (LoveDA dataset), indicating the model's sensitivity to occlusions. Similarly, in Figure6, small vehicles in dense parking lots (Vaihingen dataset) are occasionally merged into background clusters due to limited spatial resolution. Additionally, the Dynamic Snake Convolution, while effective for linear structures, introduces a computational trade-off—inference time increases by 15% compared to the baseline Segformer (Table5). These challenges highlight the need for future work on occlusion-aware attention mechanisms and lightweight DSC variants for real-time applications.

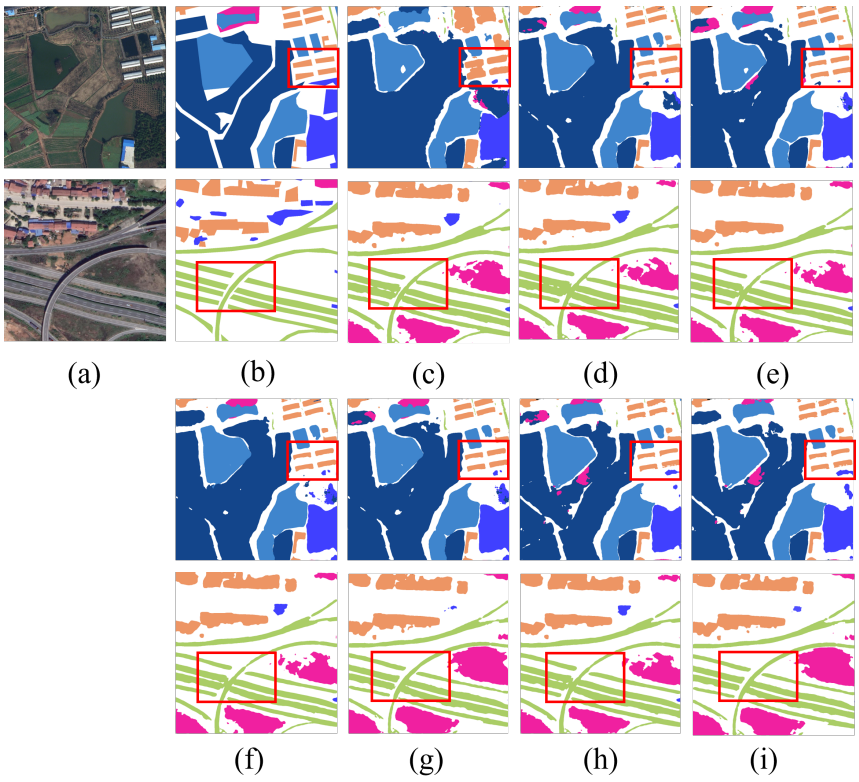


Fig. 4. Segmentation results of different Algorithms on LoveDA.(a) Raw Image; (b) Ground Truth; (c) FCN; (d) DeepLabV3; (e) Unet; (f) Unet++; (g) FPN; (h) PSPNet; (i) ours

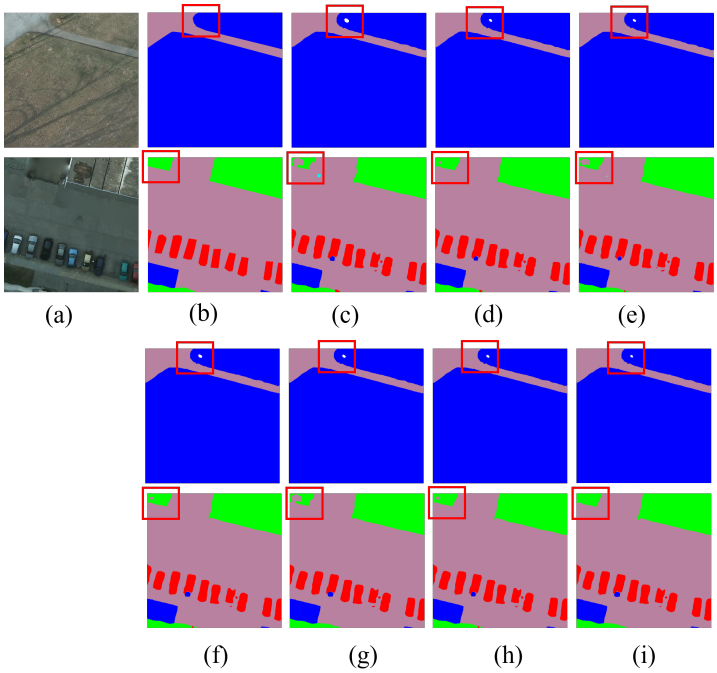


Fig. 5. Segmentation results of different Algorithms on Postdam (a)Raw Image; (b) Ground Truth; (c) FCN; (d) DeepLabV3; (e) Unet; (f)Unet++; (g) FPN; (h) PSPNet; (i) ours

5. Conclusion

This study proposes a boundary-aware semantic segmentation framework for remote sensing images by integrating Segformer's global context modeling with Dynamic Snake Convolution (DSC). The key contributions include: (1) a hybrid architecture that synergizes Segformer's hierarchical attention for multi-scale semantics and DSC's iterative offset constraints for slender structures (e.g., roads, rivers), (2) an auxiliary semantic branch to align cross-scale features and mitigate intra-class heterogeneity, and (3) comprehensive validation on LoveDA, Potsdam, and Vaihingen, showing mIoU improvements of 4.18%, 0.74%, and 2.09% over baseline models, with notable gains in fine-grained categories (e.g., 81.08% IoU for cars on Vaihingen). Despite its effectiveness, the model faces challenges in segmenting sub-10px objects (e.g., small agricultural patches) and incurs a 15% inference time overhead from DSC. Future work will focus on lightweight DSC variants for edge deployment, multi-modal fusion (e.g., SAR + optical), and occlusion-aware mechanisms to address complex urban scenes. This framework advances high-resolution land cover mapping and urban planning, with potential extensions to dynamic environmental monitoring through temporal data integration.

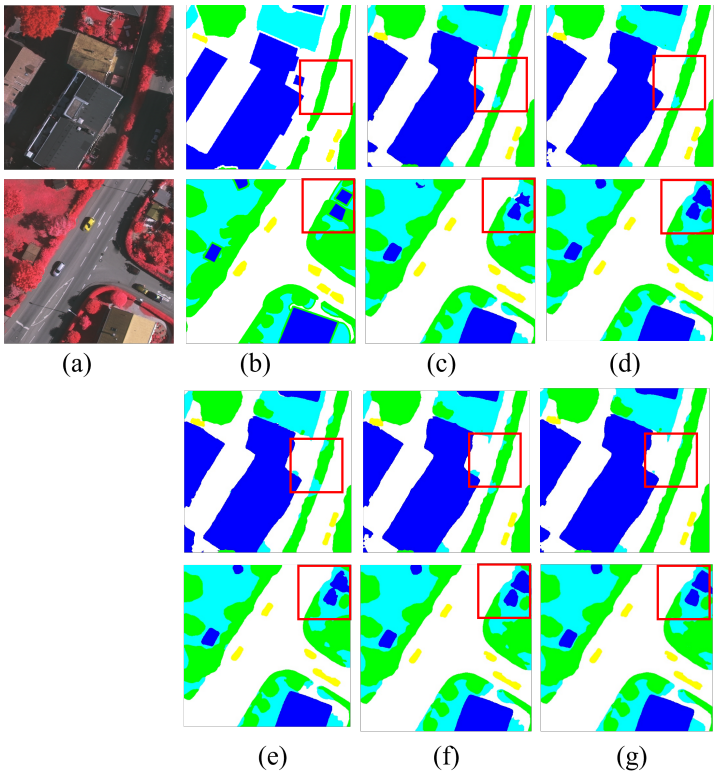


Fig. 6. Segmentation results of different Algorithms on Vaihingen (a)Raw Image; (b) Ground Truth; (c) Unet;(d) Unet-AFS; (e) PSPNet; (f) DeepLabV3; (g) ours

References

1. A., R.M., Y, W.: Optimizing intersection-over-union in deep neural networks for image segmentation. In: International Symposium on Visual Computing. pp. 234–244. Springer, Cham (2016)
2. A, S., Y, K.: Semantic segmentation of remote-sensing imagery using heterogeneous big data: International society for photogrammetry and remote sensing potsdam and cityscape datasets. ISPRS International Journal of Geo-Information 9(10), 601 (2020)
3. Borgwardt, K.M., Gretton, A., Rasch, M.J., et al.: Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics 22(14), e49–e57 (2006)
4. Ding, L., Lin, D., Lin, S., Zhang, J., Cui, X., Wang, Y., Tang, H., Bruzzone, L.: Looking outside the window: Wide-context transformer for the semantic segmentation of high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 60, 1–13 (2022)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. G, Z., T, L., Y, C., et al.: A dual-path and light-weight convolutional neural network for high-resolution aerial image segmentation. ISPRS International Journal of Geo-Information 8(12), 582 (2019)
7. Guo, Y., Liu, Y., Georgiou, T., et al.: A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval 7, 87–93 (2018)
8. H, Z., J, S., X, Q., et al.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890 (2017)
9. J, D., H, Q., Y, X., et al.: Deformable convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2–3 (2017)
10. J, G., S, S., W, J.: Dual-path feature aware network for remote sensing image semantic segmentation. IEEE Transactions on Circuits and Systems for Video Technology 34(5), 3674–3686 (2023)
11. J, L., E, S., T, D.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440 (2015)
12. J, W., Z, Z., A, M., et al.: Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation. arXiv preprint arXiv:2110.08733 (2021)
13. L, W., K, P., M, X., et al.: Sgformer: A local and global features coupling network for semantic segmentation of land cover. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16, 6812–6824 (2023)
14. LC, C., G, P., I, K., et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4), 834–848 (2017)
15. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems 33(12), 6999–7019 (2021)
16. Lian, X., Pang, Y., Han, J., Pan, J.: Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. Pattern Recognition 110, 107622 (2021)
17. Markus Gerke, I.: Use of the stair vision library within the isprs 2d semantic labeling benchmark (vaihingen). Use of the stair vision library within the isprs 2d semantic labeling benchmark (vaihingen) (2014)
18. O, R., P, F., T, B.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer, Cham (2015)
19. P, S.J., P, E., R, K.S.: Unveiling the secrets of brain tumors: A fuzzy c-means and u-net convolution approach for enhanced segmentation. INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL 19(2) (2024)

20. Pinheiro, P.O., Collobert, R.: From image-level to pixel-level labeling with convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1713–1721 (2015)
21. Rodrigues, C.M., Pereira, L., Rocha, A., et al.: Image semantic representation for event understanding. In: 2019 IEEE International Workshop on Information Forensics and Security (WIFS). pp. 1–6. IEEE (2019)
22. S, J., J, L., Z, H.: Dpcfn: Dual path cross fusion network for medical image segmentation. Engineering Applications of Artificial Intelligence 116, 105420 (2022)
23. S, Y., L, W., L, T.: Threshold segmentation based on information fusion for object shadow detection in remote sensing images. Computer Science and Information Systems 00, 23–23 (2024)
24. S, Z., J, L., H, Z., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Computer Vision and Pattern Recognition. pp. 6877–6886. IEEE (2021)
25. S, Z., J, L., H, Z., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6881–6890 (2021)
26. V, B., A, K., R, C.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(12), 2481–2495 (2017)
27. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. Advances in Neural Information Processing Systems 30, 5998–6008 (2017)
28. Voulodimos, A., Doulamis, N., Doulamis, A., et al.: Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience 2018(1), 7068349 (2018)
29. Wang, W., Zhou, C., He, H., Ma, C.: Advancing uav image semantic segmentation with an improved multiscale diffusion model. Tehnički vjesnik 31(6), 1859–1865 (2024)
30. X, H., Y, Z., J, Z., et al.: Swin transformer embedding unet for remote sensing image semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing 60, 1–15 (2022)
31. Xu, H., Zhang, X., Li, H., et al.: Seed the views: Hierarchical semantic alignment for contrastive representation learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 45(3), 3753–3767 (2022)
32. Y, M.: Research review of image semantic segmentation methods in high-resolution remote sensing image interpretation. Journal of Frontiers of Computer Science and Technology 17(7), 1526–1548 (2023)
33. Yin, S., Wang, L., Teng, L.: Threshold segmentation based on information fusion for object shadow detection in remote sensing images. Computer Science and Information Systems (00), 23–23 (2024)
34. Z, G., C, G., Z, F., et al.: Integrating masked generative distillation and network compression to identify the severity of wheat fusarium head blight. Computers and Electronics in Agriculture 227, 109647 (2024)
35. Z, L., Y, L., Y, C., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022 (2021)
36. Zeng, F., Yang, B., Zhao, M., Xing, Y., Ma, Y.: Masanet: Multi-angle self-attention network for semantic segmentation of remote sensing images. Tehnički vjesnik 29(5), 1567–1575 (2022)
37. Zhong, L., Ruijun, B., Jun, H., et al.: Aircraft detection algorithm for remote sensing images based on adaptive feature fusion and multi-scale output. Microelectronics and Computer 38(4), 40–45, 51 (2021)

Yanting Xia currently serves as a lecturer at Geely University of China. Her research interests include Embedded systems, LiDAR and neural networks. In recent years, she

has published over 10 papers in various academic journals and conference proceedings, including the Journal of Computing and Information Technology and technical Gazette. In addition, she has led several scientific research projects at the national and provincial-ministerial levels. Her research accomplishments also include one national-level textbook, one invention patent, and two software copyrights.

Lin Zhang is a lecturer at Geely University of China. His research focuses on embedded system design and development, circuit layout and design, and host computer development. In recent years, he has published three SCI international journal papers (including one in SCI Zone 1) and three papers in Chinese core journals, such as Technical Gazette, Advances in Production Engineering and Management, ComSIS, and CIT. In addition, he has led several scientific research projects at the national and provincial-ministerial levels. His research achievements also include one invention patent, one utility model patent, and two software copyrights.

Ting Guo is currently the Director of the Office at Geely University of China. She obtained her Master's Degree in Public Administration from University of Electronic Science and Technology of China in 2015. Since then, she has been engaged in teaching and administrative work at private higher education institutions, including management roles in the Department of Ideological and Political Education and the Department of General Education. Her research focuses on public teaching management and efficiency enhancement.

Jin Qi is currently the Deputy Secretary of the Party Branch and Assistant Dean at the School of Electronic Information Engineering, Geely University of China. His research interests focus on student affairs management. With 17 years of experience in higher education teaching and administrative management, he has published three academic papers, participated in four research projects, and co-edited one ideological and political education textbook.

Received: March 12, 2025; Accepted: April 28, 2025.

Computer Science and Information Systems 22(3):1011–1046 https://doi.org/10.2298/CSIS241024034B

Extending Hybrid SQL/NoSQL Database by Introducing Statement Rewriting Component

Srđa Bjeladinović

University of Belgrade, Faculty of Organizational Sciences
Jove Ilića 154, Beograd, 11000, Belgrade, Serbia
{srdja.bjeladinovic}@fon.bg.ac.rs

Abstract. Contemporary organisations often include different business subdomains, for which it is neither easy nor optimal to decide on using an exclusive database type. The hybrid SQL/NoSQL databases encompass various types of databases unified into a unique logical database. At the same time, they provide the usage benefits of working with the SQL and the NoSQL databases simultaneously. Recently, there has been an increase in research that deals with the challenges of hybrid databases' query optimisation, especially query rewriting. This trend opened up possibilities for analysing the influence of applying different statement rewriting techniques to other data manipulation statements besides queries (i.e. INSERT, UPDATE and DELETE) and its impact on the average execution times. In this paper, a process model for applying automatic hybrid statements' rewriting was designed, and the architecture for the hybrid database was extended with the newly developed Statement Rewriting Component (SRC). The tested use cases were conducted on the example of Oracle/MongoDB/Cassandra hybrid before and after introducing SRC. The tests have shown particular decreases in the average execution times of the system with the SRC.

Keywords: Hybrid database; SQL; NoSQL; statement rewriting; database architecture.

1. Introduction

Since the beginning of the 90s of the last century, it has been identified that individual systems cannot always satisfy all business needs and that it is often necessary to design, manage and maintain systems made up of several other systems. The development of complex systems, whose components are intensive software systems, usually developed by different manufacturers, was significantly influenced by the rapid growth of society and industry [1]. The constant increase of the business processes complexity, the expansion of the portfolio of products and services, the diversification of business between companies and the expansion of software specialised for specific domains of business or industrial branches have further promoted the development of systems that integrate other, independent, systems. Individual and independent systems cannot always achieve a higher business mission. Still, their collaboration and integration into complex systems enable the realisation of missions beyond their capabilities [2]. These systems are sometimes referred to as Component-Based Software [3] or Systems-of-Systems (SoS) [4]. Today, large and complex systems that contain other constituent systems [5] are represented in many fields [6][2][7], such as transportation networks, smart energy grids, space,

aeronautics, e-commerce applications, medical assistance, emergency management, and databases.

To eliminate possible terminological ambiguities, the interpretations of common terms used in this paper are defined before the term hybrid database. Generally speaking, two main aspects of databases are data model and DataBase Management System (DBMS). As defined by ANSI [8], *"DBMSs can be categorised in terms of the data model which is supported and the data language provided for interacting with this data model. A data model defines the acceptable types of data structures."* So, data models, from the DBMS point of view, represent an appropriate structure for storing and managing data. The authors [9] define a data model as a theoretical approach that determines the way of specifying and designing a specific database, while a DBMS is a particular data processing technology, i.e. a software system, which allows management of large amount of data and implements a specific data model. Further, data model instances can belong to one of three types:

- Conceptual data model - Corresponds to the conceptual schema of the entire system and describes the domain semantics without technological and implementation specifics;
- Logical data model - Describes the semantics in the context of a specific technology for data manipulation, i.e. specific structures suitable for system development (tables and columns, JSON documents, graphs, etc.). It represents the result of mapping the conceptual data model (conceptual schema) into the schema of a data model supported by a specific DBMS [10];
- Physical data model - A data model corresponds to the design of the internal, physical structure of the database based on the logical model and the specification of all non-functional requirements [9].

A hybrid is a unique logical database designed over a single conceptual data model. The conceptual data model is translated into the logical data models of the particular DBMSs, which are chosen for building the hybrid's components. A hybrid database integrates all logical data models of its components into a unique logical unit via unified processes of designing and administrating hybrid and its components, as well as using an integrated meta-repository.

Depending on the types of integrated DBMSs, logical models contained in a hybrid SQL/NoSQL database can describe tables and columns that implement the relational model, JSON document and its fields, ordered key-value pairs, column families, or graph nodes and their relationships. The hybrid SQL/NoSQL databases provide the benefits of working with the SQL and the NoSQL databases simultaneously. Through the hybrid, each component (DBMS) participates in achieving the broader business mission of the company, which could not be satisfied as an isolated DBMS.

Hybrid databases have been developed to satisfy the growing need of contemporary businesses for storage, access and processing of data, regardless of the source and degree of data structuredness. This approach in developing data-intensive systems enables the use of suitable DBMS representatives for each business domain or, more precisely, for each business subdomain. Contemporary organisations often unify business functions that use strict and, in advance, defined data structures with other business functions that use highly flexible data structures. Apart from the simultaneous use of data with different levels of structuredness, various criteria may have higher or lower importance, depending

on the business needs. For example, we can observe the information system of a marketing company, which entails monitoring and executing financial transactions (with the banks, business partners, employees, etc.) and marketing promotions on social media. We use different potential criteria for the successful execution measurement of each function. Regarding financial transactions, it is necessary to provide data integrity, i.e. to use more structured data with the highest level of consistency. In the case of social media promotion, data integrity and structuredness are often secondary-significant to the data availability and fast analytical processing.

Before the emergence of hybrid databases, the challenge of using data with different levels of structuredness was solved in one of the following two approaches: 1) limiting organisations to using an exclusive database type for all subsystems; 2) using different types of databases for each of the subsystems (or similar subsystems). In the previous example of a marketing company, as in many other examples of business, it is not always easy nor optimal to choose a single database type. The usage of exclusive database type gives the organisation benefits for some specific subsystems. At the same time, other subsystems of that company have the limitations of using a non-suitable or non-optimal database type. Because of that, the first legacy approach cannot resolve all specific demands raised by the necessity of using different types of databases.

The second approach to the problem, using different types of databases (with varying data models) whose operation is not fluidly represented as the unique logical database, implies additional expenses of connections, integration and unified administration. Compared to the two prior approaches (single database and multiple databases over multiple models), a hybrid system of databases enables integration and concurrent use of different technologies. Hybrid database users benefit from using the best functionalities of different database types. Users of hybrid do not have to worry about their integration since different types of databases represent the components of the unique logical (hybrid) database, which uses a single data model. Hybrid database usage enables the engagement of suitable data storage and management technology for every business subdomain. The most common and comprehensive representatives of the hybrids are the hybrid SQL/NoSQL databases.

Even though, in recent years, there has been a noticeable increase in the number of works and research on hybrid database optimisation and statement rewriting, additional space for research has emerged. For the optimisation of databases, in general, different approaches can be used, such as logical optimisation, physical optimisation (e.g. horizontal and vertical partitioning), various ways to access data (e.g. via tables, indexes, materialised views) and reformulating the way statements are written. The latter contains a potential not fully explored in existing papers.

This manuscript is a continuation of research and expansion of the hybrid SQL/NoSQL database and its architecture presented in previous works [11] [12]. This paper aims to give answers to the following research questions which have arisen during the continuation of the research:

- Research question 1: How do we specify the automatic usage of statement rewriting techniques in a hybrid SQL/NoSQL database?
- Research question 2: Is it feasible to develop a new dedicated component for statement rewriting and integrate it into the existing architecture for hybrid SQL/ NoSQL databases?

– Research question 3: How do the applied statement rewriting techniques influence the duration of the statement execution, based on the example of the Oracle/MongoDB/Cassandra hybrid database?

The answer to the first research question covers the extension of the existing hybrid database design methodology [11], and it introduces the designed process model and activities for rewriting entered statements (particularly INSERT, UPDATE and DELETE). Automatic statement rewriting is one of the powerful optimisation techniques, and it aims to reduce statements' average execution time of the initial architecture of SQL/NoSQL database [12] by introducing a newly developed *Statement Rewriting Component (SRC)*. The second question of this paper deals with how to implement SRC on the existing architecture for a hybrid SQL/NoSQL database and how to seamlessly integrate SRC with the execution of other components of a hybrid. The third question aims to quantify conducted practical tests of the selected use cases and compare the achieved results of the existing SQL/NoSQL architecture (without SRC) and extended SQL/NoSQL architecture (with newly developed SRC component).

However, there are some limitations in this paper. The current architecture version of the hybrid SQL/NoSQL database is still a prototype, and not all functionalities are entirely implemented. The SELECT statement is out of the scope of this paper for several reasons. The first one is that the initial architecture version for SQL/NoSQL databases (without SRC) already contains some built-in mechanisms, but only the basic ones, like indexing and partitioning. We acknowledge that query optimisation is thoroughly researched in the analysed papers. Because of that, in this paper, we focus on one particular optimising technique (statement rewriting) of the other data manipulation statements (INSERT, UPDATE and DELETE), which is, to our knowledge, not researched enough. The main limitation of this paper is that we focus exclusively on providing automatic statement rewriting as one of many techniques for statement optimisation, but at the same time, it is a very promising technique considering its potential effect on the statement's average execution time. However, the Optimisation Rules Repository, created and introduced in this paper together with SRC, present a useful base for integration and support not only with the new SRC component but also with future components of the hybrid.

In order to present the existing research in the field and the original results obtained in the process of answering the research questions, the paper has the following organisation. Section 2 overviews the existing directions of hybrid databases' design and use. Section 3 deals with the first research question by presenting the process model designed for automatically applying statement rewriting techniques in hybrid databases. Section 4 describes the architecture extension, with a newly created component for statement rewriting. A description of the SRC component's role in the architecture and an explanation of its functioning aims to answer the second research question. Section 5 lists the use cases chosen for testing and comparing the average statements' execution times for the hybrid SQL/NoSQL databases without the SRC component and for the hybrid SQL/NoSQL databases with the newly created SRC. The Oracle/MongoDB/Cassandra hybrid database was used as the test environment. Section 6 contains experimental results. Section 7 consists of conclusions and gives directions for future research.

2. Related Work

In general, hybrid databases can be defined as integrated data systems comprised of multiple autonomous databases [13] or as databases that incorporate different types of databases into a unique logical database [11]. By reviewing their presence on the market [14], it can be concluded that the dominant databases are still relational (also commonly called SQL databases after the standardised query language they use). However, NoSQL databases, suitable for working with large amounts of data, are increasingly being used [15]. For years, big companies such as Facebook, Amazon and Google have been using SQL and NoSQL databases to complement each other [16]. Recent research on how four representatives of NoSQL handle a variety of data is presented in the paper [17]. The hybrid SQL/NoSQL databases unify the use benefits of the SQL and NoSQL databases for the purposes they are designed for by overcoming individual limitations typical for particular database types [18]. At times in literature, the NoSQL databases are called non-relational [19][20][21]. However, non-relational databases can be generally treated as a broader term, including other types of databases [22]. The increased popularity of the NoSQL databases directly affected the focus of databases' hybridisation in recent years. The focus has shifted towards integrating NoSQL and SQL databases into the hybrid SQL/NoSQL database [23]. Therefore, the authors [19] state that hybrid databases aim to use data from relational and non-relational databases and to provide conjoint results in a single output. The paper [24] defines the term hybrid database or just the hybrid as "systems where there are several databases implemented that can be relational and/or NoSQL." Based on all of the covered definitions, it is very important to highlight the common trait of a hybrid. Each hybrid database, no matter how many different types of databases it includes, is created over just a single data model and thus is designed and maintained as a unique logical database. Various parts (components) of a hybrid's conceptual data model are implemented in different DBMSs, which can represent very different database types. However, users always access and interact with a hybrid database as with a single database (which implements the entire data model and its necessary logic), no matter which particular hybrid component, e.g., a specific DBMS, stores searched data.

The review in the field of hybrid databases identified four groups of papers of interest. Papers in the first group explain similarities but also crucial differences between hybrid databases and other contemporary databases, which contain multiple systems similar to the SoS. The second group comprises the articles that contribute towards setting hybrid databases' general principles (i.e. design, integration, uniform use). This group of papers set the foundations for hybrid databases. The articles whose research focus is on different aspects of performance measurements and optimisation of various database types represent the third group of research papers. In addition, the significance of these papers is reflected in the fact that different database types can be components of the hybrid. Finally, the fourth group of analysed papers directly discusses the hybrid database's statements optimisation and rewriting.

2.1. Similar but Different: Alternatives for Hybrid Databases

In recent years, significant progress has been made in the directions that have a tangential research question with hybrids, which is the integration of different systems of databases, i.e. different types of databases. However, due to noticeable differences in the way of

solving the mentioned problem, as well as obvious distinctions in these approaches, it would be risky, even indescribable, to equate them. However, because of their popularity, some of these approaches will be briefly mentioned here. These approaches are polystores, polyglot persistence, multi-model databases, and Object-NoSQL Data Mapper (ONDM) frameworks. Table 1 shows the taxonomy of the approaches from the first group, their key characteristics, similarities and differences compared to a hybrid database.

Table 1. Taxonomy of the approaches dealing with the SQL and NoSQL integration (alternatives to the hybrid databases)

Approach	Key characteristics	Similarities to hybrid DB	Differences to hybrid DB
Polystores	Orchestration of different models and improving the usage of a uniform language on multiple databases	Simultaneous use of several different types of databases; Single query language	Numerous isolated data models, which are subsequently linked; Not the unique logical database
Polyglot persistence	Using different types of databases to solve conflicting requests	Use "the best" type of database for the concrete requests	Absence of a unique language for accessing all databases; Not the unique logical database
Multi-model databases	Ease of use of different models within a single database	Using different models for different requirements; Using one logical database	Only one DBMS; Number of different data models limited by particular DBMS
ONDM	Object model instead of an approach for integrating SQL and NoSQL	A sense of using one logical database	Only one data model (object) instead of a variety of different ones; List similarities with the hybrid databases

Polystores is an approach that deals with the simultaneous use of several different types of databases. They don't have one common data model for the entire logical database, but instead numerous isolated data models, which are subsequently linked. The principles of polystores are described in detail in the paper [25], in which the authors emphasise the importance of using a uniform language over multiple data models. They have presented a BigDAWG prototype composed of SciDB, Accumulo, Postgres and S-Store DBMS. The research mentioned above was extended in subsequent works [26][27][28][29], while the authors [30] have developed purpose-built modelling tool, named TyphonML, which automatically generates CRUD API for polystores. The consequence of the simultaneous, not necessarily related, design and the use of multiple data models is reflected in the ploy-

stores' inability to provide a unique administration process for all databases in use, which implies an additional difficulty in controlling and reducing unwanted data redundancy effectively. The maiden authors of polystores [25] mentioned this as a direction for further research. In addition to the above, synchronisation and simplicity of system expansion, as two of the indicators of integration and usability [31], are aggravating in the case of polystores architecture.

Polyglot persistence represents one of the approaches of using different types of databases to solve conflicting requests, in which only one database cannot solve all tasks. A detailed description of this approach and its specific variants are given in the paper [32]. The authors of the mentioned work identify the following subcategories, which differ in the realisation of the polyglot persistence approach: (I) Application-coordinated Polyglot Persistence (ACPP); (II) Service-oriented Polyglot Persistence (SOPP); (III) Polyglot Persistence as a Service (PPS). Polyglot persistence defines four types of cardinalities between application modules and DBMS: One-to-one, Many-to-one, One-to-many and Many-to-many. The first two cardinalities (1-1; M-1) eliminate the possibility of using different types of databases because they have only one destination database type. The other two cardinalities (1-M; M-M) imply using different types of databases but with the parallel use of multiple query languages, which the authors clearly emphasised in their work [32]. Polyglot persistence with cardinalities 1-1 and M-1 is not of interest to this paper because, with these cardinalities, clients are limited to using only one type of database. On the other hand, cardinalities 1-M and M-M in polyglot persistence approaches, with all their variants (ACPP, SOPP and PPS), allow the use of several different types of databases but require the simultaneous knowledge and use of several query languages. Consequently, the used databases cannot be treated as unique logical databases because it is necessary to separate the data by database types, which results in additional difficulties of integration and reduction of redundancy.

Multi-model databases were created to offer ease of use of different models within a single database. For example, Redis, which is primarily a key-value database, supports document-oriented and graph models, while Elasticsearch supports search engine and document-oriented types of databases [14]. A detailed review of multi-model databases was given by the authors [33]. Authors Lu and Holubová investigated the multi-model database from several aspects: the way of handling a variety of data [34] and the comparison with polystores [35]. Some authors recently compared multi-model databases with polyglot persistence [36][37]. Elaborate analysis of the evolution of multi-model databases and directions for further development are detailed in the paper [38]. The authors Lajam and Mohammed [32] state that despite the support for multiple models in one database, multi-model databases still cannot adequately compete with polystores, i.e. systems made up of several different types of databases, primarily in terms of satisfying non-functional requirements such as scalability and performance. Apart from those mentioned above, a significant limiting factor in comparisson to hybrids is the closedness of each individual DBMS. Although certain DBMSs support several different models, the extension of the set of supported models is strictly locked and completely dependent on the DBMS manufacturer. Vendor lock-in can lead to a model not being implemented in the observed DBMS in the future, either. Those mentioned above noticeably limit the possibilities of designers to compose a database of several models and maintain and expand the set of models used by the user.

ONDM research focuses on mapping and integrating object and NoSQL models and extends the functionalities introduced by Object-relational mapping (ORM) frameworks [39]. However, the primary focus of these frameworks is not the direct integration of SQL and NoSQL types of databases. Instead, to a certain extent, they achieve it via an object model.

Table 2. Pros and cons of the analysed approaches

Approach	Pros	Cons
Polystores	Simultaneous use of different types of databases; Usage of a uniform language	The inability to provide a unique administration process for all databases; Difficulty in controlling and reducing data redundancy; Synchronisation and simplicity of system expansion
Polyglot persistence	The use of several different types of databases	Use of several query languages; Difficulties of integration and reduction of redundancy
Multi-model databases	Suported diferrent models; One logical database; The convenience of accessing and managing only one DBMS	Limitations of (broaden) usage of different models in a single DBMS; Vendor lock-in and closedness of each individual DBMS; It is harder to achieve non-functional requirements such as scalability and performance
ONDM	A single data model (object) tries to “fit all”	Not many authors treat this approach as a genuine alternative to a hybrid; Only one data model (object) with its limitations instead of a variety of different ones

Table 2 summarises the pros and cons of the four analysed approaches. The conclusion that can be drawn is that each approach, of course, has advantages and disadvantages, the same as the hybrid databases. Nevertheless, it can be unequivocally concluded that, despite their benefits, the four analysed approaches cannot be treated as adequate alternatives to hybrid databases. The reason is either because of the problem they are trying to solve or because of the way they approach the solution. In particular, although polystores support the parallel use of different types of databases using a unique language, the absence of the possibility of designing the entire database as a unique logical whole without unwanted redundancy makes it impossible to equalise polystores and hybrids. The main difference between the polyglot persistence approach and the hybrids is reflected in not supporting a single language for working with all types of databases. Individual multi-model databases and ONDM show significant disadvantages compared to the hybrids,

which result from the limits of a single DBMS usage, i.e. single data model usage (object model).

2.2. General Principles of Hybrid Databases

The second group of papers deals with hybrid databases' general principles. The authors [40][41][42] performed a trend review of the contemporary databases' designs, including common, current and future development directions. The hybrid design and use challenges originate from the heterogeneous characteristics of different database types integrated into the unique logical entity. The authors [43] dealt with developing dedicated design methodologies for hybrid databases. The authors [44] analysed software test techniques for the hybrid databases. The authors [23] highlighted the use benefits of the hybrid SQL/NoSQL databases and presented dedicated concepts of the extended ER model. These components are useful while modelling hybrid databases. Primarily, theoretical integration possibilities of the relational (MySQL) and graph database (Neo4j) are discussed in the papers [13][45]. In their paper [46], the authors focused on the diversity of NoSQL models (key-value, document-oriented, column family, graph). They highlighted that relational databases' traditional design approaches cannot be directly applied to NoSQL design. They presented the Mortadelo framework based on the model-driven transformation process. It transforms the generic data model into the intermediate logical model (specific for some of the four NoSQL models). The latter should then be transformed into the implementation code of the particular DBMS. The authors [47] presented the migration approach from the SQL into the hybrid SQL/NoSQL database using ontology to define data schema. In their paper, the authors [48] presented how a conceptual data model could describe Big Data stored in the NoSQL database. Unlike the Mortadelo framework that uses specific logical models for each NoSQL database type, the authors [48] created a generic logical layer suitable for work with three types of NoSQL databases (document-oriented, column family, and graph). By applying the QVT (Query-View-Transformation) rules, this layer enables efficient transformation execution into the physical model, decreasing the influence of the destination NoSQL database's technical specifications. The authors [43] analysed the metamodel integration possibilities of different database types. Through the application of the lightweight extension approach, this paper describes the way of adding new elements and constraints to the existing metamodels. By applying one conceptual, one logical and one physical data model, this approach enables the integration of different types of databases. The authors depicted this in the example of the system comprising one relational database and two document-orientated databases. Some authors [49] approached the topic of hybrid databases from the aspect of domain-specific language (DSL). The limitation of the domain-specific languages is the lower prevalence than of the standardised SQL, supported by leading manufacturers or relational databases. Also, authors [50][51] researched some of the challenges of working with heterogeneous databases. The authors [50] have emphasised differences in the consistency and transaction limitations between SQL and NoSQL DBMS.

Furthermore, they have developed a comprehensive approach for managing distributed transactions with guaranteed ACID in heterogeneous data store environments, regardless of whether the individual data stores support ACID. The authors [51] focused on the challenge of mapping syntactically and semantically related attributes among schemas. The

heterogeneous data stores used as the target databases can also be useful for this research because they can present the components of the hybrid databases.

2.3. Database Optimisation and Performance Measurement

The third group of papers deals with different aspects of performance measurements and optimisation of different database types. The authors [52][53] compared query execution in relational and non-relational databases. The paper [53] compared the three DBMS's login and usage performances (PostgreSQL, MongoDB and Cassandra), while the article [52] analysed in detail query execution of the three most common SQL database representatives (Oracle, MySQL and MS SQL Server) and four representatives of the NoSQL databases (MongoDB, Redis, Cassandra and GraphQL). The authors have developed and used a data model of train stations and stops. They have concluded that for Slovakia, it is justified to use an SQL database for storing and managing data. In contrast, for countries with a larger amount of train data, such as the Netherlands or Germany, it was suggested the usage of a NoSQL database. The authors [54] have decreased the performance gap between the SQL and NoSQL databases by introducing a dedicated binary format for JSON. The applicability of this solution is reflected in the hybrid databases whose components support JSON format. However, this also results in usage limitations on the hybrids that contain NoSQL databases without JSON support. Detailed analysis of the evolution of using different JSON functionalities, in both native and binary formats, and their influence on the performance of Oracle DBMSs was given in the paper [55]. The authors Kemper and Neumann [56] approached the optimisation of different databases in use from the aspect of the gap elimination that emerges while using traditional OLTP and OLAP systems. They suggested the creation of a hybrid OLTP & OLAP system that would contain data versioning. As stated by the authors, introducing data versioning would enable the separation of data manipulation from query execution while using a hybrid system's database for both purposes. This solution allows for the execution of the BI queries *"on an arbitrarily current database snapshot system"* while eliminating the consumption of the resources derived from the additional activities necessary for data adjustment and transfer from the traditional OLTP systems into the OLAP systems [56]. From the theoretical aspect, the authors [57] dealt with optimising a large amount of data with a different degree of structuredness. In addition to the presented mathematical model, the authors showed postulates of the DSL, which is based on the unification concept with the aim of easier information search.

The author [58] researched the hybrid database design directions. These were inspired by finding a solution to the challenge of hardware components' optimisation and their specificity use with the aim of offloading database operators. The author has been using parsing and rewriting components of the existing DBMS (the paper mentions PostgreSQL as a potential candidate) and optimiser (whose main part is cost optimiser that calculates operations' expense based on the data from the dictionary). With those components, the author optimised the execution plan considering different execution engine types. Also, the paper focuses on the execution plan adjustment to the hardware components without considering hybrid SQL/NoSQL database optimisation specificities. The authors [59] dealt with the optimisation of the hybrid databases' hardware components, with a focus on CPU/FPGA, while the priority on CPU/GPU was given in the papers by the authors

[60][61][62]. Even though, in general, the mentioned papers [60][61][62] discuss the optimisation of hybrid systems and hybrid databases, they will not be further analysed in this paper, given their focus on the hardware optimisation.

2.4. Hybrid Database's Statements Optimisation and Rewriting

The fourth group of papers deals with hybrid database statements optimisation and rewriting. The authors [63] dealt with hybrid optimisation of "classical" databases and MapReduce. They analysed the advantages and limitations of databases and the MapReduce systems and highlighted the benefits of using hybrids made up of the mentioned types of repositories. Finally, they presented a dedicated and improved version of the query optimiser named AquaPlus. The purpose of the presented optimiser is to use database features as much as possible (like index and partitioning), intending to reduce the amount of data needed to be processed by the MapReduce hybrid component. The authors [64] had a compatible approach. In their paper, they researched the effect of physical optimisation, predominantly partitioning, on the average time of query execution in the SQL database (Oracle DBMS). After that, the authors compared the SQL database's improved performance with the graph database (Neo4j). They concluded that the gap was decreased primarily in the complex queries (subqueries and JOINs). A noticeable step towards improving query optimisation in a system made of heterogeneous databases was achieved by the authors [65]. Even though they do not explicitly use the term "hybrid databases" but "virtual data source" instead, they have focused their research on the hybrid domain, whose components are relational and NoSQL databases. The authors suggested the introduction of the mediation component, whose aim is to optimise queries for efficient execution over multiple databases of different types. They used a joint schema to describe all data sources in the system. In addition, they enabled parallel execution over data sources and optimal plan generation by using dynamic programming. However, these authors focused solely on queries, and they did not deal with other statements. Li and Gu [66]) developed a useful solution to the nested query optimisation problem over the hybrid operation. In the mentioned paper, they solved the integration problem of MySQL, MongoDB, and Redis as the hybrid database components and simultaneous query execution over the relational and NoSQL databases. They presented the Multiple Sources Integration architecture (MSI) that supports databases' integration. Also, they graphically presented the communication between components for optimisation with the SQL parser on one and the SQL router on the other side. In addition to explaining the algorithm logic that was used for the nested query optimisation, the authors state that *"the expression of the query conditions must be carried out according to the distributive law, which also needs to be simplified based on the Espresso algorithm... and it is planned to be discussed in another paper"*[66]. Even though it is stated that the presented architecture supports optimisation, the article focuses only on one optimisation segment, and that is the nested queries.

The analysed papers contributed to the purposefulness of further research on the statement rewriting in hybrid databases and its following effects on the architecture changes. It opened up possibilities for and motivated authors to expand further the research of statement rewriting not on queries (because they have been too frequently researched) but on other statements (INSERT, UPDATE and DELETE) in the hybrid databases, which, according to our knowledge, are not explored in detail so far.

3. Process Model for Statement Rewriting of Hybrid SQL/NoSQL Database

Given that the authors of the presented papers took into consideration queries only, additional opportunities emerged for the research of other statements enhancing (INSERT, UPDATE and DELETE), and especially for researching the effects of the statements rewriting on the average execution time of a hybrid SQL/NoSQL database.

All leading manufacturers of Relational DataBase Management Systems (RDBMS) support the standardised SQL language. SQL is a declarative query language which enables uniform usage of data manipulation statements, creation and update of different types of objects of a relational scheme, as well as control of transaction execution. The procedural extensions of SQL language are not standardised. The manufacturers use specific extensions (i.e. Oracle use PL/SQL, and Microsoft use T-SQL) instead. Unlike the SQL databases, the NoSQL databases do not have a standardised query language, not even for CRUD operations. The complexity of the challenge is that in addition to the lack of unified language for all NoSQL databases, there is no agreement about the language of particular NoSQL databases' subtypes (key-value, document-oriented, column family, graph databases). Therefore, none of the NoSQL database subtypes have their unique language. The absence of the NoSQL database standardised language made it difficult to uniform statements, which are executed in the NoSQL components of the hybrid SQL/NoSQL databases. Also, the lack of NoSQL language standardisation limited the scope of possible solutions. Further language unification by the NoSQL databases' manufacturers will open possibilities for additional extension and improvement of the solution suggested by this paper.

With the aim of introducing the automatic application of statement rewriting (as an optimisation support technique) for all three statements mentioned above, a dedicated process model was designed. Figure 1 depicts the UML activity diagram for the statement rewriting process. The designed activity diagram starts with the statement input, followed by its decomposition and analysis.

Some of the hybrid database's main strengths are the integrated conceptual data model (common for all hybrid components), and the user disburden of knowing what component stores needed data. Given the scenario in which the input statement is executed over multiple target components, i.e. different DBMSs of the hybrid SQL/NoSQL database, the next activity is statement decomposition on its integral parts. The approach presented in this paper, as well as the mentioned approach it extends, uses exclusive language for the statement input over the hybrid database, and that language is the standardised SQL. SQL language, popular among users, and the supported architecture eliminate the need to know or use other domain-specific languages for each particular hybrid component. The user can access all the data as if stored in a single relational database. The user uses only the SQL syntax, regardless of whether the destination of the input statement is the SQL component, NoSQL component or both. A detailed description of the architecture with the newly developed component for the statement rewriting, which enables the stated operation, is given in Section 4.

The preparation activity of a statement rewriting starts after analysing and decomposing the entered statement and its integral parts. Depending on the input statement (SELECT, INSERT, UPDATE, DELETE), the process execution transfers to the appropriate branch following the decision node in the activity diagram. The decision node is

used for exclusive branching depending on the input statement type. Exclusive branching is present since the execution of different types of nested statements (i.e. UPDATE with subquery or DELETE with subquery) is not supported by the current version of the architecture. These are the limitations of the current version, as well as the direction for future work.

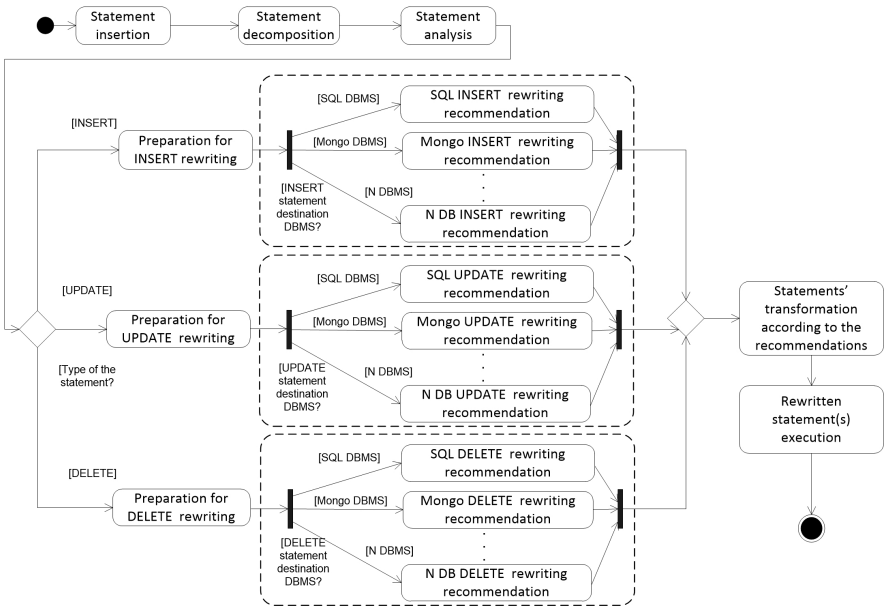


Fig. 1. The UML activity diagram for automatically applying recommendations for statement rewriting on the SQL and NoSQL components of the hybrid SQL/NoSQL database

The statement rewriting recommendations for all hybrid DBMSs that use SQL language are given within one activity of a single branch, marked with [SQL DBMS]. It is the uniformity of SQL language that made this possible. Each statement type will have a single [SQL DBMS] branch.

A different situation is observed with the NoSQL components. Given the absence of a standardised NoSQL language, it is necessary to give specific recommendations for each exact NoSQL DBMS, which is depicted in Figure 1 through the appropriate flows with conditions ([Mongo DBMS]. . . [N DBMS]). Each of these flows gives specific recommendations for a particular NoSQL DBMS in the syntax it supports. A "three-dot" symbol on the diagram suggests that the hybrid database can consist of an arbitrary number of different NoSQL components, and N DBMS represents the Nth NoSQL DBMS. The fork node enables parallel rewriting of different parts of the input statement. These parts could be executed into various destination components of the hybrid database.

Branch flows with accepted recommendations meet in the join nodes for each statement type. Next, all flows from all statements meet in the merge node. After merging flows, statement transformation takes place. It follows the identified recommendations for

statement rewriting. If necessary, the keywords of the entered SQL statement are mapped according to the language of the destination database, i.e. the hybrid component. Syntax transformation does not take place if the destination database is SQL, given that the input statement is already in SQL language. Thus, a prepared and (according to the accepted recommendations) rewritten statement is then executed. Statement execution represents the last activity on the diagram depicted in Figure 1.

The presented diagram displays the generic logic of introducing the statement rewriting over different components of a hybrid database. A hybrid Oracle/MongoDB/Cassandra database is the chosen representative to depict some of the supported statement rewriting use cases enabled by this approach. As the name suggests, this hybrid consists of three components: Oracle DBMS, the representative of the relational DBMS; MongoDB, the representative of the document-oriented DBMS; and Apache Cassandra, the representative of the wide-column DBMS. The selection was made based on the popularity rankings of these DBMSs. At the time of writing this paper, Oracle was the most popular SQL system [67], MongoDB was the most popular document-oriented and NoSQL system overall [68], while Apache Cassandra was the most popular wide-column DBMS and the third most popular NoSQL system [69]. Although, from the technical point of view, there is a possibility to have multiple SQL components, the core idea behind a hybrid is to use the most suitable database type (i.e. hybrid component) for a particular organisational subdomain and a particular set of business requests. Because of that, we don't find it appropriate to introduce more than one SQL DBMS but intentionally use the chosen one SQL DBMS for all requests that the SQL database type should cover. However, the possibility of switching to another SQL system is supported. That scenario boils down to the migration process from one database to another (for example, from Oracle to SQL Server, from MySQL to PostgreSQL, etc.). However, the migration between different SQL databases is not in the scope of this paper.

On the other hand, to demonstrate the usage of different NoSQL subtypes (particularly document-oriented and wide-column), two NoSQL components are present in the hybrid database, which was developed to practically test the architecture before and after introducing SRC. Expanding the test database with even more additional subtypes would be useful, but that is planned for upcoming research. The main reasons for that are the existing limitations of the test environment and the complexity of the parallel introduction of additional components on the old architecture without SRC and the improved architecture with SRC, which exceeds the extent of the conducted research.

Section 6 consists of the average execution time for each tested use case. The supported recommendations for certain statements rewriting, classified into Oracle/ MongoDB/Cassandra hybrid database components, are given in Table 3. These recommendations represent supported statement transformation techniques in the current version of the system. The list of the supported rewriting rules is extendible and will be expanded in future research.

The process model in Figure 1 is comprehensive, and it depicts rewriting techniques' application activities for all DML statements. Even the first version of the hybrid SQL/ NoSQL database incorporated basic optimisation techniques for the SELECT statement (query writing syntax, indexes, partitioning, etc.). The optimised SELECT is already shown through the test queries' execution results in earlier papers [11] [12]. For this reason, the focus of this paper is on the rewriting of other DML statements (INSERT, UP-

DATE, DELETE) through the application of the newly developed SRC component, as well as on the effects of stated rules usage on chosen use cases.

Table 3. Supported statement rewriting techniques for the hybrid Oracle/MongoDB/Cassandra database

Type of statement	Statement scenario	Recommendations for statement rewriting: *SQL component - Oracle*	Recommendations for statement rewriting: *NoSQL component - MongoDB*	Recommendations for statement rewriting: *NoSQL component - Cassandra*
INSERT	Inserting multiple rows	INSERT ALL syntax; *The decrease in the number of calls to the database, compared with the execution of multiple but individual INSERT statements*	BULK INSERT syntax; *The decrease in the number of calls to the database, compared with the multiple calls of the insertOne() method*	BATCH INSERT syntax; *The decrease in the number of calls to the database, compared with the multiple calls of the individual INSERT method*
UPDATE	Updating multiple rows	PARALLEL update; *Under certain conditions, hint PARALLEL enables the rewriting via parallel execution of the UPDATE statements instead of the default sequential execution*	*updateMany()* method; *Update multiple documents that satisfy the filter specified as the first argument instead of executing individual updateOne() methods numerous times*	BATCH UPDATE syntax; *The decrease in the number of calls to the database, compared with the multiple calls of the individual UPDATE method*
DELETE	Deleting all rows (DELETE without a WHERE clause)	TRUNCATE statement; *Using the mentioned DDL statement gives the benefits of quickly deleting a large amount of data while preserving the table structure*	*drop()* method; *The drop() method enables the quick deletion of all data, as well as the collection. Due to its structure flexibility, the new record insertion automatically recreates the collection and thus does not represent a significant resource cost for INSERT*	TRUNCATE statement; *Using the TRUNCATE statement gives the benefits of quickly deleting a large amount of data while preserving the table structure. It is a similar method to the SQL component (Oracle)*

For the rewriting of the INSERT statement, supported syntaxes are INSERT ALL (for the Oracle SQL component), BULK INSERT (for the MongoDB component) and BATCH INSERT (for the Cassandra component). The principle is the same, and the expected benefit is in the shortened average execution time of the rewritten statement due to the one call to the database, compared to the multiple calls for the execution of many individual

INSERT statements (before rewriting). This principle represents the essence of the enhancement regardless of whether INSERT ALL (Oracle), BULK INSERT (MongoDB) or BATCH INSERT (Cassandra) are used.

The UPDATE statement rewriting rules contain a recommendation for parallel execution by using the PARALLEL hint for the SQL component. For the MongoDB component, using the dedicated *updateMany()* method instead of the multiple *updateOne()* method should provide better results. The limitations of Cassandra's UPDATE statement are reflected in the obligatory WHERE clause, which must contain all the primary key fields. Besides, the IN operator usage is not supported in conjunction with the primary key fields. As a result, Cassandra does not support multiple target row selection within a single UPDATE statement. For the Cassandra component of a hybrid, the applied recommendation is BATCH UPDATE. The BATCH syntax is not much more complex than multiple UPDATE execution, but it reduces the number of calls to the database (one versus numerous). Each input statement should satisfy preconditions for the rewriting rule to be applicable. A detailed description of these preconditions is in Section 4.

Even though the DELETE statement supports parallel execution, as INSERT and UPDATE in SQL database, Table 3 shows a more efficient technique. However, its application scope is noticeably smaller. When the deletion of all table records is needed (the DELETE statement without the WHERE clause), the TRUNCATE statement can be executed while preserving the table itself, its structure and its constraints. The expected benefits of the average execution duration are reflected in the more efficient realisation of the DDL statement (TRUNCATE belongs to this category) instead of the multiple DELETE statement execution. The limitation of using TRUNCATE is that *rollback* is not accessible, given that *auto-commit* follows TRUNCATE by default (as well as other DDL statements). The TRUNCATE recommendation applies to Oracle and Cassandra components of the used test hybrid database. Applying the TRUNCATE statement is additionally powerful with the Cassandra NoSQL component. In Cassandra, it is not feasible to execute delete from a table without the WHERE clause (in contrast to SQL databases). Because the WHERE clause with the primary key is mandatory, every deletion of all records requires a preceding SELECT statement to get the IDs of all records. In contrast to that, the TRUNCATE statement is executed without preceding SELECT. In the described case of data deletion in MongoDB, the *drop()* method can be used. Although this action implies collection deletion, during new document input into the non-existent collection, the stated collection is automatically created and thus does not represent a significant resource cost for INSERT. The other option would be the use of the *remove()* operation.

Table 3 shows implemented suggestions for specific statement rewriting within the identified use cases for working with multiple records at once. The presented recommendations are specified in the syntax of three components of a prototype hybrid database (Oracle/MongoDB/Cassandra), purposely built to test the new architecture. The list of recommendations cannot be treated as final. Table 3 presents the rules implemented so far for the syntax optimisation of the entered statements using the rewriting technique, which is automatically performed by the SRC component of the new architecture. Besides that, the scenario of using different SQL DBMS as the SQL component of a hybrid requires additional effort to specify and implement recommendations syntactically adapted to the chosen DBMS. For instance, although the TRUNCATE command is supported by other

popular SQL DBMS systems (such as MS SQL Server, PostgreSQL, MySQL, etc.), we are aware that this is not the case with all the recommendations given.

Another example is that the MS SQL Server does not support the syntax of BULK/ BATCH INSERT, unlike Oracle and Cassandra. Instead, it allows specifying values of the multiple new records in parentheses after the VALUES clause. Similarly, although PostgreSQL does not explicitly support the BULK/BATCH UPDATE syntax, it adds a FROM clause to the UPDATE statement to achieve the same effect.

It is important to note that the supported statement rewriting techniques cannot always apply. The focus is on statements whose execution affects "multiple" records (for insert, update or delete statements). Therefore, the term *recommendation* is on the activity diagram. At the same time, the recommendation represents the crucial part of each rewriting rule, as shown in Table 4.

Whether the recommendation will be applied depends on the statement type and fulfilment of the specific preconditions. Despite all syntax preconditions fulfilling (the number of statements, the existence of particular clauses and similar), sometimes additional conditions must be met. For example, in the selected SQL component (Oracle) for applying PARALLEL onto the UPDATE statement (same for INSERT or DELETE), it is necessary to enable parallel execution of the DML statements on the system or session level by running the command (*alter system/session enable parallel dml*). Preconditions such as this can affect the application outcome of the given recommendation (similar to how statistical data of statement execution and calculated cost of accessing the data in alternative ways affect the index application in the query). Therefore, specific rules come down to the recommendation, and it is impossible to give generic enough and for the execution engine an utterly binding way of applying these recommendations. On the other hand, the absence of PARALLEL hint usage due to the unsupported parallel statement execution does not affect the success of the statement realisation. As in the case of stating the hint for inadequate and non-optimal indexes during query execution, the stated PARALLEL hint gets neglected, and the statement is executed without an error occurring due to the forwarded hint.

To a certain extent, decision-making is automated (for supported techniques) through the hybrid database dedicated architecture that supports the presented process model and new statement rewriting component. Regardless, statement optimisation is a complex process. It is often limited by adjustment options as well as the influence on the way the optimiser and execution engine of the DBMS work. The hybrid database extension with the newly created SRC component uses the optimisation advantages of particular DBMSs, with limitations within those DBMSs. The described condition is evident when individual DBMSs are observed, especially within hybrid databases. If the provided hint in the SQL database has a higher execution cost value than its alternatives, the execution engine will not use it. The same principle follows a hybrid database because the stated derives from the execution engine of the specific component.

At this time, creating a fully comprehensive solution for optimising all types of databases that a single hybrid can encompass is extremely challenging. It is not an easy task to determine the number of domain-specific languages used by all NoSQL databases currently available on the market. Even if there were an estimate, without the language standardisation by the leading NoSQL manufacturers, there would not be a comprehensive solution for implementing the rewriting techniques for the entire hybrid database.

Therefore, this paper aims to demonstrate the feasibility of extending the dedicated architecture for work with a hybrid database by introducing the newly developed SRC components that operate under the process model presented in this Section to improve performance.

4. Extending Hybrid SQL/NoSQL Database with Statement Rewriting Component (SRC)

The architecture presented in the paper [12] was taken as the starting point of the system that uses the hybrid database. This architecture enables integration and uniform use of all hybrid database components. It gives the user benefits of using different technologies, as well as the convenience of working with a unique logical database. The limiting factor of the initial architecture, which also represents the direction for further research, is the lack of support for the statement optimisation, more precisely, statement rewriting. This paper presents the extension of the initial architecture with the design of the SRC component. Figure 2 shows the extended SQL/NoSQL database architecture with the SRC component, which implements the principles of the newly introduced process model in Figure 1.

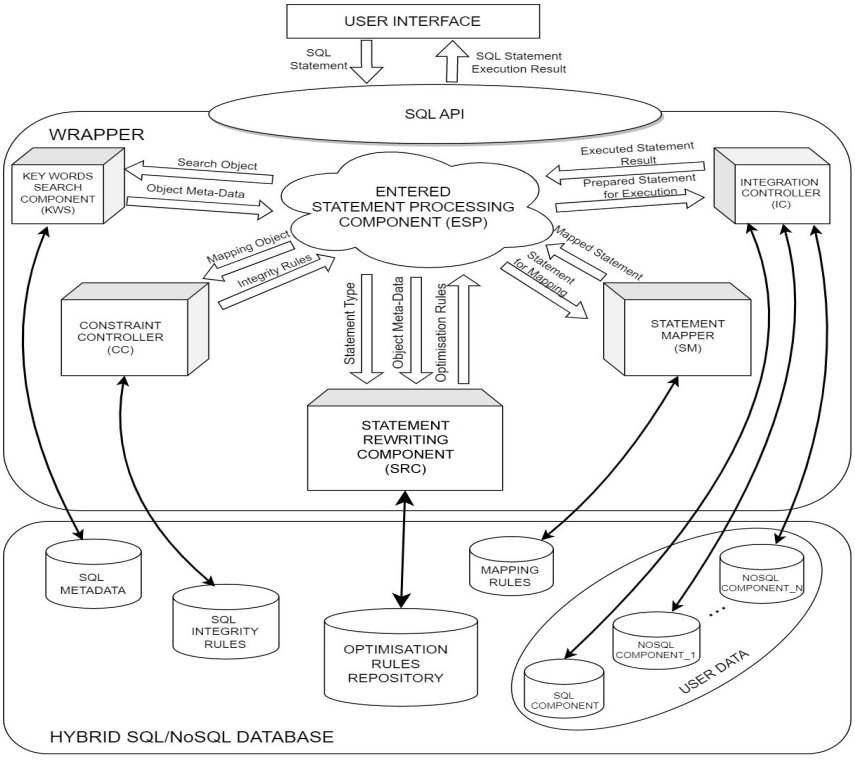


Fig. 2. The SQL/NoSQL database architecture extended with the newly developed component for the statement rewriting (SRC)

Pictured architecture represents the extension of the traditional three-tier architecture. Without going into the details of user interface organisation and implementation (graphical interface, statement input console and similar), the purpose of its presentation layer is to display data and to enable statement input and execution by the user. The middle layer is represented through the Wrapper, and it contains earlier established components necessary for providing support with the hybrid database. SQL language was chosen for the entire hybrid database and all types of databases that constitute its components, regardless of whether a particular database type supports SQL language. The motivation was to provide the comfort of using a single query language and to free the user from thinking about which component has requested data.

SQL API is in charge of communication to the presentation layer. The Wrapper manages all middle-layer components, including SQL API. Following the acceptance of the SQL statement, it is necessary to analyse, decompose, optionally map and forward the statement or its parts for execution to the hybrid destination components (i.e. specific DBMSs unified by the hybrid). This activity is in the jurisdiction of the Entered Statement Processing Component (ESP), the central communication component of the Wrapper. To achieve this, ESP communicates with other Wrapper components, sends them requests, processes their return values and manages the whole process from the reception of the SQL statement to its execution and display of the returned results to the users. From the function description, in the most general sense, the ESP component is most similar to the traditional three-tier architecture controller. The components ESP communicates with are (present in the initial architecture) Key Words Search Component (KWS), Constraint Controller (CC), Statement Mapper (SM), and Integration Controller (IC), as well as, in this paper introduced, the newly developed Statement Rewriting Component (SRC).

After analysing the SQL statement entered by a user, the ESP component communicates with the KWS component by sending the objects' names (read from the appropriate clause). It receives metadata about the objects as a response from the KWS component. The metadata contains the object type (table, column, key-value pair, etc.) and the database type the object belongs to (SQL, document-oriented NoSQL, column family NoSQL, etc.). It also contains the specific DBMS in which the object is implemented (Oracle, MongoDB, Cassandra, etc.). Here, it is necessary to highlight one of the essential characteristics of the hybrid database: the whole hybrid database, with all its components, represents the unique logical database. Given that all hybrid objects, regardless of what component they belong to, are integrated with the joint data model, a hybrid can't contain two objects with the same name but of a different type. Therefore, for every object name forwarded to the KWS, the ESP receives a single object type, a single database type it belongs to and one specific destination DBMS. Based on the return values, the ESP component will decide if it is necessary to execute statement mapping. When an entered statement or part of that statement has a destination in a database type that doesn't support SQL language, a mapping will occur. In contrast, when the entered statement and its integral parts have a destination in a database that supports SQL language, a mapping doesn't happen.

The next component the ESP addresses in the communication chain is the CC. The CC is in charge of centralised management of all hybrid components' integrity rules. Since SQL databases provide a higher degree of constraint control [70], it is the SQL databases' integrity rules (that encompass entity integrity rules and referential integrity

rules) that are implemented as common characteristics for the whole hybrid database. The lack of NoSQL support for the mentioned rules is overcome by the integrated placement of constraints in the dedicated hybrid's SQL component for storing integrity rules (named Integrity Rules). It contains the rules of entity integrity, i.e. it takes care of the primary key of each object (columns or fields, depending on the type of database) and its complexity (whether it contains one or more columns or fields). In addition to the entity rules, the Integrity Rules repository contains the referential integrity rules (i.e. foreign keys). It keeps data about referenced and referencing columns (or fields), regardless of which types of database objects participate in the referencing. This functionality overcomes the lack of integrity rules support in the NoSQL databases. It enables fluid referencing between the SQL and NoSQL database objects, as well as between different types of NoSQL objects.

Metadata, Integrity Rules, Mapping Rules and the newly added Optimisation Rules Repository represent dedicated SQL databases for metadata storage. These should not be confused with databases that are part of the hybrid database and contain user data. The reason for choosing the SQL databases for metadata storage is based on the strict ACID properties' support, which is imminent for the consistent use of metadata.

In the earlier version of the hybrid architecture, after obtaining the integrity rules from the CC component, the ESP component communicated with the SM component by sending the entered statement and metadata of its objects.

The extension of the improved architecture operating logic enables the communication of the ESP component with the newly developed SRC component. The earlier version didn't contain the SRC component. Instead, when needed and after receiving the integrity rules from the CC component, the ESP component sent the request to the SM component to perform statement mapping into the domain-specific language. In the extended architecture, the ESP component communicates with the newly developed SRC component before interacting with the SM component. The ESP component sends the statement type and metadata of destination objects to the SRC component. The SRC component accesses the newly introduced Optimisation Rules Repository. The Optimisation Rules Repository contains necessary preconditions, recommendations and application rules of specific optimisation techniques for particular statements (in this case, for statement rewriting).

Table 4 shows the structure of the Optimisation Rules Repository. This table contains selected examples of the statement rewriting rules for INSERT, UPDATE and DELETE of the hybrid Oracle/MongoDB/Cassandra database.

If the need for introducing new rules occurs, the new record will be added to the Optimisation Rules Repository. One record will be added for each statement type of each existing database. Additionally, if the hybrid database expends to the additional components (databases), new rules, in the form of new records in the Optimisation Rules Repository, will be added for each statement type of each new database. The stated repository, in addition to the specific optimisation recommendation (*Recommendation* column), contains a statement type (*Stat_type* column), a component type (*Comp_type* column) and columns with preconditions. Each precondition corresponds to a column of the same name. In Table 4, optimisation rules, in this case rewriting rules examples, have depicted preconditions in two columns (*Multi_rows* and *WHERE_clause* columns), and other rules can have additional preconditions (marked with '...'). Only essential columns for the chosen examples are shown in Table 4. Columns whose headings are preconditions' names represent a specific optimisation technique known at Oracle under Hard-coded values.

In the Hard-coded values column, the CHECK constraint defines the range of valid values. It contains the value *'YES'* for preconditions for which fulfilment is obligatory, while the value *'NO'* is for preconditions in which fulfilment must not be satisfied. The value *'YES/NO'* is for optional preconditions, i.e. that precondition doesn't influence the use of the particular rule, and the value *'N/A'* is for preconditions not applicable to the observed rule.

Table 4. Example of an Optimisation Rules Repository

Rule_ID	Stat_type	Comp_type	Multi_rows	WHERE_clause	...	Recommendation
101	INSERT	SQL	YES	N/A		INSERT_ALL
102	INSERT	NoSQL/MongoDB	YES	N/A		BULK_INSERT
103	INSERT	NoSQL/Cassandra	YES	N/A		BATCH_INSERT
...						
201	UPDATE	SQL	YES	YES/NO		PARALLEL
202	UPDATE	NoSQL/MongoDB	YES	YES/NO		UPDATE_MANY
203	UPDATE	NoSQL/Cassandra	YES	YES/NO		BATCH_UPDATE
...						
301	DELETE	SQL	YES	NO		TRUNCATE
302	DELETE	NoSQL/MongoDB	YES	NO		DROP
303	DELETE	NoSQL/Cassandra	YES	NO		TRUNCATE
...						

Selected examples from the table will be described. For instance, for the specific *INSERT_ALL* rule to be applied, the precondition of inserting multiple rows must be fulfilled (column *Multi_rows* has a *'YES'* value). In contrast, the prerequisite *WHERE_clause* does not apply to this rule (*WHERE_clause* has the value *'N/A'*) because the INSERT syntax does not support the WHERE clause. Similarly, the same precondition needs to be fulfilled (*Multi_rows*) for the insert of multiple rows into the MongoDB component and Cassandra component, while, once again, *WHERE_clause* is not applicable.

TRUNCATE, DROP and *TRUNCATE* are respective rewriting rules for the SQL, MongoDB and Cassandra *DELETE* statements for deleting all rows without filtering the records. That is why, for the observed *DELETE* rules, column *Multi_rows* has a *'YES'* value, and *WHERE_clause* has a *'NO'* value.

The rewriting rules for the *UPDATE* statement of Oracle, MongoDB and Cassandra components of the observed Oracle/MongoDB/Cassandra hybrid are *PARALLEL, UPDATE_MANY,* and *BATCH_UPDATE*, respectively. All three mentioned rules are applicable for multiple rows updates (column *Multi_rows* has a *'YES'* value) and can be applied regardless of whether all records are updated or just filtered data (column *WHERE_clause* has a *'YES/NO'* value).

Based on forwarded metadata, the SRC component determines if it can perform statement rewriting and how. The SRC component reads applicable optimisation rules of the input statement from the Optimisation Rules Repository and forwards them to the ESP component. After receiving the return values from the SRC component, the statement execution flow of the new hybrid with SRC is equivalent to the process in the earlier architecture version. The ESP component sends statements for mapping to the SM com-

ponent. By reading the rules from the Mapping Rules repository, it maps statements into the domain-specific languages and returns them to the ESP. The ESP component sends mapped statements to the IC component. The IC component has the role of managing and controlling the execution of the statements in one or more components. The IC sends the statement execution results and feedback to the ESP component. The ESP component then adjusts the results format to be user-friendly and forwards it to the presentation layer, i.e. to the appropriate user interface. The described activities encompass the whole process from the SQL statement input, analysis, decomposition, rewriting, optional mapping, statement execution in the hybrid's destination component and user notification.

5. The Use of the Hybrid Database with SRC on Oracle/MongoDB/Cassandra Example

The data model, Figure 3, was developed to demonstrate the usage of the current version of the hybrid SQL/NoSQL database with the new SRC component. The UML Class diagram depicts the created model. The hybrid Oracle/MongoDB/Cassandra database selected for testing implements the shown model.

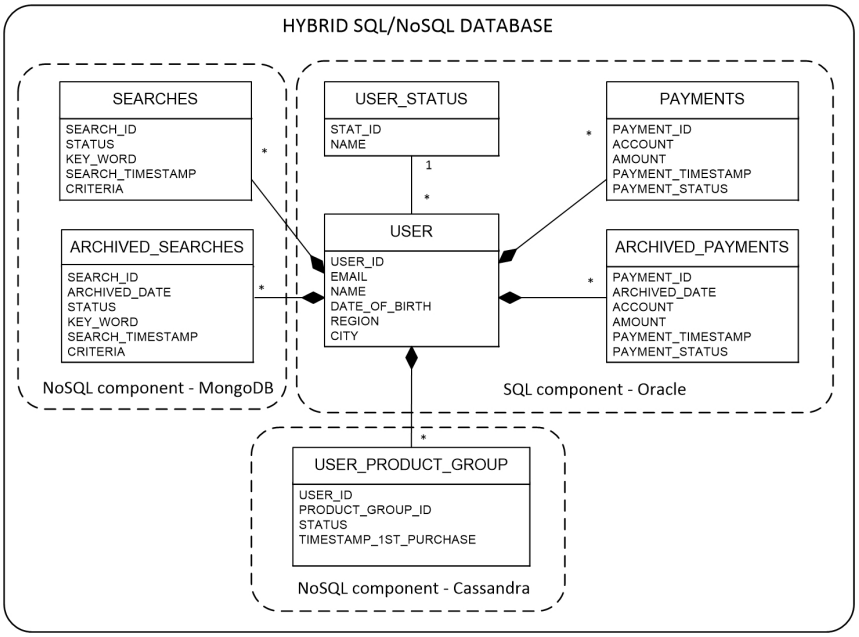


Fig. 3. UML Class diagram for tested domain

The domain of the model is online search and product payment. Figure 3 represents only a part of the model which was needed to realise the use cases chosen for testing (for example, the ordering process was not necessary to show). The model consists of the following classes: *User, User_status, Payments, Searches, Archived_payments,*

Archived_searches, and *User_product_group*, and it is implemented in two versions of the system: the previous hybrid Oracle/MongoDB/Cassandra database without SRC and the current hybrid Oracle/MongoDB/Cassandra database with SRC.

Table 5. Tested use cases for INSERT

Use case id	Statement type and hybrid component	Statement description	Statement before rewriting	Statement after rewriting
UC_1	INSERT into SQL component	INSERT rows into *archived_payments*	INSERT INTO archived_payments VALUES... ... INSERT INTO archived_payments VALUES...	INSERT ALL INTO archived_payments... ... INTO archived_payments...
UC_2	INSERT into NoSQL component (MongoDB)	Insert rows into *archived_searches*	db. archived_searches.insertOne(...) ... db. archived_searches.insertOne(...)	var bulk = db.archived_searches.initializeOrderedBulkOp(); bulk.insert(...); ... bulk.insert(...); bulk.execute();
UC_3	INSERT into NoSQL component (Cassandra)	Insert rows into *user_product_group*	INSERT INTO user_product_group (...) VALUES ... INSERT INTO user_product_group (...) VALUES ...	BEGIN BATCH INSERT INTO user_product_group(...) VALUES ... ... INSERT INTO user_product_group(...) VALUES ... APPLY BATCH,

The unconditional consistency of sensitive data, information about users, their statuses, and payments requires storing them in the SQL component of the hybrid. For users' searches, availability and fast reporting have a higher level of importance than the necessary consistency, so the mentioned part of the system (*Searches*) is implemented in the NoSQL component of the hybrid system (precisely in the MongoDB component). To demonstrate the functioning of the test hybrid SQL/NoSQL database with more than one NoSQL component, an additional Cassandra component was introduced. Cassandra component implements the table *User_product_group*, which contains data of the products group (searched products, bought products, etc.) in correlation with a particular user.

Table 6. Tested use cases for UPDATE

Use case id	Statement type and hybrid component	Statement description	Statement before rewriting	Statement after rewriting
UC_4	UPDATE SQL component	Users with the *status_id = 1* update to *status_id = 2*	UPDATE user SET status_id = 2 WHERE status_id = 1	UPDATE /*+ PARALLEL(4)*/ user SET status_id = 2 WHERE status_id = 1
UC_5	UPDATE NoSQL component (MongoDB)	All searches with the status "Accepted" update to values "Done"	db.searches.updateOne ({Status: "Accepted"}, {$set: {Status: "Done"}}) ... db.searches.updateOne ({Status: "Accepted"}, {$set: {Status: "Done"}})	db.searches.updateMany ({Status:"Accepted"}, {$set: {Status: "Done"}})
UC_6	UPDATE NoSQL component (Cassandra)	All products' status of the user with *id = 5* updates to value 'Searched'	SELECT distinct product_group_id FROM user_product_group where user_id = 5 ... UPDATE user_product_group SET status = 'Searched' WHERE user_id = 5 and product_group_id=... ... UPDATE user_product_group SET status = 'Searched' WHERE user_id = 5 and product_group_id=...	SELECT distinct product_group_id FROM user_product_group where user_id = 5; BEGIN BATCH UPDATE user_product_group SET status= 'Searched' WHERE user_id = 5 and product_group_id=... ... UPDATE user_product_group SET status= 'Searched' WHERE user_id = 5 and product_group_id = ... APPLY BATCH;

Payments and *Searches* are identifiable and existentially dependent on the entity *User*. In the diagram, they make a possessive Composition relationship with the *User* class. *Archived_payments* is a table inside the SQL component, while *Archived_searches* represent documents in the MongoDB component of the hybrid database. A large amount of data is cyclical, in certain time intervals, being input into *Archived_payments* and *Archived_searches* from *Payments* and *Searches*, respectively. This is how two use cases are profiled, one for data insertion (usually several dozens of thousands of records) into the SQL component (table *Archived_payments*) and the other one for the entry of, once again, a large amount of data into the NoSQL component (*Archived_searches* structure). After realising the mentioned use cases, records are deleted from *Payments* and *Searches*, which represent use cases for deleting all records from the SQL (Oracle) and NoSQL (MongoDB) components, respectively. For the SQL and NoSQL (MongoDB) compo-

nents' update, the chosen use cases were the user status change (foreign key) in the *User* table (the SQL component) as well as the performed searches update (the NoSQL component). Three use cases represent insert, update and delete in the Cassandra component as well. Table 5, Table 6 and Table 7 show the snapshot of use cases chosen for testing based on the hybrid Oracle/MongoDB/Cassandra components.

For every use case, Table 5, Table 6 and Table 7 display the use case ID, a statement type, a destination component of the hybrid, a statement description and a statement syntax before and after applying the supported rules. The statement rewriting rules, shown in Table 4 and described in Section 4, were applied to nine chosen use cases. All nine selected use cases were executed over the previous version of the hybrid architecture without the SRC component and, after that, over the extended version of the hybrid architecture, which contains the SRC component.

Tests were carried out on the PC with an Intel i7 CPU, with a 2.9 GHz speed, 16 GB RAM and SSD hard disk. The testing system has Windows OS, Oracle DBMS version 19c for the SQL component and MongoDB version 3.6 for the NoSQL component of the hybrid. The average statement execution time was taken as the performance indicator.

NetBeans IDE was used for statement inputs, executions and time measurements. Each test had 12 iterations. In order to eliminate the outliners, tests with the shortest and the longest execution times for each statement were discarded. The average time contains the execution times of the remaining ten iterations.

Table 7. Tested use cases for DELETE

Use case id	Statement type and hybrid component	Statement description	Statement before rewriting	Statement after rewriting
UC_7	DELETE from SQL component	Delete all rows from *payments*	DELETE FROM payments	TRUNCATE TABLE payments
UC_8	DELETE from NoSQL component (MongoDB)	Delete all rows from *searches*	db.searches.deleteMany()	db.searches.drop()
UC_9	DELETE from NoSQL component (Cassandra)	Delete all rows from *user_product_group*	SELECT distinct user_id FROM user_product_group... DELETE FROM user_product_group WHERE user_id IN ...	TRUNCATE user_product_group

Measurements of use cases UC_1, UC_2, UC_3, UC_7, UC_8 and UC_9 were conducted on datasets of 5,000, 10,000, 30,000, 50,000, 75,000 and 100,000 records. The INSERT and DELETE use cases for SQL (UC_1 and UC_7), NoSQL – MongoDB (UC_2 and UC_8) and NoSQL – Cassandra (UC_3 and UC_9) components were performed over the same amount of records. Use cases UC_4, UC_5 and UC_6 were tested over 100,000,

300,000, 500,000, 750,000 and 1,000,000 records. Increasing the dataset relative to the remaining statements was chosen due to the nature of the described model. Since archived tables are periodically filled (INSERT), and operational tables are emptied (DELETE), a larger amount of records was selected for UPDATE to cover the amount of data that can be moved in several cycles. What follows is the display and the analysis of the average execution times of the use cases chosen for testing, focusing on the execution times before and after the statement rewriting rules usage.

6. Experimental Results

The average measured execution times of the use cases chosen for testing are shown in Figure 4, Figure 5 and Figure 6. The X-axis of the diagrams shows the number of records affected by the particular statement execution. The Y-axis shows the average statement execution time in seconds. In addition, every chart has three parts. The first part of the diagram, marked as (a), shows the duration of the statement execution in the SQL (Oracle) hybrid component before and after optimisation, more precisely, statement rewriting. The second part of the diagram, marked as (b), shows the average duration of the observed statement execution in the NoSQL (MongoDB) hybrid component before and after statement rewriting, while the third part, labelled as (c), represents the average execution time in the second NoSQL component (Cassandra), also before and after statement rewriting. Before applying statement rewriting, the use case is executed in a hybrid database without SRC (previous architecture), and the after statement rewriting presents the statement execution in a hybrid database with SRC (extended architecture). Although we acknowledge that the statement rewriting represents one of many optimisation techniques, for the easiness of presenting and analysing the following results, by “before/after optimisation”, we will mean “before/after applying particular statement rewriting rule”.

Figure 4 shows the average execution time of the INSERT statement. In the SQL (Oracle) component, the average execution times of the INSERT statement before optimisation were 14.836 (5,000 records), 25.543 (10,000 records), 61.547 (30,000 records), 83.721 (50,000 records), 119.236 (75,000 records) and 196.008 (100,000 records) seconds. Following the optimisation, INSERT in the Oracle component lasted on average 3.018 (5,000 records), 7.123 (10,000 records), 18.856 (30,000 records), 28.641 (50,000 records), 43.378 (75,000 records) and 71.23 (100,000 records) seconds. The average times for data insertion into the MongoDB component before optimisation were 1.414 (5,000 records), 3.655 (10,000 records), 6.295 (30,000 records), 8.527 (50,000 records), 11.885 (75,000 records) and 15.697 (100,000 records) seconds. However, after optimisation, it took 0.529 (5,000 records), 0.601 (10,000 records), 0.735 (30,000 records), 1.147 (50,000 records), 1.473 (75,000 records) and 1.739 (100,000 records) seconds. In the Cassandra component, the values before applying the optimisation recommendation were 5.75 (5,000 records), 7.729 (10,000 records), 14.815 (30,000 records), 22.281 (50,000 records), 30.937 (75,000 records) and 39.995 (100,000 records) seconds. After SRC executed statement rewriting, the average execution times were significantly decreased to 0.869 (5,000 records), 1.257 (10,000 records), 1.746 (30,000 records), 2.345 (50,000 records), 2.981 (75,000 records) and 3.374 (100,000 records) seconds.

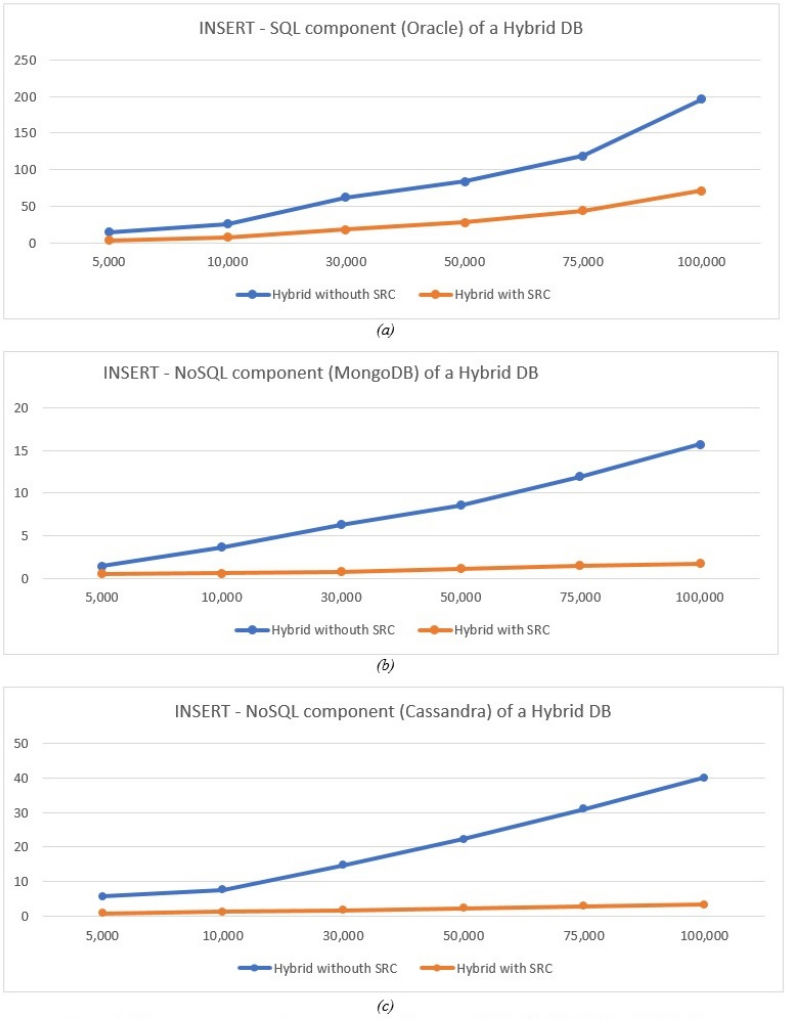


Fig. 4. The average measured execution times of use cases UC_1 (a), UC_2 (b) and UC_3 (c)

The optimised (rewritten) INSERT statement uses INSERT ALL and BULK INSERT for the SQL and NoSQL components, respectively, and achieves shorter execution times over a hybrid without SRC, as expected. In addition, the average data insertion time in the NoSQL component is noticeably shorter than in the SQL component on a comparable number of records. The explanation is that records in the SQL components have a strict schema structure and additional constraints to satisfy. Cassandra is representative of the wide-column NoSQL databases. As shown by the achieved results, concerning the schema structure strictness and consistency, Cassandra is between MongoDB and Oracle but closer to Oracle. That can be concluded by achieving a noticeably higher aver-

age execution time than MongoDB, especially for the non-optimised statements. On the other hand, Cassandra still manages shorter execution times than Oracle, benefiting from the wide-column principle of storing data. Cassandra achieved a shorter pre-optimised execution time (on 100,000 records than Oracle on 30,000), which became even more emphasized with the rewritten statements (quicker on 100,000 records than Oracle on 10,000).

It is important to point out that the INSERT statement with SQL optimisation recommendations has undergone a slight syntax adjustment. The INSERT ALL statement is generally created by concatenating the INTO table_name clause to the one INSERT statement. That way, multiple uses of the INSERT statement are eliminated. Although this approach has a fixed cost in concatenating the INTO clauses before executing the statement, the concatenation itself does not require much time. However, the limitations of this technique are the significant increase in the number of statement characters, which was visible in the average execution time even with only 1,000 inserted records. Not to discredit the mentioned rule through numerous characters concatenation, the syntax has been adapted by dividing the INSERT ALL into 1,000 records chunks. An INSERT ALL was performed every 1,000 records in as many iterations as needed to insert all records. Because of that, the mentioned fixed cost of statement concatenation was multiplied. However, the average execution time over a larger amount of data highlighted the benefits of executing one INSERT ALL over 1,000 records.

Figure 5 shows the average UC_4, UC_5 and UC_6 use cases' execution times. The UPDATE statement before optimisation in the destination SQL component averaged the following durations: 0.981 (100,000 records), 2.679 (300,000 records), 4.849 (500,000 records), 8.032 (750,000 records) and 10.835 (1,000,000 records), expressed in seconds. In the hybrid with SRC, the performance was as follows: 0.843 (100,000 records), 2.156 (300,000 records), 2.766 (500,000 records), 3.513 (750,000 records) and 4.519 (1,000,000 records). The limitation of applying the PARALLEL hint is the inability to guarantee its usage. However, tested UC_4 was using the PARALLEL hint. Even though the slight advantage of using PARALLEL was detected even on 100,000 records, the benefit of the parallel record update was more noticeable on 500,000 records. With the increase in the number of records, the average execution time has decreased compared to the non-optimised execution. This time saving occurred as the consequence of the parallel, instead of sequential, statement execution. Before optimisation, UC_5, on the hybrid without SRC, was executed in 18.649 (100,000 records), 65.005 (300,000 records), 94.751 (500,000 records), 142.677 (750,000 records) and 188.009 (1,000,000 records) seconds. Following the optimisation, UC_5 achieved drastically decreased average execution times. The average times for UC_5 after optimisation are 1.473 (100,000 records), 3.381 (300,000 records), 5.054 (500,000 records), 7.907 (750,000 records) and 10.035 (1,000,000 records) in seconds. By far, the greatest absolute and relative savings in average execution time was achieved by the optimised UC_5. The reason for that is the powerful *updateMany()* mechanism, which significantly comes to the fore in contrast to a million executions of the *updateOne()* method.

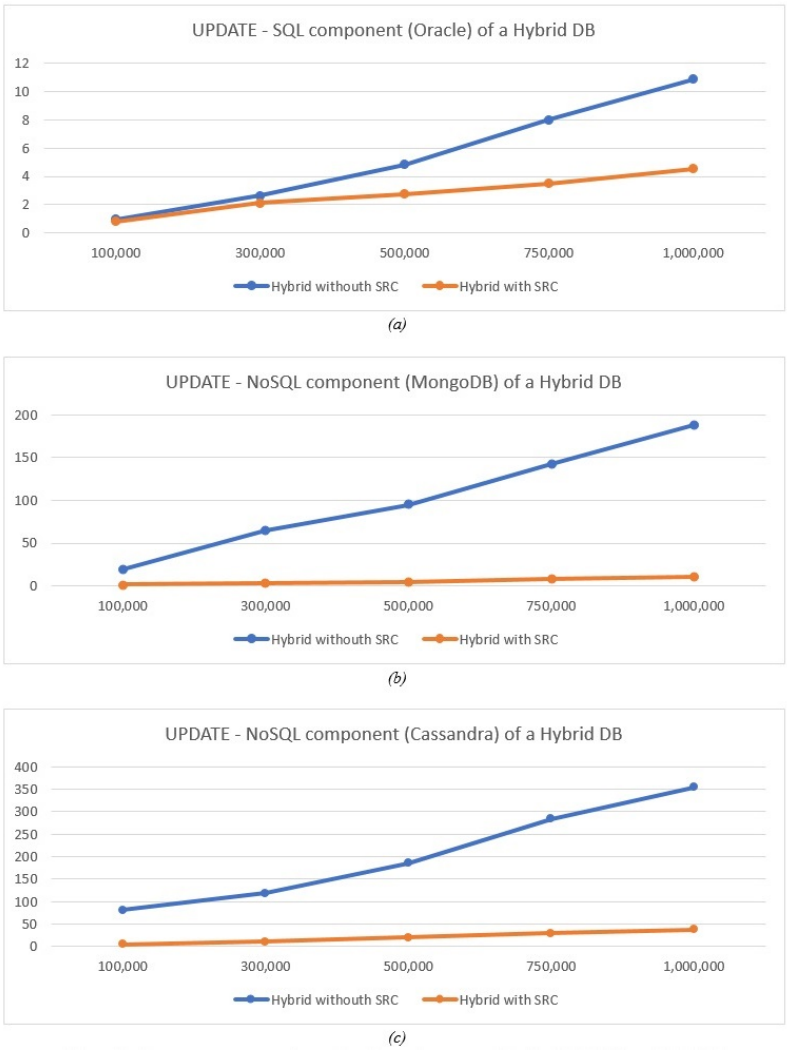


Fig. 5. The average measured execution times of use cases UC_4 (a), UC_5 (b) and UC_6 (c)

After the optimisation, a significant decrease in average execution time also occurred within the Cassandra component. UPDATE statement in the Cassandra component, inside the architecture without SRC, achieved 80.766 (100,000 records), 118.905 (300,000 records), 185.432 (500,000 records), 284.124 (750,000 records) and 354.174 (1,000,000 records) seconds. In comparison, the rewritten statement in Cassandra inside the new architecture with SRC achieved 4.343 (100,000 records), 11.172 (300,000 records), 19.781 (500,000 records), 29.7 (750,000 records) and 37.211 (1,000,000 records) seconds. Because in the tested version of Cassandra and its driver, BATCH UPDATE was successfully

executed with no more than 300,000, four calls of this syntax (for 1,000,000 records) were executed. Still, they were noticeably quicker than the non-optimised multiple UPDATE, which has an obligatory WHERE clause with all primary key fields and without the support of IN.

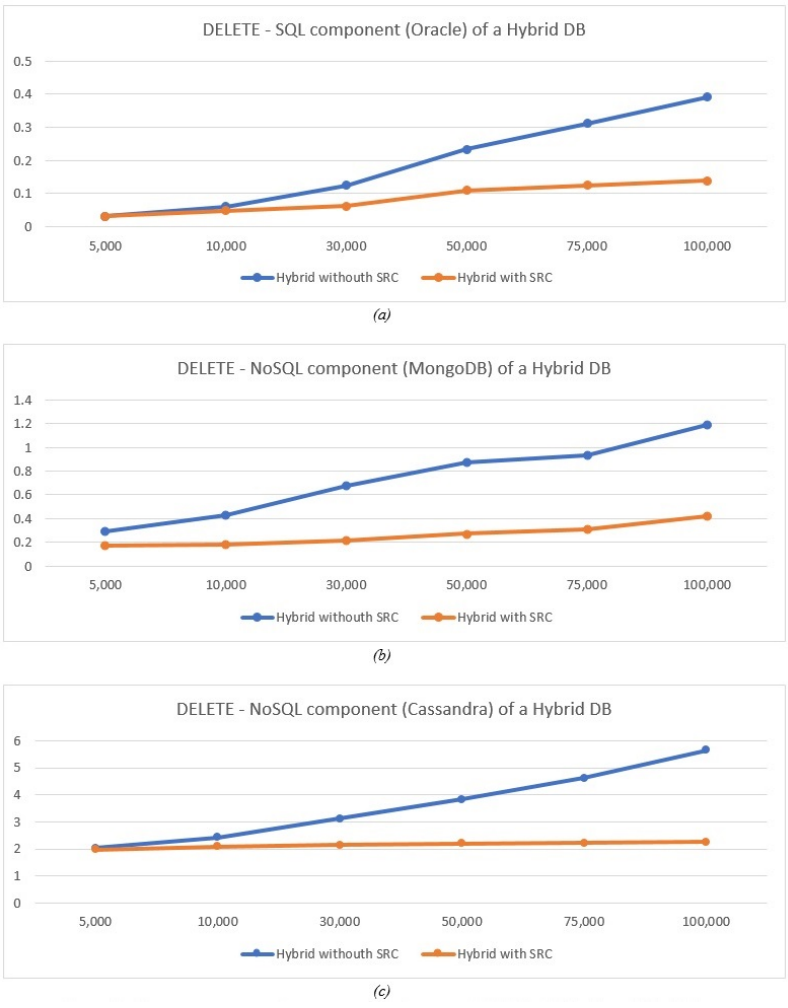


Fig. 6. The average measured execution times of use cases UC_7 (a), UC_8 (b) and UC_9 (c)

Figure 6 depicts the optimisation effects of UC_7, UC_8 and UC_9 on the average statement execution times. The deletion of all records in the SQL component was carried out in 0.031 (5,000 records), 0.061 (10,000 records), 0.125 (30,000 records), 0.234 (50,000 records), 0.312 (75,000 records) and 0.391 (100,000 records) seconds. The

optimised statement has achieved 0.031 (5,000 records), 0.047 (10,000 records), 0.062 (30,000 records), 0.109 (50,000 records), 0.125 (75,000 records) and 0.138 (100,000 records) seconds. The identical average time of record deletion, before and after optimisation, over the dataset of 5,000 records showed the efficiency of the DELETE statement when the WHERE clause was not forwarded. However, with the increase in the number of records for deletion, especially over 30,000 records and more, the advantage of using the DDL statement, which does not go through individual rows, came into the spotlight.

The *deleteMany()* operation, which already represents an improvement over the basic *deleteOne()* method, needed 0.291 (5,000 records), 0.428 (10,000 records), 0.677 (30,000 records), 0.874 (50,000 records), 0.929 (75,000 records) and 1.192 (100,000 records) seconds and it represents the UC_8 performance before optimisation. The optimised UC_8 took, on average, 0.176 (5,000 records), 0.181 (10,000 records), 0.219 (30,000 records), 0.271 (50,000 records), 0.309 (75,000 records) and 0.421 (100,000 records) seconds. Even though less drastically, the optimised *drop()* over the MongoDB component also led to performance improvement, expressed through the average execution times.

Using the TRUNCATE statement in the Cassandra component decreased the average execution time in the architecture with the SRC. Old architecture without SRC achieved average of 2.032 (5,000 records), 2.433 (10,000 records), 3.109 (30,000 records), 3.823 (50,000 records), 4.616 (75,000 records) and 5.656 (100,000 records) seconds, while the new one, with SRC, achieved 1.965 (5,000 records), 2.081 (10,000 records), 2.137 (30,000 records), 2.191 (50,000 records), 2.225 (75,000 records) and 2.25 (100,000 records) seconds. Although the recommendation for the DELETE statement wasn't as dominant as the BATCH technique for INSERT or DELETE, it still managed time decrease. It is noticeable that with the smaller datasets (for example, 5,000 records), there is nothing to separate DELETE FROM and TRUNCATE. Still, with the increased dataset volume, a slight advantage is on the rewritten statement side.

7. Conclusions and Future Work

The findings presented in this paper represent the continuation of the hybrid SQL/NoSQL databases research. The motivation was to give answers to the research questions which emerged during the previous phase of research. The main goal was to explore the feasibility and justification of extending the hybrid SQL/NoSQL database by creating new Statement Rewriting Component and Optimisation Rules Repository components and integrating them into the well-proven hybrid's architecture. As support for applying rewriting techniques, a process model (in the UML notation) was developed.

Without striving to cover all possible optimisation rules for all types of databases, which would be an almost impossible task at the moment, selected statement rewriting rules were chosen for INSERT, UPDATE and DELETE statements. Test use cases demonstrate the average statement duration over the hybrid database with and without SRC. In some use cases, over certain records, a hybrid database with the SRC component didn't necessarily achieve a shorter execution time (i.e. UC_7 over 5,000 records). However, the observed trend was that the hybrid database with the SRC component required less time for execution when the number of records was increased compared to the hybrid without SRC.

The conclusion that arises is that with the smaller number of records, a hybrid with SRC cannot always necessarily achieve a decrease in the average execution time in comparison to the hybrid without SRC. However, when working with a larger amount of data (several tens or hundreds of thousands of records), experimental tests indicated a significant decrease in the average execution times for the selected use cases of the presented domain. These results were gathered on a particular Oracle/MongoDB/Cassandra hybrid, which was chosen for simulating execution in the architecture with and without the SRC component. However, we acknowledge that these results present one instance of outcomes and that the architecture extended with the SRC is still in the prototype phase. Nevertheless, the results showed the purposes of the introduced extension and the expected performance-gaining trend of using a hybrid with SRC.

There are several identified directions of future work to overcome the limitations of the current version of hybrid SQL/NoSQL databases. The first one is to implement additional functionalities into the current prototype architecture. That will enable expansion of for-now supported basic optimisation techniques for SELECT and their combining with other DML statements (i.e. SELECT with complex subquery, but also UPDATE and DELETE with subquery). The current version doesn't support DDL statements, and incorporating these functionalities would enable users to easily create and alter all types of database objects, no matter what component of a hybrid would store it, instead of just manipulating with data in the existing objects. An important direction of future research and advancing the presented approach would be expanding the number of different components inside the hybrid while introducing particular implementations for the other subtypes of databases (for example, key-value and graph) and the syntax support for other SQL DBMS (for example, PostgreSQL, MS SQL Server etc.) as well as other NoSQL systems. Additional enhancement of the present architecture could include broadening optimisation techniques in the Optimisation Rules Repository because the number of statement optimisation rules is not final and can be extended. In the end, expanding testing Oracle/MongoDB/Cassandra hybrid database to a more complex system with multiple components of many other types of databases is planned for the future.

References

1. C. A. Lana, M. Guessi, P. O. Antonino, D. Rombach, and E. Y. Nakagawa. A systematic identification of formal and semi-formal languages and techniques for software-intensive systems-of-systems requirements modeling. *IEEE Systems Journal*, 13(3):2201–2212, 2019.
2. V. de Oliveira Neves, A. Bertolino, G. De Angelis, and L. Garcés. Do we need new strategies for testing systems-of-systems? In *Proceedings of the SESoS'18: SESoS'18:IEEE/ACM 6th International Workshop on Software Engineering for Systems-of-Systems*, pages 29–32, New York, NY, USA, 2018. ACM.
3. A. Bertolino and R. Mirandola. Software performance engineering of component-based systems. In *Proceedings of the Fourth International Workshop on Software and Performance, WOSP 2004*, pages 238–24, Redwood Shores, California, USA, 2004. Association for Computing Machinery, NY, United States.
4. A. Bertolino, G. De Angelis, and F. Lonetti. Governing regression testing in systems of systems. In *Proceedings of 2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pages 144–148, Berlin, Germany, 2019. IEEE.

5. S. Park, Y. Shin, S. Hyun, and D. Bae. Simva-sos: Simulation-based verification and analysis for system-of-systems. In *Proceedings of the 15th International Conference of System of Systems Engineering (SoSE)*, pages 575–580, Budapest, Hungary, 2020. IEEE.
6. M.A. Olivero, A. Bertolino, F.J. Dominguez-Mayo, M.J. Escalona, and I. Matteucci. Addressing security properties in systems of systems: Challenges and ideas. In R. Calinescu and F. Di Giandomenico, editors, *Software Engineering for Resilient Systems - SERENE 2019*, volume 11732 of *Lecture Notes in Computer Science*, pages 138–146. Springer, Cham, 2019.
7. H. Cadavid, V. Andrikopoulos, and P. Avgeriou. Improving hardware/software interface management in systems of systems through documentation as code. *Empirical Software Engineering*, 28, 2023.
8. ANSI. Ansi/x3 /sparc dbms framework. Report of the Study Group on Database Management Systems, 1977.
9. B. Lazarević, Z. Marjanović, N. Aničić, and S. Babarogić. *Baze podataka*. FON, Belgrade, Serbia, 2006.
10. A. Borgida, M. Casanova, and A. H. F. Laender. Logical database design: from conceptual to logical schema. In L. LIU and T. ÖZSU, editors, *Encyclopedia of Database Systems*, pages 1645–1649. Springer, Boston, MA, US, 2009.
11. S. Bjeladinovic. A fresh approach for hybrid sql/nosql database design based on data structuredness. *Enterprise Information Systems*, 12(8-9):1202–1220, 2018.
12. S. Bjeladinovic, Z. Marjanovic, and S. Babarogic. A proposal of architecture for integration and uniform use of hybrid sql/nosql database components. *Journal of Systems and Software*, 168:110633, 2020.
13. H.R. Vyawahare, P.P. Karde, and V.M. Thakare. A hybrid database approach using graph and relational database. In *Proceedings of the 2018 IEEE International Conference on Research in Intelligent and Computing in Engineering*, pages 2555—-2564, Univ Don Bosco, San Salvador, EL SALVADOR, 2018. IEEE.
14. SolidIT. Db-engines ranking. Web site: DB-engines ranking, 2024. [Online]. Available on: https://db-engines.com/en/ranking (Retrieved: January 2024).
15. K. Sudhakar. Difference between sql and nosql databases. *International Journal of Management, IT and Engineering*, 8(6):444–452, 2018.
16. A. Faraj, B. Rashid, and T. Shareef. Comparative study of relational and nonrelations database performances using oracle and mongodb systems. *International Journal of Computer Engineering Technology (IJCET)*, 5(11):11–22, 2014.
17. A. Vágner. How do nosql databases handle variety of big data? In XS. Yang, S. Sherratt, and Joshi A. Dey, N., editors, *Proceedings of Ninth International Congress on Information and Communication Technology ICICT 2024*, volume 1012 of *Lecture Notes in Networks and Systems*, pages 459–469. Springer, Singapore, 2024.
18. L. Zhang, K. Pang, J. Xu, and B. Niu. Json-based control model for sql and nosql data conversion in hybrid cloud database. *Journal of Cloud Computing*, 11(23), 2022.
19. S. Goyal, P.P. Srivastava, and A. Kumar. An overview of hybrid databases. In *Proceedings of the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 285–288, Greater Noida, India, 2015.
20. C. Gyorodi, R. Gyorodi, and R. Sotoc. A comparative study of relational and nonrelational database models in a web-based application. *International Journal of Advanced Computer Science and Applications*, 6(10):78–83, 2015.
21. B. James and P.O. Asagba. Hybrid database system for big data storage and management. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, 7(3/4):15–27, 2017.
22. N. Jatana, S. Puri, M. Ahuja, I. Kathuria, and D Gosain. A survey and comparison of relational and non-relational database. *International Journal of Engineering Research Technology*, 1(6):1–5, 2012.

23. M. Villari, A. Celesti, M. Giacobbe, and M. Fazio. Enriched e-r model to design hybrid database for big data solutions. In *Proceedings of the 2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 163–166, Messina, Italy, 2016. IEEE.
24. D. Martinez-Mosquera, R. Navarrete, and S. Lujan-Mora. Modeling and management big data in databases—a systematic literature review. *Sustainability*, 12(2):634, 2020.
25. J. Duggan, A. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, and S. Zdonik. The bigdawg polystore system. *ACM SIGMOD Record*, 44(2):11–16, 2015.
26. E. Kharlamov, T. Mailis, K. Bereta, D. Bilidas, S. Brandt, E. Jimenez-Ruiz, S. Lamparter, C. Neuenstadt, O. Özçep, A. Soylu, C. Svingos, G. Xiao, D. Zheleznyakov, D. Calvanese, I. Horrocks, M. Giese, Y. Ioannidis, Y. Kotidis, R. Moller, and A. Waaler. A semantic approach to polystores. In *Proceedings of the 2016 IEEE International Conference on Big Data*, pages 2565–2573, Washington, DC, USA, 2016. IEEE.
27. S. Dasgupta, K. Coakley, and A. Gupta. Analytics-driven data ingestion and derivation in the awesome polystore. In *Proceedinsg of the 2016 IEEE International Conference on Big Data*, pages 2555–2564, Washington, DC, USA, 2016. IEEE.
28. A. Maccioni, E. Basili, and R. Torlone. Quepa: Querying and exploring a polystore by augmentation. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2133–2136, San Francisco, California, USA, 2016. Sigmod.
29. J. McHugh, P.E. Cuddihy, J.W. Williams, K.S. Aggour, V.S. Kumar, and V. Mulwad. Integrated access to big data polystores through a knowledge-driven framework. In *Proceedings of the 2017 IEEE International Conference on Big Data*, pages 1494–1503, Boston, MA, USA, 2017. IEEE.
30. F. Basciani, J. Di Rocco, L. Iovino, and A. Pierantonio. Typhonml: Tool support for hybrid polystor. *Science of Computer Programming*, 232:103044, 2023.
31. N. Niu, L. D. Xu, and Z. Bi. Enterprise information systems architecture - analysis and evaluation. *IEEE Transactions On Industrial Informatics*, 9(4):2147–2154, 2013.
32. O. Lajam and S. Mohammed. Revisiting polyglot persistence: From principles to practice. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 13(5):872–882, 2022.
33. E. Płuciennik and K. Zgorzałek. The multi-model databases – a review. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, and D. Kostrzewa, editors, *Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation. BDAS 2017.*, volume 716 of *Communications in Computer and Information Science*, pages 141–152. Springer, Cham, 2017.
34. J. Lu and I. Holubová. Multi-model databases. *ACM Computing Surveys*, 52(3):1–38, 2019.
35. J. Lu, I. Holubová, and B. Cautis. Multi-model databases and tightly integrated polystores. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2301–2302, New York, NY, USA, 2018. Association for Computing Machinery.
36. F. Ye, X. Sheng, N. Nedjah, J. Sun, and P. Zhang. A benchmark for performance evaluation of a multi-model database vs. polyglot persistence. *Journal of Database Management*, 34(3):1–20, 2023.
37. D. Van Landuyt, J. Benaouda, V. Reniers, A. Rafique, and W. Joosen. A comparative performance evaluation of multi-model nosql databases and polyglot persistence. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 286—-293, New York, NY, USA, 2023. Association for Computing Machinery.
38. I. Holubová, M. Vavrek, and S. Scherzinger. Evolution management in multi-model databases. *Data Knowledge Engineering*, 136:101932, 2021.
39. Van Landuyt D. Rafique A. Joosen W. Reniers, V. Object to nosql database mappers (ondm): A systematic survey and comparison of frameworks. *Information Systems*, 85:1–20, 2019.

40. N. Roy-Hubara and A. Sturm. Design methods for the new database era: a systematic literature review. *Software and Systems Modeling*, 19:297–312, 2020.
41. A. Kalayda. Promising directions for the development of modern databases. *Journal of Physics: Conference Series*, 2131(022087):1–6, 2021.
42. B. Bender, C. Bertheau, T. Kŏrppen, H. Lauppe, and N Gronau. A proposal for future data organisation in enterprise systems—an analysis of established database approaches. *Information Systems and e-Business Management*, 20:441—494, 2022.
43. I. Zečević, P. Bjeljac, B. Perišić, S. Stankovski, D. Venus, and G. Ostojić. Model driven development of hybrid databases using lightweight metamodel extensions. *Enterprise Information Systems*, 12(8-9):1221–1238, 2018.
44. H.N. Aleem, M.M. Baig, and M.M. Khan. Efficient software testing technique based on hybrid database approach. *International Journal of Advanced Computer Science and Applications*, 10(7):349—-356, 2019.
45. H.R. Vyawahare, P.P. Karde, and V.M. Thakare. Hybrid database model for efficient performance. *Procedia Computer Science*, 152(8-9):172–178, 2019.
46. A. de la Vega, D. García-Saiz, C. Blanco, M. Zorrilla, and P. Sánchez. Mortadelo: A model-driven framework for nosql database design. In E. Abdelwahed, L. Bellatreche, M. Golfarelli, D. Méry, and C. Ordonez, editors, *Model and Data Engineering (MEDI 2018)*, volume 11163 of *Lecture Notes in Computer Science*, pages 41–57. Springer, Cham, 2018.
47. M. Sokolova, F. Gómezb, and L. Borisoglebskayaa. Migration from an sql to a hybrid sql/nosql data model. *Journal of Management Analytics*, 7(1):1–11, 2019.
48. F. Abdelhedi, A.A. Brahim, F. Atigui, and G. Zurfluh. Logical unified modeling for nosql databases. In *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017)*, pages 249–256, Porto, Portugal, 2017. HAL Science.
49. K. Mershad and A. Hamieh. Sdms: smart database management system for accessing heterogeneous databases. *International Journal of Intelligent Information and Database Systems*, 14(2):115–152, 2021.
50. L. Nikolic, V. Dimitrieski, and M. Celikovic. An approach for supporting transparent acid transactions over heterogeneous data stores in microservice architectures. *Computer Science and Information Systems*, 21(1):167—-202, 2024.
51. O. Mehdi, H. Ibrahim, S. Affendey, E. Pardede, and J. Cao. Exploring instances for matching heterogeneous database schemas utilizing google similarity and regular expression. *Computer Science and Information Systems*, 15(2):295–320, 2018.
52. R. Čerešňák and M. Kvet. Comparison of query performance in relational a non-relation databases. *Transportation Research Procedia*, 40:170–177, 2019.
53. K. Fraczek and M. Plechawska-Wojcik. Comparative analysis of relational and non-relational databases in the context of performance in web applications. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, and D. Kostrzewa, editors, *Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation. BDAS 2017.*, volume 716 of *Communications in Computer and Information Science*, pages 153–164. Springer, Cham, 2017.
54. Z.H. Liu, B. Hammerschmidt, D. McMahon, Y. Liu, and H.J. Chang. Closing the functional and performance gap between sql and nosql. In *Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16)*, pages 227–238, San Francisco, USA, 2016. Sigmod.
55. S. Bjeladinović, M. Škembarević, O. Jejić, and M. Asanović. An analysis of using binary json versus native json on the example of oracle dbms. *IPSI Transactions on Internet Research*, 19(2):92–103, 2023.
56. A. Kemper and T. Neumann. One size fits all, again! the architecture of the hybrid oltpolap database management system hyper. In M. Castellanos, U. Dayal, and V. Markl, editors, *Enabling Real-Time Business Intelligence (BIRTE 2010)*, volume 84 of *Lecture Notes in Business Information Processing*, pages 7–23. Springer, Berlin, Heidelberg, 2011.

57. L. Thiry, H. Zhao, and M. Hassenforder. Categories for (big) data models and optimisation. *Journal of Big Data*, 5(21), 2018.
58. B. Scheuermann. Design of a reconfigurable hybrid database system. In *Proceedings of the 18th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines*, pages 247–250, Charlotte, NC, USA, 2010. IEEE.
59. M. Owaida, D. Sidler, K. Kara, and G. Alonso. Centaur: A framework for hybrid cpu-fpga databases. In *Proceedings of the 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM 2017)*, pages 211–218, Napa, CA, USA, 2017. IEEE.
60. S. Breß, E. Schallehn, and I. Geist. Towards optimization of hybrid CPU/GPU query plans in database systems. In M. Pechenizkiy and M. Wojciechowski, editors, *Advances in Intelligent Systems and Computing*, volume 185 of *Advances in intelligent systems and computing*, pages 27–35. Springer Berlin Heidelberg, 2013.
61. S. Cremer, M. Bagein, S. Mahmoudi, and P. Manneback. Improving performances of an embedded relational database management system with a hybrid cpu/gpu processing engine. In C. Francalanci and M. Helfert, editors, *Data Management Technologies and Applications. DATA 2016*, volume 737 of *Communications in Computer and Information Science*, pages 160–177. Springer, Cham, 2017.
62. M. Gowanlock, B. Karsin, Z. Fink, and J. Wright. Accelerating the unacceleratable: Hybrid cpu/gpu algorithms for memory-bound database. In *Proceedings of the 15th International Workshop on Data Management on New HardwareJuly (DaMoN'19)*, pages 1–11, Amsterdam Netherlands, 2019. ACM.
63. Z. Pang, S. Wu, H. Huang, Z. Hong, and Y. Xie. Aqua+: Query optimisation for hybrid database-mapreduce system. *Knowledge and Information Systems*, 63:905—938, 2021.
64. W. Khan, W. Ahmad, B. Luo, and E. Ahmed. Sql database with physical database tuning technique and nosql graph database comparisons. In *Proceedings of the 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2019)*, pages 110–116, Chengdu, China, 2019. IEEE.
65. R. Sellami and B. Defude. Complex queries optimisation and evaluation over relational and nosql data stores in cloud environments. *IEEE Transactions on Big Data*, 4(2):217–230, 2018.
66. C. Li and J. Gu. An integration approach of hybrid databases based on sql in cloud computing environment. *Software: Practice and Experience*, 49(3):11–16, 2018.
67. SolidIT. Db-engines ranking – relational dbms. Web site: DB-engines ranking, 2024. [Online]. Available on: https://db-engines.com/en/ranking/relational+dbms (Retrieved: January 2024).
68. SolidIT. Db-engines ranking – document store dbms. Web site: DB-engines ranking, 2024. [Online]. Available on: https://db-engines.com/en/ranking/document+store (Retrieved: January 2024).
69. SolidIT. Db-engines ranking – wide-column store dbms. Web site: DB-engines ranking, 2024. [Online]. Available on: https://db-engines.com/en/article/Wide+Column+Stores (Retrieved: December 2024).
70. C. Nance, T. Losser, R. Iype, and G. Harmon. Nosql vs rdbms - why there is room for both. In *Proceedings of the Southern Association for Information Systems Conference*, pages 111–116, 2013.

Srđa Bjeladinović is an Assistant Professor at Faculty of Organizational Sciences, University of Belgrade. He received his M.Sc. and Ph.D. degrees in Information Systems from University of Belgrade. His research interests are databases, information systems development methodologies, integrated software solutions and ERPs. In recent years he has been researching NoSQL and hybrid databases.

Received: October 24, 2024; Accepted: February 20, 2025.

Computer Science and Information Systems 22(3):1047–1060 https://doi.org/10.2298/CSIS240905047Y

Analyzing the Operational Efficiency of Online Shopping Platforms Integrated with AI-Powered Intelligent Warehouses

Wang Yuqin[1], Chin-Shyang Shyu[2], Cheng-Sheng Lin[3], Chao-Chien Chen[4], Thing-Yuan Chang[5], and Liang Shan[6,⋆]

[1] Wuhan Business University School of Tourism Management, China
295751380@qq.com
[2] Department of Recreation & Sport Management, Shu-Te University, Taiwan
shaw@stu.edu.tw
[3] Department of Agricultural Technology, National Formosa University, Taiwan
sheng8876@nfu.edu.tw
[4] Department of Leisure and Recreation Management, Asia University, Taiwan
peter72@asia.edu.tw
[5] Department of Information Management, National Chin-Yi University of Technology, Taiwan
c07408@ncut.edu.tw
[6] Wuhan Business University School of Tourism Management, China
46112164@qq.com

Abstract. The application and development of AI intelligent technology is the key to solving the current shortage of manpower in online malls and industries, and also improving the overall efficiency of online shopping Platforms operations. Modern people's living and consumption habits have changed, and online shopping focuses on delivery speed, accuracy and delivery quality. Therefore, Intelligent warehousing management plays an important role in the operation of online shopping Platforms. This study mainly uses the network DEA model to analyze the operating efficiency of online shopping platforms Integrated with AI-Powered Intelligent Warehouses. The online shopping platforms operation and management is divided into the AI intelligent warehousing stage and the sales department stage, and based on the operating efficiency weights of the two stages, the total efficiency of the online shopping platforms is measured by a weighted average. According to the empirical analysis results, the AI-Powered intelligent warehousing efficiency weights of all online shopping platforms are higher than the sales efficiency weights, indicating that the key to the overall efficiency of online shopping platforms is the AI-Powered intelligent warehousing operation and management efficiency. Although D1 (Shopee Taiwan) has the highest sales efficiency value and annual transaction volume, its efficiency value in the AI-Powered intelligent warehousing stage is lower than D3, resulting in the total operating efficiency being lower than D3. Therefore, it can be seen that the operating efficiency of the AI-Powered intelligent warehousing department is currently The key to the success or failure of online shopping platforms operations.

Keywords: AI-Powered intelligent warehousing, network DEA model.

⋆ Corresponding author

1. Introduction

The rapid advancement of the Internet and technology has shifted global shopping habits, including those in Taiwan. Currently, online shopping has become a widely preferred shopping alternative [9], [19], [7]. For instance, according to a 2022 survey by the Market Intelligence & Consulting Institute (MIC), there was a notable increase in online spending among consumers in Taiwan. Statistically, 43.4% of respondents reported an annual online shopping expenditure between NTD 10,000 and 60,000, marking a 4.1% increase. Furthermore, 18% of respondents spent over NTD 60,000 in 2022, representing a 4.9% rise. Of this group, 9.1% spent between NTD 60,000 and NTD 100,000, while 8.9% were heavy spenders with expenditures exceeding NTD 100,000. These statistics underscore the growing acceptance of online shopping, driven by the widespread use of the Internet and mobile devices. This transition has prompted traditional brick-and-mortar shop operators to undergo digital transformation, shifting their focus from physical stores to their official websites, online shopping platforms, or e-commerce to meet consumer demands. This sentiment is supported by [4], who note the rapid growth of online sales businesses utilizing e-commerce. Additionally, the COVID-19 pandemic, which began in 2020, boosted the online shopping trend. Concerns about virus transmission prompted consumers to minimize in-person shopping trips at physical stores, further fueling the growth of online shopping platforms. As uncertainty persists amid the novel COVID-19, consumers have become more cautious about shopping in public. Moreover, the widespread use of today's technology makes consumers' preference for online shopping predictable [5]. Despite advancements, online shopping platforms still encounter challenges. [31] highlighted one such challenge: the increasingly low availability of warehouse workers for warehouse operations, impacting the picking process. Moreover, [33] stressed the importance of a comprehensive system for online sales, encompassing online communication, logistics management, and customer service. In the context of low-price competition, [33] cautioned that any misstep in the process could lead to a subpar customer experience. [34] also highlighted the pivotal role of warehouses in e-commerce distribution, noting the widespread adoption of warehouse-distribution integration for efficiency. Estimating full-link delivery time remains a key concern in this integrated model, affirming the significance of warehouses in logistics operations.

With the evolution of Internet technology and artificial intelligence (AI), intelligent warehouses, also known as a smart warehouse, are gaining traction. An intelligent warehouse utilizes advanced technologies to improve efficiency, effectiveness, and overall operations. [2] asserted that these facilities aim to enhance overall service quality, productivity, and efficiency while minimizing costs and failures. [6] identified order-picking as a major time-consuming warehouse process, accounting for over 55% of warehouse operational costs. Consequently, efficient warehouse management through technologies has emerged as a significant topic. [23] similarly advocated for enhanced warehousing process management, aligning with the trend of informatization and lean material warehouse management. Hence, exploring the operational efficiency of online shopping platforms, particularly in the context of intelligent warehouses, warrants attention.

Prior literature suggested that existing studies on e-commerce or the operational efficiency of online shopping platforms predominantly focused on inputs and final outputs [14], [13], [28]. For instance, [14] conducted an in-depth analysis using a two-stage methodology to examine the specific influence of competition on the operational effi-

ciency of garment companies (with B2C mode) operating in an e-commerce environment. Their findings revealed that competition significantly increased pure technical efficiency but did not positively impact scale efficiency. However, a limited number of studies have analyzed the impacts of automated picking, packaging, and transportation on the operational efficiency of online shopping platforms during the AI-powered intelligent warehouse phase, where inputs translate to outputs. Thus, this study aims to assess the performance of online shopping platforms in Taiwan in terms of operational efficiency during the AI-powered intelligent warehouse phase, market efficiency, and overall operational efficiency from 2018 to 2023. The outcomes of this investigation can provide a foundation for developing strategies for substantial improvement.

2. Literature Review

2.1. Online shopping platform

The advancement of Internet technology has unlocked new business opportunities for companies to capitalize on new opportunities for profit growth. E-commerce, stemming from the rapid evolution of Internet technology, empowers enterprises to explore expanded business opportunities unrestricted by temporal or spatial limitations. E-commerce is at the forefront of transforming marketing strategies, based on new technologies, and facilitates product information and improved decision-making [26]. Facilitated by shifting consumer behaviors, online shopping platforms have emerged, granting consumers access to various products or service offerings. Consumers can browse products, place orders, make payments, authorize deductions, and communicate online with the platform's customer service. [15] defined e-commerce from four different perspectives:

1. From a communications perspective, e-commerce encompasses delivering goods, services, information, or payments via computer networks or other electronic means.
2. From a business process perspective, e-commerce denotes the application of technology to automate business transactions and workflows.
3. From a service perspective, e-commerce serves as a tool addressing the desire of organizations, consumers, and management to reduce service costs, enhance product quality, and increase the speed of service delivery.
4. From an online perspective, e-commerce provides the capability of buying and selling products and information on the Internet and other digital services.
   The classification above underscores the intrinsic connection between e-commerce and consumers.

Consequently, the majority of subsequent studies on online shopping platforms revolved around the topic: What motivates consumers to engage in online shopping [30], [27], [11], [32], aiming to identify the factors that influence consumers' online shopping behavior for further improvement. However, online shopping entails several drawbacks, as outlined below:

1. Customers are unable to physically inspect the products before purchase.
2. Delivery times may be excessively prolonged at times.
3. Additional shipping charges may inflate the product's overall cost.
4. Sellers often lack personalized attention, increasing the risk of fraud.

5. Concerns arise regarding the security of online banking and credit card passwords.
6. Issues regarding product quality may arise [20].

These disadvantages highlight the pivotal role of warehouses in the operations of online shopping platforms. As sales on online shopping platforms surge, the transition to e-commerce warehouses becomes inevitable due to their advantages, including automated and intelligent operations. Beyond enhancing logistics efficiency and quality, e-commerce warehouses can minimize error rates and time costs, thereby delivering superior customer experiences and efficient delivery services.

2.2. Intelligent Warehouse

For online shopping platforms, warehouse management encompasses the systematic control of the flow, distribution, packaging, picking, storage, and classified processing of goods within the warehouse. Warehouses play an important role in the industry supply chain, as well as the production of any industrial unit. The whole flow of a company runs smoothly with respect to its warehouse, as it is used to store, manage, and track the goods [17]. To enhance efficiency, resources and procedures should be fully leveraged. As technology advances, warehouse management has shifted from traditional to automatic warehousing. Subsequently, the advent of Industry 4.0 has paved the way for intelligent warehouses, facilitated by advancements in the Internet of Things (IoT), AI, cloud computing, and big data analytics. Integrating software and hardware, along with sensors, automated drivers, and IoT, intelligent warehouses are expected to reduce labor costs and enhance accuracy and traceability. As outlined by [24], a smart warehouse embodies the future of intelligent warehousing, fostering flexible, reconfigurable, and agile warehousing environments by automating warehousing activities and running diagnostics for equipment repair and improvement, facilitated by seamless communication among computer systems, mobile devices, machinery, automated guided vehicles (AGVs), and equipment on the warehouse floor. Key components of a smart warehouse include (1) cyber-physical systems, (2) cloud computing services, (3) Internet of Things (IoT), (4) automated control platforms, (5) warehouse management systems (WMS), and (6) collaborative robots ("cobots"). In other words, intelligent warehouses leverage diverse sensors, radio-frequency identification (RFID), and network technology, alongside automated logistics equipment, to ensure automatic data scraping, identification, early warning, and management in inbound and outbound processes, thereby substantially reducing warehousing costs, optimizing efficiency, and enhancing intelligent management. For instance, the AI-powered automated storage/retrieval system developed by the [8] anticipates demands by stocking popular products in advance using AI-driven data computation based on factors such as season, weather, festivals, and regions. This helps enterprises seize business opportunities proactively. Furthermore, once orders are placed, the system optimizes batch picking and sequencing based on order relevance models, ensuring timely collection and shipping for each order, thus fulfilling tasks precisely. In today's landscape, where consumers demand rapid and accurate order delivery, intelligent warehouses enhance accuracy and traceability and minimize order errors and omissions, ensuring timely product delivery. Additionally, they significantly reduce labor costs by facilitating rapid and accurate order fulfillment. Despite their advantages, intelligent warehouses face certain development constraints, as outlined by [10]:

1. The cost of building a smart warehouse is more expensive than the traditional warehouses.
2. The transition to a smart warehouse is time-consuming and requires a lot of effort.
3. The transition to a smart warehouse requires the support of top management.

Therefore, comprehensively assessing the impact of intelligent warehouse integration on the overall operational efficiency of online shopping platforms is another key focus of this study.

2.3. Relevant Studies

Relevant literature reveals that studies on intelligent warehouses mainly focus on the application and development of smart warehouse management systems across different countries [18], [29], [12], [21], [35]. For instance, [18] proposed an Internet-of-Things (IoT)-based architecture for real-time warehouse management, dividing the warehouse into multiple domains. Architecture viewpoints were employed to present models based on context diagrams, functional views, and operational views tailored to stakeholder needs. Furthermore, certain studies explore new technologies capable of multi-tasking and providing economic benefits [16], [3], [25], [22]. For example, [3] explored the use of drones for various applications in smart warehouse management. Some studies delved into the impact of IoT on warehouse management, discussing both its advantages and disadvantages. Despite the broad spectrum of studies explored above, few studies specifically examined the impact of intelligent warehouses as inputs on enterprise operational efficiency. As [10] highlighted, the cost of constructing a smart warehouse is a crucial consideration. For online shopping platforms striving for operational success, optimizing the inputs and outputs during the operational phase of intelligent warehouses indirectly influences the market efficiency of these platforms in the second phase and their overall operational efficiency. Therefore, further exploration in this regard is warranted.

3. Research Method

3.1. Establishment of a model structure

As modern high-tech automation industries progress, the majority of prior studies on the operational efficiency of online shopping platforms primarily focused on the relative efficiency of inputs (e.g., labor, capital, and other variable inputs) translating into final outputs (e.g., revenue). However, they overlooked the impact of automated picking, packaging, and transportation on the operational efficiency of online shopping platforms during the AI-powered intelligent warehouse phase, where inputs translate into final outputs. Due to input and output process differences between online shopping platforms, the corresponding automated management skills and resource allocation vary. In the operational phase of AI-powered intelligent warehouses, labor costs are reduced, and transportation durations are shortened. Furthermore, by leveraging AI technology, automation is achieved in transportation, warehousing, packaging, loading and unloading, flow, processing, distribution, and data services, thereby gathering relevant big data. AI-powered intelligent warehouses automatically optimize and rectify management based on big data, further enhancing the

operational efficiency of online shopping platforms. Optimizing inputs and outputs during the operational phase of these warehouses indirectly impacts the market efficiency of online shopping platforms in the second phase, thereby influencing the overall operational efficiency. Thus, online shopping platforms exhibit varying operational efficiencies across different operational phases, which, in turn, mutually influence one another. This underscores the need for further exploration in this area. This study employed the Dynamic Network Slacks-Based Measure (SBM) Data Envelopment Analysis (DEA) model proposed by [1] for empirical analysis. The main focus was on evaluating the operational efficiency of AI-powered intelligent warehouses utilized by online shopping platforms in Taiwan, assessing the market efficiency of these platforms, and gauging their overall operational efficiency. This assessment was based on the total profits generated by a two-stage dynamic model over multiple years. In this study, the operational benefits derived from AI-powered intelligent warehouses in the first phase served as the basis for formulating operational strategies and allocating resources for online shopping platforms in the second phase. This study primarily examined the performance of Taiwan's online shopping platforms in terms of the operational efficiency of AI-powered intelligent warehouses, market efficiency, and overall operational efficiency from 2018 to 2023.

3.2. Network DEA model for online shopping platforms

Assumption: An online shopping platform comprises three interconnected departments facilitated by an AI-powered intelligent warehouse operation (or intermediate goods). Each department is characterized by its own set of inputs and outputs. As depicted in Figure 1, Link1$-$ $>$2 refers to the utilization of a portion of Department 1's output as part of Department 2's input. Similarly, Link1$-$ $>$3 and Link2$-$ $>$3 carry similar implications, as explained earlier.

As depicted in the figure, the network DEA model addresses the online shopping platform's internal production translation, representing a secondary production activity featuring interconnection and mutual influence between departments.

In this study, the Decision Making Unit (DMU) symbolizes the online shopping platform in the input-output analysis. The efficiency of the online shopping platform during the AI-powered intelligent warehouse phase and the platform's market efficiency are denoted as θ_o^{A*} and θ_o^{B*}, respectively. The mathematical model is formulated as follows:

$$\begin{aligned}
&\theta_o^{A*} = max \frac{\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{d0}^{AB}}{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{i=1}^{i_s} v_i^s x_{io}^s}, \\
&\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{dj}^{AB} \leq \sum_{i=1}^{i_A} v_i^A x_{ij}^A + \sum_{i=1}^{i_s} v_i^s x_{ij}^s \forall j, U_{ij}^{\alpha} \leq L_{ij}^{\alpha} \forall j, \\
&\eta_d^{AB}, v_i^A, v_i^s \geq \varepsilon
\end{aligned} \tag{1}$$

In Equation (1), the denominator of the objective function signifies the inputs during the AI-powered intelligent warehouse phase, while the numerator represents the outputs.

$$\begin{aligned}
&\theta_o^{B*} = max \frac{\sum_{r=1}^{r_B} u_r^B y_{r0}^B}{\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} (1-\alpha_{io}) v_i^s x_{io}^s}, \\
&\sum_{r=1}^{r_B} u_r^B y_{rj}^B \leq \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{dj}^{AB} + \sum_{i=1}^{i_c} v_i^B x_{ij}^B + \sum_{i=1}^{i_s} (1-\alpha_{io}) v_i^s x_{ij}^s \forall j, \\
&U_{ij}^{\alpha} \leq \alpha_{ij} \leq L_{ij}^{\alpha} \forall j, \\
&\eta_d^{AB}, u_r^B, v_r^B, v_i^B, v_i^s \geq \varepsilon
\end{aligned} \tag{2}$$

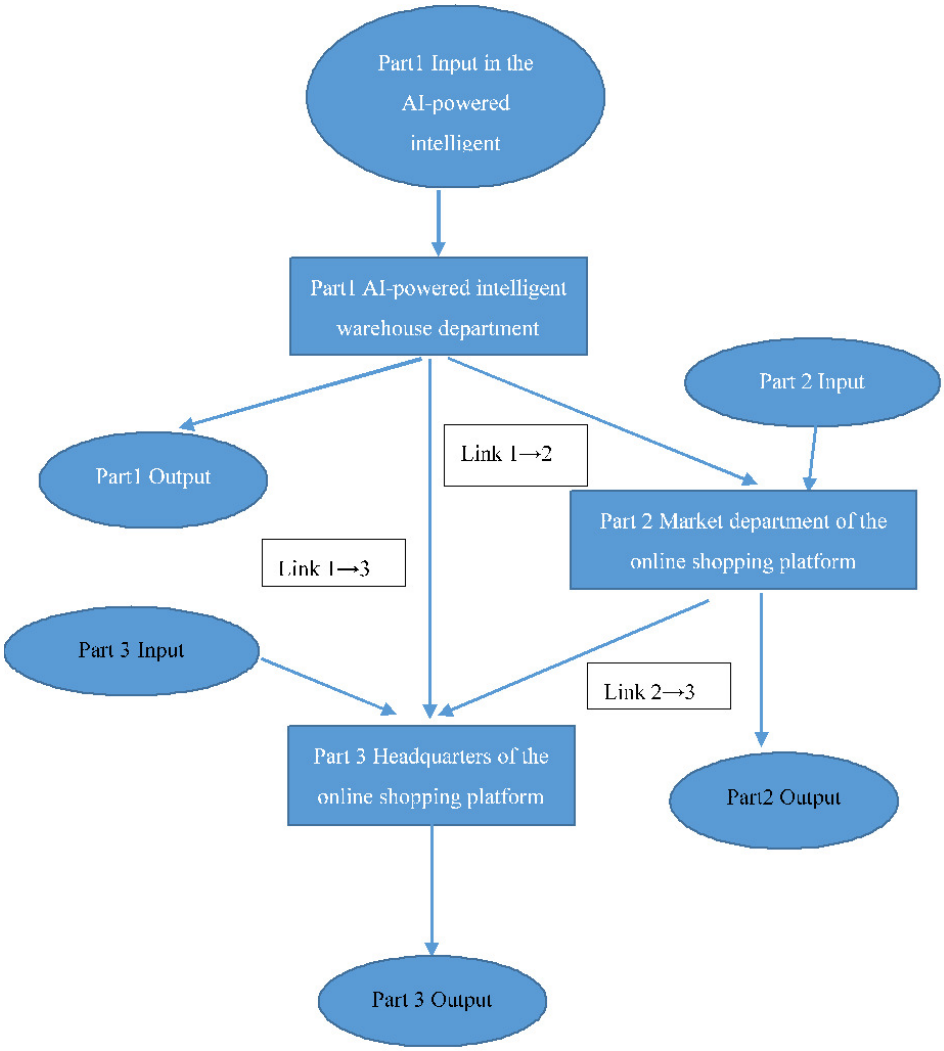


Fig. 1. Network DEA model of the online shopping platform

In Equation (2), the denominator of the objective function represents the inputs during the market operation phase of the online shopping platform, while the numerator signifies the outputs.

$$w_A = \frac{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{i=1}^{i_s} \alpha_{io} v_i^s x_{io}^s}{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} v_i^s x_{io}^s},$$
$$w_B = \frac{\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} (1-\alpha_{io}) v_i^s x_{io}^s}{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} v_i^s x_{io}^s} \tag{3}$$

As per the mathematical equations outlined above, concerning the weight values for the AI-powered intelligent warehouse phase of the online shopping platform and the market operation phase, assuming equal weight values are assigned to the intermediate goods

between these two phases, denoted as (Z1), (Z2), and (Z3): $\eta_1^A = \eta_1^B = \eta$; $\eta_2^A = \eta_2^B = \eta$; $\eta_3^A = \eta_3^B = \eta$. Additionally, it is assumed that the weights assigned to shared inputs are equal: $v_4^A = v_4^B = v_4$. Based on these assumptions, the total operational efficiency value (θ_o^{AB}) for the aforementioned two phases of the online shopping platform is the weighted aggregate of the efficiency (θ_o^A) during the AI-powered intelligent warehouse phase of the online shopping platform and the market efficiency (θ_o^B)), as shown in Equation (4):

$$\theta_o^{AB} = w_A\theta_o^A + w_B\theta_o^B = \frac{\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{d0}^{AB} + \sum_{r=1}^{r_B} u_r^B y_{r0}^B}{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} v_i^s x_{io}^s} \tag{4}$$

The mathematical model for the overall efficiency of the online shopping platform is presented in Equation (5):

$$\begin{aligned} &\theta_o^{ABC*} = max \frac{\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{d0}^{AB} + \sum_{r=1}^{r_B} u_r^B y_{r0}^B}{\sum_{i=1}^{i_A} v_i^A x_{io}^A + \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{do}^{AB} + \sum_{i=1}^{i_B} v_i^B x_{io}^B + \sum_{i=1}^{i_s} v_i^s x_{io}^s}, \\ &\sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{dj}^{AB} \leq \sum_{i=1}^{i_A} v_i^A x_{ij}^A + \sum_{i=1}^{i_s} v_i^s x_{ij}^s \forall j, \\ &\sum_{r=1}^{r_B} u_r^B y_{rj}^B \leq \sum_{d=1}^{d_{AB}} \eta_d^{AB} Z_{dj}^{AB} + \sum_{i=1}^{i_c} v_i^B x_{ij}^B + \sum_{i=1}^{i_s} (1-\alpha_{io}) v_i^s x_{ij}^s \forall j, \\ &U_{ij}^{\alpha} \leq \alpha_{ij} \leq L_{ij}^{\alpha} \forall j, \\ &\eta_d^{AB}, u_r^B, v_r^B, v_i^B, v_i^s \geq \varepsilon \end{aligned} \tag{5}$$

This study mainly analyzes the top seven online shopping platforms in Taiwan in terms of turnover for the year 2023, including D1: Shopee (Taiwan); D2: PChome eBay; D3: momo.com; D4: PChome 24h; D5: Books.com; D6: Yahoo!Kimo; and D7: Taiwan Rakuten Ichiba. Input and output variable data were sourced from the Global Information Network of the Ministry of Economic Affairs, R.O.C. and from the financial statements disclosed by these online shopping platforms. The data spans from 2021 to 2023.

This paper employed a network DEA model to define input and output variables, as outlined in Table 1. During the AI-powered intelligent warehouse phase, inputs include the warehouse area and AI automation costs, while the shipment volume serves as the output variable. During the second phase, the inputs of the sales department comprise the output (shipment volume) from the first phase, the number of employees, and operating expenses. The output variable during this phase is earnings per share (EPS), with the net profit acting as a carry-over from the previous years to analyze efficiency changes.

4. Network DEA Empirical Results

This study employed DEA-SOLVER Professional 16.0 to define the online shopping platform's production model as variable returns to scale. Additionally, a network DEA model was utilized to investigate the operational efficiency of online shopping platforms in Taiwan, specifically focusing on the operational efficiency of AI-powered intelligent warehouses within these platforms and their sales departments. Furthermore, this study assessed departments and online shopping platforms exhibiting low operational efficiency and proposed optimization strategies accordingly.

Based on the empirical findings presented in Table 2 concerning the efficiency during the AI-powered intelligent warehouse phase, D3, momo.com, reported the highest operational efficiency value, reaching 1.000. This indicates that momo.com has effectively

Table 1. Definition of input and output variables

	Variable	Variable definition description	Unit
Inputs during the AI-powered intelligent warehouse phase	Area of the intelligent warehouse	Floor area of the warehouse	Ping
	AI automation equipment inputs	RFID, light sensors, and infrared sensors	Million US
Intermediate output	Shipment volume	Annual average shipment volume of the intelligent warehouse	Pieces
Inputs during the online sales phase	Number of employees	Number of employees	People
	Operating expenses	Including water and electricity charges	Million US
Interdepartmental Link	Operating income	Net revenue after deducting sales returns and discounts	Million US
Final output	EPS	Net profit after tax number of common shares outstanding	US
Inter-period carry over	Profit	Presented in the financial statements Net profit after tax	Million US

leveraged the automation systems of its AI-powered intelligent warehouse to optimally allocate input-output resources. Moreover, D3 outperforms D1, Shopee, in terms of the operational efficiency of the intelligent warehouse, despite the latter's higher popularity and sales volume, suggesting potential for improvement in D1's AI-powered intelligent warehouse operations. Conversely, the lowest efficiency in the AI-powered intelligent warehouse operations phase was observed for D7, Taiwan Rakuten Ichiba, attributed to its comparatively smaller scale of automation equipment input and suboptimal input-output configuration, according to the descriptive statistics.

As AI technology continues to advance year by year, from 2021 to 2023, all seven online shopping platforms experienced growth in the operational efficiency of intelligent warehouses, highlighting their reliance and emphasis on AI-powered intelligent warehouses. Furthermore, AI-powered intelligent warehouses accelerate the pace and development of both the logistics and online shopping industries by integrating various functions such as transportation, distribution, and information services, thereby optimizing product distribution and logistics resource allocation. Additionally, by integrating with the logistics industry, AI-powered intelligent warehouses centralize dispersed product resources, harnessing overall and scale advantages to modernize, specialize, and complement traditional logistics enterprises, thus fostering the operational efficiency of AI-powered intelligent warehouses.

As social and economic landscapes evolve and consumer shopping behaviors change, the demand for online shopping increasingly prioritizes efficiency, accuracy, and speed. AI-powered intelligent warehouses, leveraging IT, integrate sensory systems into processes such as transportation, warehousing, packaging, loading and unloading, flow, pro-

Table 2. Analysis of the operational efficiency of online shopping platforms during the AI-powered intelligent warehouse phase between 2021 and 2023

DMU	2021	2022	2023	Efficiency value	Ranking
D1	1.000	1.000	0.903	0.986	2
D2	0.632	1.000	0.981	0.852	3
D3	1.000	1.000	1.000	1.000	1
D4	0.756	0.684	0.976	0.826	4
D5	0.561	0.651	0.792	0.702	6
D6	0.512	0.686	0.852	0.738	5
D7	0.486	0.628	0.725	0.638	7
Average value	0.703	0.822	0.896	0.815	NA
Maximum value	1.000	1.000	1.000	1.000	NA
Minimum value	0.486	0.628	0.725	0.638	NA
Standard deviation	0.012	0.026	0.01	0.020	NA

cessing, and distribution. Equipped with logistics and distribution systems that collect and analyze data, they process and optimize decisions based on the analysis results, thereby meeting online shoppers' cognitive and purchasing demands. Consequently, the operational efficiency of AI-powered intelligent warehouses in online shopping platforms significantly influences both the operational efficiency of sales departments of these platforms and consumers' willingness to shop online.

According to Table 3, D1 emerges as the leader in online sales efficiency among the selected online shopping platforms, with an operational efficiency value of 1.000. Analysis suggests that this success is attributed to the platform's marketing strategies, such as free shipment services, product discounts, premier service quality, convenience, rapid delivery, and efficient smart warehousing and logistics. Despite ranking second in the operational efficiency of AI-powered intelligent warehouses, D1 excels in offering online shoppers convenience, superior service quality, and discounts, thus securing its top position in sales efficiency. Conversely, the lowest efficiency of AI-powered intelligent warehouses and less appealing product discounts and service quality of D7 indirectly affect the sales market efficiency of the online shopping platform.

According to the above mathematical equations, the weighted average total efficiency values of the online shopping platforms between 2021 and 2023 are calculated in Table 4. The weight assigned to the AI-powered intelligent warehouse phase is observed to exceed that of the market sales phase for each online shopping platform, surpassing 0.5. This indicates that the overall operational efficiency of online shopping platforms is primarily influenced by the efficiency during the AI-powered intelligent warehouse phase, which is a pivotal factor behind the success of online shopping platforms. Furthermore, the outputs and shipment volume of AI-powered intelligent warehouses indirectly impact the orders and overall operational efficiency during the market sales phase.

According to the table, D1 has the highest average number of orders and popularity per year among online shopping platforms. However, its overall operational efficiency ranks second, primarily due to D1's lower weight and efficiency in the AI-powered intelligent warehouse phase compared to D3. Hence, it can be inferred that the operational efficiency

Table 3. Analysis of the operational efficiency of online shopping platforms during the online sales phase between 2021 and 2023

DMU	2021	2022	2023	Efficiency value	Ranking
D1	1.000	1.000	1.000	1.000	1
D2	0.986	1.000	0.978	0.962	3
D3	1.000	1.000	0.968	0.992	2
D4	0.852	0.882	0.932	0.886	4
D5	0.631	0.728	0.812	0.756	6
D6	0.587	0.802	0.822	0.761	5
D7	0.460	0.608	0.708	0.629	7
Average value	0.733	0.846	0.881	0.802	NA
Maximum value	1.000	1.000	1.000	1.000	NA
Minimum value	0.460	0.608	0.708	0.629	NA
Standard deviation	0.013	0.018	0.011	0.015	NA

of AI-powered intelligent warehouses is crucial in determining the overall efficiency of online shopping platforms.

Table 4. Weighted average overall efficiency of the online shopping platforms between 2021 and 2023

DMU	Efficiency of the AI-powered intelligent warehouse	Market sales efficiency	Weight for the AI-powered intelligent warehouse phase	Weight for the market sales phase	Overall efficiency	Ranking
D1	0.986	1.000	0.652	0.348	0.991	2
D2	0.852	0.962	0.621	0.379	0.894	3
D3	1.000	0.992	0.702	0.298	0.998	1
D4	0.826	0.886	0.616	0.384	0.849	4
D5	0.702	0.756	0.562	0.438	0.726	6
D6	0.738	0.761	0.526	0.474	0.749	5
D7	0.638	0.629	0.508	0.492	0.634	7
Average value	0.815	0.802	0.598	0.402	0.809	NA
Maximum value	1.000	1.000	0.702	0.492	0.998	NA
Minimum value	0.638	0.629	0.508	0.298	0.634	NA
Standard deviation	0.015	0.020	0.008	0.010	0.015	NA

5. Conclusions

Consumer habits, coupled with modern AI technology, IoT, and big data technology, have ushered in an innovative and intelligent high-tech business system. This system offers consumers convenient, secure, and rapid shopping experiences while enhancing operational efficiency and reducing the operating costs of online shopping platforms.

Since customers of modern online malls focus on transaction convenience and shortened delivery time, the empirical analysis results of this study show that online malls that incorporate AI-Powered intelligent warehouse management will have higher overall operating efficiency. The operation and management of AI-Powered intelligent warehousing increases the speed of picking, packaging, and shipping of traded goods, which not only saves delivery time, but also saves a lot of labor costs to improve online mall operating profits and overall operating efficiency.

In recent years, with the development of high technology, artificial intelligence (AI) and machine learning (ML) have grown rapidly, which has had a huge impact on industries such as online shopping platforms and retail. Industries that can combine the development of AI-Powered intelligent technology will be the key to competitiveness.

References

1. Dynamic dea with network structure: A slacks-based measure approach. Omega 42(1), 124–131 (2014)
2. Abdullah, H., Abdul Shukor, S., Raslie, H., Ngah, E., Ahmad Jamain, N.: A bibliometric analysis and future research directions on language and literacy. International Journal of Academic Research in Business and Social Sciences 13 (09 2023)
3. Ali, S., Khan, S., Fatma, N., Özel, C., Hussain, A.: Utilisation of drones in achieving various applications in smart warehouse management. Benchmarking: An International Journal 31 (04 2023)
4. Anardani, S., Azis, M., Asyhari, M.: The implementation of business intelligence to analyze sales trends in the indofishing online store using power bi. Brilliance: Research of Artificial Intelligence 3, 300–305 (12 2023)
5. and, H.M.A.H.: Determinants of intention to continue usage of online shopping under a pandemic: Covid-19. Cogent Business & Management 8(1), 1936368 (2021)
6. Bottani, E., Montanari, R., Rinaldi, M., Vignali, G.: Intelligent Algorithms for Warehouse Management, pp. 645–667. Springer International Publishing, Cham (2015)
7. Chaudhary, S.: Effect of e-commerce on organization sustainability. IOSR Journal of Business and Management 19, 15–24 (07 2017)
8. Fenercioğlu, A., Soyaslan, M., Közkurt, C.: Automatic storage and retrieval system (as/rs) based on cartesian robot for liquid food industry (01 2011)
9. Fitriani, S., Valentika, N.: Developing marketplace-based online store as an adaptation to online purchase trends. AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis 3, 96–104 (05 2023)
10. van Geest, M., Tekinerdogan, B., Catal, C.: Smart warehouses: Rationale, challenges and solution directions. Applied Sciences 12(1) (2022)
11. Groß, M.: Mobile shopping: A classification framework and literature review. International Journal of Retail & Distribution Management 43, 221–241 (03 2015)
12. Hamdy, W., Mostafa, N., Alawady, H.: An intelligent warehouse management system using the internet of things. International Journal of Engineering & Technology Sciences 32, 59–65 (12 2020)
13. John, V., Vikitset, N.: Impact of b2c e-commerce on small retailers in thailand: An investigation into profitability, operating efficiency, and employment generation. SSRN Electronic Journal (01 2019)
14. Ju, S., and, H.T.: Competition and operating efficiency of manufacturing companies in e-commerce environment: empirical evidence from chinese garment companies. Applied Economics 55(19), 2113–2128 (2023)
15. Kalakota, R., Whinston, A.B.: Electronic commerce: a manager's guide. Addison-Wesley Longman Publishing Co., Inc., USA (1997)

16. Khalid, A.H., Sapry, H.R.M., Jaafar, J., Ahmad, A.R.: The Application of Technology Organization Environment Framework in the Approaches for Smart Warehouse Adoption, pp. 57–61. Springer Nature Switzerland, Cham (2024)
17. Khan, M.G., Huda, N.U., Zaman, U.K.U.: Smart warehouse management system: Architecture, real-time implementation and prototype design. Machines 10(2) (2022)
18. Khan, M.G., Huda, N.U., Zaman, U.K.U.: Smart warehouse management system: Architecture, real-time implementation and prototype design. Machines 10(2) (2022)
19. Kleisiari, C., Duquenne, M.N., Vlontzos, G.: E-commerce in the retail chain store market: An alternative or a main trend? Sustainability 13(8) (2021)
20. Kumar, S.: Online shopping-a literature review. In: Proceedings of National Conference on Innovative Trends in Computer Science Engineering. pp. 129–131 (2015)
21. Lee, C., Lv, Y., Ng, K.K., Ho, W., Choy, K.: Design and application of internet of things-based warehouse management system for smart logistics. International Journal of Production Research pp. 1–16 (10 2017)
22. Liu, X., Cao, J., Yang, Y., Jiang, S.: Cps-based smart warehouse for industry 4.0: A survey of the underlying technologies. Computers 7(1) (2018)
23. Mao, J., Xing, H., Zhang, X.: Design of intelligent warehouse management system. Wireless Personal Communications 102, 1355–1367 (2018)
24. Min, H.: Smart warehousing as a wave of the future. Logistics 7(2) (2023)
25. Piardi, L., Costa, P., Oliveira, A., Leitão, P.: Mas-based distributed cyber-physical system in smart warehouse. IFAC-PapersOnLine 56(2), 6376–6381 (2023)
26. Rosário, A., Raimundo, R.: Consumer marketing strategy and e-commerce in the last decade: A literature review. Journal of Theoretical and Applied Electronic Commerce Research 16(7), 3003–3024 (2021)
27. Saxena, E., Gupta, D.: Factors influencing online shopping behaviour: A review of motivating and deterrent factors. The Marketing Review 18, 3–24 (06 2018)
28. Shan, H., Yang, K., Shi, J.: A strategic perspective analysis for improving operational inefficiency of e-commerce based on integrated bsc and super-sbm model. In: Proceedings of the 2019 3rd International Conference on Management Engineering, Software Engineering and Service Sciences. p. 128–134. Association for Computing Machinery (2019)
29. Shashidharan, M., Anwar, S.: Importance of an efficient warehouse management system. Turkish Journal of Computer and Mathematics Education 12(5), 1185–1188 (2021)
30. Srivastava, A., Thaichon, P.: What motivates consumers to be in line with online shopping?: a systematic literature review and discussion of future research perspectives abhinav srivastava park thaichon. Asia Pacific Journal of Marketing and Logistics 35 (04 2022)
31. Strack, F., Werth', L., Deutsch, R.: Reflective and impulsive determinants of consumer behavior. Journal of Consumer Psychology 16(3), 205–216 (2006)
32. Venkatesan, V., Mythili, G.: A study of factors affecting online shopping in chennai. Indian Journal of Public Health Research & Development 10, 24 (04 2019)
33. Yang, M.H.: Analysis of consumers' shopping behavior in physical stores and online malls. (2021)
34. Zhao, X., Wang, S., Wang, H., He, T., Zhang, D., Wang, G.: Hst-gt: Heterogeneous spatial-temporal graph transformer for delivery time estimation in warehouse-distribution integration e-commerce. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. p. 3402–3411. CIKM '23, Association for Computing Machinery (2023)
35. Žunić, E., Delalić, S., Hodžić, K., Beširević, A., Hindija, H.: Smart warehouse management system concept with implementation. In: 2018 14th Symposium on Neural Networks and Applications (NEUREL). pp. 1–5 (2018)

Wang Yuqin is a lecturer at the School of Tourism Management, Wuhan Business University. Research expertise: Tourism and Hospitality Management, Higher Education.

Chin-Shyang Shyu received his Ph.D. from National Pingtung University in Taiwan(2010) and is currently researching in the fields of agricultural farm management, marketing management and outdoor event planning.

Cheng-Sheng Lin received his PhD from National Chung Hsing University in Taiwan (2006). He is currently an Assistant Professor at Formasa University. Research expertise: agricultural digital technology, economic benefit analysis and agricultural digital marketing.

Chao-Chien Chen is a highly regarded professor in the Department of Leisure and Recreation Management at Asia University in Taiwan. With a passion for research, Dr. Chen has developed extensive expertise in recreation behavior, sustainable tourism, and ecotourism. He has authored numerous articles in top-tier academic journals and has presented at conferences both domestically and internationally, sharing his insights and knowledge with fellow experts in the field.

Thing-Yuan Chang is a Professor of the Department of Information Management at National Chin-Yi University of Technology, and concurrently serves as a director of Dahan Institute of Technology and Ta Ming High School, and a consultant of the Taichung Computer Association. He obtained his Ph.D. from the Institute of Information Management, National Central University. His research field focuses on IoT applications, enterprise resource planning systems, and information systems integration. His academic papers have been published in journals like Engineering Applications of Artificial Intelligence, NTU Management Review, the International Journal of Electronic Commerce Studies, Sensors and Materials, etc. To date, he has obtained five invention patents and eleven utility model patents from the Republic of China.

Liang Shan is an Associate Professor at the School of Tourism Management, Wuhan Business University. Research expertise: Tourism and Hospitality Management, Higher Education.

Received: September 05, 2024; Accepted: January 05, 2025.

Computer Science and Information Systems 22(3):1061–1080 https://doi.org/10.2298/CSIS241030040C

Deep Learning-Driven Decision Tree Ensembles for Table Tennis: Analyzing Serve Strategies and First-Three-Stroke Outcomes

Che-Wei Chang [1], Sheng-Hsiang Chen[2], Peng-Yu Chen[3], and Jing-Wei Liu [4*]

[1,3] Department of Recreational Sport, National Taiwan University of Sport, No. 16, Sec. 1, Shuangshi Rd., North Dist., Taichung City
404401 Taiwan (R.O.C.)
chewei@gm.ntus.edu.tw
60931ponpon@gmail.com

[2,4] Department of Sport Information and Communication, National Taiwan University of Sport, No. 16, Sec. 1, Shuangshi Rd., North Dist., Taichung City
404401 Taiwan (R.O.C.)
harvestpaleale@gmail.com
liujingwei.ntus@gmail.com

Abstract. This paper presents a novel artificial intelligence system that integrates deep learning-driven decision tree ensemble algorithms (DLDDTEA) for table tennis match analysis. By analyzing videos of professional matches featuring Lin Yun-Ju and Ma Long, the system extracts key insights into player techniques, hitting positions, and scoring outcomes. DLDDTEA processes the video data and constructs a predictive model to determine optimal serve positions and estimate point win/loss probabilities within the first three exchanges. The results revealed distinct serve strategies and techniques: Lin Yun-Ju favors backhands, whereas Ma Long prefers forehands. Based on these findings, this study offers specific training and strategic recommendations for both players. Thus, the proposed system offers a comprehensive framework for table tennis match analysis, enabling players to gain a deeper understanding of their strengths and weaknesses, ultimately facilitating the development of more effective training and competitive strategies.

Keywords: deep learning, decision tree, video analysis, table tennis match model, notational analysis, convolutional neural networks.

1. Introduction

In table tennis matches, the serve, return of service, and subsequent stroke are collectively known as the "First Three Strokes." The techniques and tactics employed during the initial three exchanges significantly influence the match outcomes. The first three strokes in table tennis are crucial. The serve, in particular, is critical as it creates opportunities for the subsequent return and is a key tactic for restricting the opponent. Wang [1] analyzed the table tennis matched held at the 2012 London and 2016 Rio Olympics and elucidated the significant importance of serve position and return techniques. Specifically, the serve

* Corresponding author

is a preemptive strategy, used either to score directly or to create attacking opportunities. Conversely, the return of service plays a crucial role in the contest of these first three strokes. Yu and Gao [2] analyzed the matches of the 2019 World Table Tennis Championship Men's Singles and found that forehand serves and aggressive returns were the highest-scoring techniques. Yin et al. [3] employed the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to objectively and accurately analyze the techniques and tactics used in the first three strokes of table tennis matches.

In [4] and [5] identified significant differences in the serve and return techniques between male and female table tennis players. Male players tend to favor positions closer to the net, whereas female players prefer those near the baseline. Regarding return techniques, males use more push strokes near the net, whereas females employ more push strokes near the baseline. Regardless of sex, players require excellent technical skills, physical fitness, tactical variations, and psychological resilience to compete effectively. Through video analysis of table tennis training, Gumilar et al. [6] found that top players enhance the quality of their third stroke after serving to directly increase their scoring chances. [2] and [7] suggested that high-quality returns can transform a defensive situation into an offensive one, increasing scoring opportunities.

Grycan et al. [8] studied the winning actions, techniques, and tactics used by the leading male table tennis players from 1970 to 2021, finding that while the first three strokes remained crucial throughout this period, the importance of serving as a direct scoring stroke decreased. Therefore, this study conducted deep-learning based analysis of match videos to construct a model of the first three strokes used by professional male players Lin Yun-Ju and Ma Long, including the frequency of their techniques and the probabilities of winning or losing points within these exchanges. Using a decision-tree algorithm, this study constructs an artificial intelligence (AI)-based table tennis match model based on the placement strategies of these players. This AI model can be extended to other table tennis players, enabling them to construct their own match models and adjust their training and match strategies.

This paper introduces an AI algorithm that leverages deep learning-driven decision tree ensemble algorithms (DLDDTEA) to analyze table tennis matches. Specifically, convolutional neural networks (CNNs) are used to extract technical information, hitting positions, and scoring outcomes from video footage of matches featuring Lin Yun-Ju and Ma Long. Subsequently, a decision tree ensemble method, based on principles similar to those underlying the C4.5 algorithms, is employed to construct a predictive model. This model is can identify optimal serve positions and estimating the probability of winning or losing points within the first three strokes.

2. Literature Review

This literature review is divided into three subsections: Section 2.1 reviews studies on notational analysis, Section 2.2 analyzes those on the applications of decision trees in sports analysis, and Section 2.3 discusses those employing AI for image analysis.

2.1 Notational Analysis

Several studies have employed notational analysis (also known as tagging analysis) to investigate tactics used in table tennis and other sports. Malagoli et al. [9] used this method to study the matches of 20 top table tennis players, focusing on the playing styles of Asian and European players. Their results indicated that Asian players are more aggressive and have more effective services, providing valuable insights for coaches and athletes regarding both technical and tactical applications. Djokić et al. [10] analyzed 20 matches from the German league and the European TOP 16 semifinals and finals, focusing on serve analysis of top table tennis players. They found that short forehand serves were most prevalent (76.9%), primarily targeting the opponent's backhand area. The direct scoring rate from the serves was 11.6%, whereas the third and fifth strokes featured scoring rates of 22.4 and 10.9%, respectively. Serve errors were primarily observed in the third (25.0%) and fifth strokes (22.4%), with an overall error rate of 1.5%. A correlation between service types and match outcomes was also found. Zhou [11] used notational analysis to examine data from 200 matches of the Chinese table tennis team and found that the scoring rate for attacks within the end line (AIEL) was significantly higher than that for attacks outside the end line (AOEL). Additionally, the timing of attacks significantly influenced the scoring rate, with earlier attacks yielding higher scores. AIEL primarily involved backhand flips, whereas AOEL mainly comprised backhand drives. Pradas de la Fuente [4] applied notational analysis to study the technical and tactical differences between male and female table tennis players. They found that male players use forehand techniques more frequently, whereas female players prefer defensive techniques. Tactically, male players are more aggressive, particularly when using flip techniques, whereas female players tend to be more defensive. Furthermore, male players' movements are faster and more explosive, whereas female players focus more on stability and defense. Guarnieri et al. [12] collected data from 25 Paralympic table tennis matches between 2012 and 2018 and used notational analysis and Kinovea software to analyze players' stroke types, ball-bounce areas, and stroke outcomes. They found that C1 players primarily used backhand and forehand drives, whereas C5 players mainly used backhand and forehand pushes and backhand topspin.

In other sports such as volleyball and football, notational analysis has been used to study technical and tactical indicators, revealing gender-specific preferences in techniques and movements. Huang [13] employed notational analysis to study the techniques of male and female single finalists in the 1990 Grand Slam tennis tournaments held on different court surfaces, revealing significant differences in the players' techniques across various surfaces. Another key finding was that there was no significant difference in the ratio of good-to-bad serves between the first and second serves for either male or female players. However, the overall scoring rate from serves was substantial. Jiang [14] used video notational analysis to study techniques and winning factors in men's single tennis, enhancing the reliability of the study by increasing the observation frequency and ensuring content validity. Gambhir [15] summarized the use of notational analysis at the 16th International Table Tennis Science Congress, highlighting its application for studying the kinematic characteristics, techniques, and health of table tennis players. Malagoli Lanzoni et al. [16] used notational analysis to record and analyze 20 table tennis matches involving 40 male and 40 female players from the top 111 (female) and 120 (male) ITTF world rankings. They found that the most common serve types for both sexes were the forehand topspin and serve. Females preferred backhand

blocks and pushes, whereas males preferred forehand topspin counters. Additionally, females often used a single step and preferred not to move their feet while striking, whereas males used crossover and pivot steps. Serves typically targeted areas close to the net and returns were often directed to the opponent's backhand corner. In [17] and [18] used global-positioning-system-based tracking combined with notational analysis to record tactical and physical indicators, and analyzed team possession, passing, and shooting performances. Herold et al. [19] used machine learning with notational analysis to help coaches analyze the attack efficiency and tactics of professional male football players.

2.2 Application of Decision Trees in Sports Analysis

Sigari et al. [20] developed a method for classifying sports videos using four simple classifiers: adjacent nodes, linear discriminant analysis, decision trees, and probabilistic neural networks. The experimental results indicated a correct classification rate of 78.8%. Kostuk and Willoughby [21] used decision-tree-based analysis to examine the choice between scoring and not scoring in the later stages of curling matches. Analysis of world-class curling competitions revealed that North American players often chose not to score in the final moments, whereas European players opted to score, concluding that not scoring in the later stages was a better choice. Pai et al. [22] combined support vector machine and decision tree models to predict basketball game outcomes, and achieved an average accuracy of 85.25%. Mumcu and Mahoney [23] applied decision trees to generate systematic and informed decisions in three sports-marketing scenarios. Çene et al. [24] compared decision trees, Technique for Order Preference by Similarity to Ideal Solution, and Performance Index Rating methods to analyze individual game data of players from the 2017–2018 European Basketball League season and identified the best and worst-ranked players. Yıldız [25] used decision trees to classify top football teams in Spain, Italy, and England with 77% accuracy. Gu and He [26] employed the fuzzy decision tree algorithm to analyze and predict member attrition in the fitness industry, achieving a classification and prediction accuracy of 97.8%. Tsai et al. [27] employed various methods, including logistic regression, support vector machine, decision tree C4.5, classification and regression tree, random forest, and extreme gradient boosting (XGBoost), to analyze the accuracy of stress state detection in table tennis players using electroencephalogram analysis. XGBoost achieved an accuracy of 86.49% for three-level stress classification, outperforming other methods by up to 11.27%. Ghosh et al. [28] used decision trees, learning vector quantization, and support vector machine to predict outcomes from a Grand Slam tennis database and found that decision trees outperformed the other two models. Chiang et al. [29] used decision tree analysis to study the segmented swimming styles of 11–12-year-old Japanese boys and girls in a 200-m individual medley and identified the winning strokes. They found that breaststroke and backstroke were the most successful strokes for boys, whereas breaststroke and butterfly were beneficial for girls. Madinabeitia et al. [30] used decision tree analysis to classify 7,345 individual statistics from 335 games in the 2018/2019 Spanish Men's Basketball League season, identifying low-contribution foreign players (FLC; 23.8% as shooting guards), high-contribution foreign players (FHC; 32.1% as centers), and low-contribution Spanish players (SLC; 32.9% as small forwards), thereby providing coaches with insights into team formation. Zuccolotto et al. [31] used the classification and regression tree

algorithm for decision trees to analyze NBA 2020/2021 season data, visually and robustly representing the scoring probabilities of players or teams. Papageorgiou et al. [32] used decision tree analysis on data of 90 NBA players from the 2019–2022 seasons, evaluated the performance of 14 machine-learning models for predicting players' overall performance rankings using 18 advanced basketball statistics and key performance indicators.

2.3 AI-based Image Analysis

Mat Sanusi et al. [33] employed smartphone sensors, Microsoft Kinect, and neural networks to develop the Table Tennis Tutor program, which detects correct and incorrect strokes during table tennis training. Liu and Ding [34] used CNNs and long short-term memory (LSTM) networks to create a table tennis trajectory and spin prediction algorithm, achieving an accuracy rate exceeding 98% and thereby enhancing the performance of table tennis robots. Qiao [35] combined a deep deterministic policy gradient (DDPG), CNN, and LSTM to develop deep-learning techniques for automatically detecting and analyzing technical and tactical indicators from match videos, including stroke type, ball trajectory, spin speed, and landing points. They achieved feature-extraction, target-tracking, and trajectory-prediction accuracies of 89, 93, and 91%, respectively. Song et al. [36] applied k-means clustering to divide player win rates into three stages: service, receive, and rally attacks. They then used a hybrid LSTM–back propagation neural network (LSTM–BPNN) model to predict match outcomes. Finally, they used Shapley additive explanations (SHAP) to analyze the impact of three technical indicators (stroke position, stroke technique, and serve strategy) and three tactical indicators (scoring patterns, return strategies, and the first three stroke analyses) on match outcomes. The results showed that the hybrid LSTM–BPNN model achieved a 92.5% accuracy for predicting match outcomes. Liu et al. [37] used neural networks to analyze the top Taiwanese singles player, Lin Yun-Ju, using a dataset comprising 22 international match videos from 2015 to 2021. Using the 3S (Speed, Spin, Spot) theory for analysis, they found that a slow service speed combined with a long-backhand service spot led to a higher win rate. Conversely, half-long forehand serve spot resulted in a higher loss rate. These findings suggest that Lin could adjust his serving style to improve his win rate.

Despite these promising advancements, several challenges have hindered the widespread adoption of AI for table tennis video analysis. First, training robust AI models requires substantial amounts of annotated data, which are time-consuming and labor-intensive to collect and label. Second, their performance can be significantly affected by environmental factors such as lighting and background conditions. Third, the complex nature of deep-learning models makes it difficult to interpret their decision-making processes, limiting their applicability in scenarios requiring transparent explanations, such as refereeing. Finally, the sequential nature of table tennis actions can produce redundant detections when each frame is individually analyzed. Addressing these challenges is crucial for advancing the application of AI in table tennis and unlocking its full potential. In summary, the literature on decision trees reveals their capacity to organize collected data into graphical tree structures, clearly displaying the outcomes at different nodes and thereby identifying the most advantageous options.

3. Research Methodology

Deep-learning techniques, such as CNNs, have revolutionized sports analysis. Recent studies, including that by Li et al. [38], have demonstrated the efficacy of CNNs in automatically extracting technical and tactical features from table tennis videos. This research builds upon earlier work utilizing notational analysis, as exemplified by [9] and [10], which relied on manual coding of matches. Furthermore, machine-learning techniques have been successfully applied for data analysis in other sports, such as curling [21] and basketball [22], to predict outcomes and uncover underlying patterns.

This study employs a multifaceted approach using DLDDTEAs, combining C4.5 decision-tree algorithm, CNNs, and notational analysis to construct a table tennis match analysis model. This model leverages three key variables—techniques, placement, and outcomes—derived from matches featuring Lin Yun-Ju and Ma Long [38-41]. The primary objective was to identify the most effective combinations of these variables that yielded high scoring rates. The research methodology comprised three main stages: data collection, data processing, and model building. These stages were further divided into seven distinct steps, as illustrated in Fig. 1.

Fig. 1. Framework of the methodology employed for analyzing table tennis matches featuring Lin Yun-Ju and Ma Long

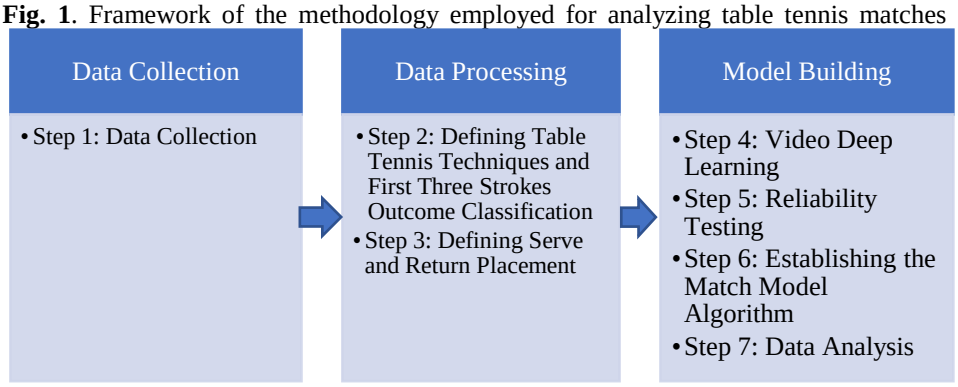


Fig. 2 illustrates the training process of the multifaceted DLDDTEA model, including the image segmentation and video preprocessing techniques. This process involves the following steps: First, the image input parameters are defined. These parameters specify the image segments for nine distinct areas on each side of the table tennis net, primarily to record serve and return positions. Additionally, two grip types are defined: Forehand Backspin and Backhand Backspin. Images for ten common table tennis techniques are also defined, including Backspin, Topspin, Counter Loop, Chiquita, Short Push, Long Push, Flick, Fast Drive, Defense, and Lob. Next, CNNs are employed to convert the video footage into images, which are then classified according to the defined techniques and ball positions. The classification results are categorized as Serve Points, Receiving and Scoring, Third Stroke Attack, Serve Errors, and Continued Rally. Finally, a decision tree algorithm is used to further classify the techniques and ball positions into Serve Points, Receiving and Scoring, Third Stroke Attack, and Serve Errors.

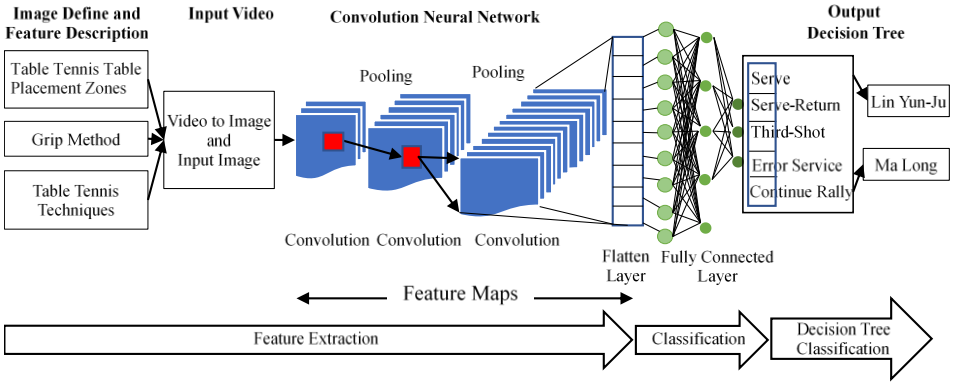


Fig. 2. Flowchart of the proposed multifaceted DLDDTEA approach

Step 1: Data Collection

This study analyzed match videos of professional male table tennis players, Lin Yun-Ju and Ma Long, from 2020, 2022, and 2023, totaling three matches.

Step 2: Defining Table Tennis Techniques and First Three Strokes Outcome Classification

There are ten commonly used table tennis techniques: forehand and backhand backspin, topspin, counter, flick, push, chop, drive, block, and lob [42]. Additionally, the various outcomes of the first three strokes were classified. The algorithm codes for each technique and their outcomes are listed in Table 1.

Table 1. Algorithm codes for table tennis techniques and outcomes

Forehand and Backhand Backspin		Techniques		Results	
Code	Grip	Code	Techniques	Code	Results
1	Forehand Backspin	B	Backspin	S	Serve Points
2	Backhand Backspin	T	Topspin	R	Receiving and Scoring
		CL	Counter Loop	T	Third Ball Attack
		C	Chiquita	S	Serve Error
		SP	Short Push	CR	Continued Rally
		LP	Long Push		
		F	Flick		
		FD	Fast Drive		
		D	Defense		
		L	Lob		

Step 3: Defining Serve and Return Placements

The table tennis table was divided into nine hitting zones, with serve and return placements coded from 1 to 9, as shown in Fig. 3.

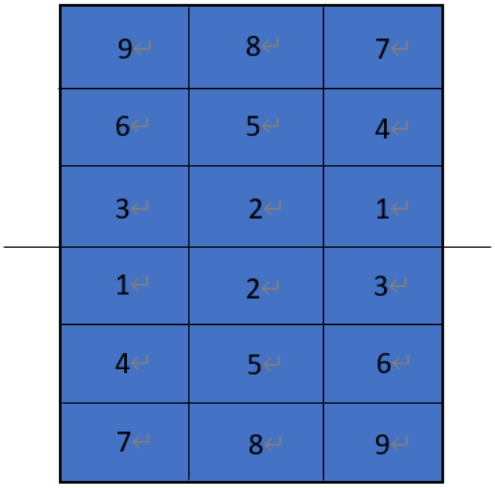


Fig. 3. Serve and return placement zones with their corresponding codes

Step 4: Video-based Deep Learning
CNNs were employed to automatically extract the technical and tactical features of both players during the first three strokes [38]. Through video analysis, the techniques, hitting placements, and point outcomes during the first three strokes were extracted.

Step 5: Reliability Testing
Two table tennis players with more than ten years of experience were invited to watch the videos together. They marked and recorded the match situations of the two players according to the methods described in Steps 2 and 4. Initially, a match was randomly selected and one game was viewed and marked. Thereafter, reliability testing was conducted using Holsti's [43] intercoder agreement and reliability formulas, as shown in Equations (1) and (2). When the reliability exceeded 0.8, the reliability standard is met and comprehensive coding can begin [44].

$$\text{Average inter} - \text{coder agreement} = \frac{2M}{N1+N2} \quad (1)$$

$$\text{Reliability} = n \times \frac{\text{Average inter}-\text{coder agreement}}{\{1+[(n-1)\times\text{Average inter}-\text{coder agreement}]\}} \quad (2)$$

In Equation (1), M represents the number of complete agreements, N1 is the number of agreements by Coder 1, and N2 is the number of agreements by Coder 2. In Equation (2), n is the number of coders involved.

Using these equations, we obtained a reliability measurement of 0.93 > 0.80. The calculation process is as follows:

Average intercoder agreement: ((0.84 + 0.81 + 0.86) / 3 = 0.83)

Reliability: ((3 * 0.83) / [1 + (3 - 1) * 0.83] = 0.93)

This indicated that the consistency among the three coders reached the standard level, allowing for comprehensive coding.

Step 6: Establishing the Match Model Algorithm
The study utilized a notational analysis method to categorize ten types of table tennis techniques and divided the table into nine zones for coding serve and return positions.

This approach aimed to reduce human error and enhance data consistency. Deep-learning techniques using CNN and decision tree models were used to automatically extract technical and tactical features from match videos, rapidly process large datasets, and reduce subjective analyses. The analysis of three match videos from 2020, 2022, and 2023 featuring Lin Yun-Ju and Ma Long involved the following coding and calculation classification processes:

1. All records were treated as a single node.
2. Based on Steps 1–3, the match videos were compared, and for each variable (table tennis techniques, serve and return placements, and first three stroke outcomes), the appropriate split points were identified based on the video analysis.

Reliability testing with experienced players ensured coding consistency, achieving a reliability of 0.93, which was above the standard of 0.8. A decision tree algorithm C4.5 was used to create predictive models to identify the optimal serve positions and win-loss probabilities within the first three shots. The analysis was continued until each ball hit and its outcome satisfied the classification for each node.

4. Results and Discussion

The analysis of the three videos encompassed 238 serves executed by the two players. Table 2 lists the statistics of the techniques used by Lin Yun-Ju and Ma Long, including the top three most frequently used ones

Table 2. Classification of techniques used by Lin Yun-Ju and Ma Long

Table Tennis Techniques	Lin Yun-Ju		Ma Long	
	Times	%	Times	%
Forehand Topspin	3	1.42%	8	3.65%
Backhand Topspin	21	9.95%	11	5.02%
Forehand Backspin	24	11.37%	52	23.74%
Backhand Backspin	16	7.58%	16	7.31%
Forehand Counter	18	8.53%	15	6.85%
Backhand Counter	9	4.27%	13	5.94%
Backhand Flick	63	29.86%	2	0.91%
Forehand Short Push	7	3.32%	33	15.07%
Backhand Short Push	28	13.27%	11	5.02%
Forehand Long Push	5	2.37%	12	5.48%
Backhand Long Push	1	0.47%	0	0%
Forehand Flick	1	0.47%	11	5.02%
Forehand Drive	1	0.47%	1	0.46%
Backhand Drive	4	1.90%	17	7.76%
Forehand Defense	4	1.90%	6	2.74%
Backhand Defense	6	2.84%	11	5.02%
Forehand Lob	0	0%	0	0%
Backhand Lob	0	0%	0	0%
Total	211	100%	219	100.0%

From Table 2, it is evident that neither player used the lob technique in the first three strokes, as it is primarily a defensive technique. This indicates that both players adopted

an aggressive approach during the first three strokes.

The most frequently used techniques in the first three strokes were backhand topspin (9.95%), forehand backspin (11.37%), backhand chiquita (29.86%), and backhand short push (13.27%). Backhand topspin and forehand backspin are off-table techniques, whereas backhand chiquita and backhand short push are on-table techniques. This suggests that, when receiving serves and attacking during the third stroke, Lin Yun-Ju primarily used the backhand chiquita for on-table balls, with the backhand short push as a secondary option. For off-table balls, he mainly uses the backhand topspin for topspin balls and forehand backspin for backspin balls.

Ma Long's most frequently used techniques in the first three strokes were the forehand backspin (23.74%) and forehand short push (15.07%). Forehand backspin is an off-table technique, whereas forehand short push is an on-table technique. This indicates that Ma Long primarily used the forehand backspin for off-table balls and the forehand short push for on-table balls. As both are forehand techniques, this suggests that Ma Long prefers using his forehand when receiving serves and during the first three strokes.

There were 37 serve points between the two players (Fig. 4). Lin Yun-Ju scored 23 serve points with the following placement distribution: 1 point for Location 2, 6 points for Location 5, 5 points for Location 6, 5 points for Location 7, 3 points for Location 8, and 3 points for Location 9. In response to Lin Yun-Ju's serves, Ma Long's errors included 12 backspins, 5 topspins, 1 chiquita, 3 short pushes, and 2 flicks. Ma Long scored 14 serve points with the following placements: 1 point for Location 1, 7 points for Location 4, 1 point for Location 6, 1 point for Location 7, 3 points for Location 8, and 1 point for Location 9. In response to Ma Long's serves, Lin Yun-Ju's errors included 2 backspins, 2 topspins, 9 chiquitas, and 1 long push. Therefore, from the serve-point data, it is evident that Lin Yun-Ju scored approximately one-third more than Ma Long, demonstrating a clear advantage in serving. Lin Yun-Ju's points were more dispersed across half-long and long balls, whereas Ma Long's service-return errors were primarily concentrated on the backspin loops. Among Ma Long's 14 serving points, half were concentrated at Location 4, and Lin Yun-Ju's service-return errors were mainly chiquitas. Thus, during training, Lin Yun-Ju should focus on improving his chiquita return technique for balls coming at Location 4.

In terms of receiving and scoring, 62 records were available for the two players (Fig. 5). Among them, Lin Yun-Ju scored 26 points from receiving serves, with 4 forehand and 22 backhand shots. Regarding the techniques used for these points, 4 points were obtained from backspin, five from topspin, 13 from chiquitas, 3 from short pushes, and 1 from long pushes. In terms of placement, 1 point was obtained at Placement 2, 1 at Placement 3, 2 at Placement 4, two at Placement 5, 1 at Placement 6, 9 at Placement 7, 3 at Placement 8, and 7 at Placement 9. Ma Long scored 36 points from receiving serves with 19 forehand and 17 backhand shots. Regarding the techniques used, he scored 23 points from backspin, 3 from topspin, 1 from chiquitas, 6 from short pushes, 2 from long pushes, and 1 from flicks. Regarding placement, 1 point was scored at Placement 2, 3 at Placement 4, 4 ay Placement 5, 5 at Placement 6, 5 at Placement 7, 5 at Placement 8 and 13 points at Placement 9.

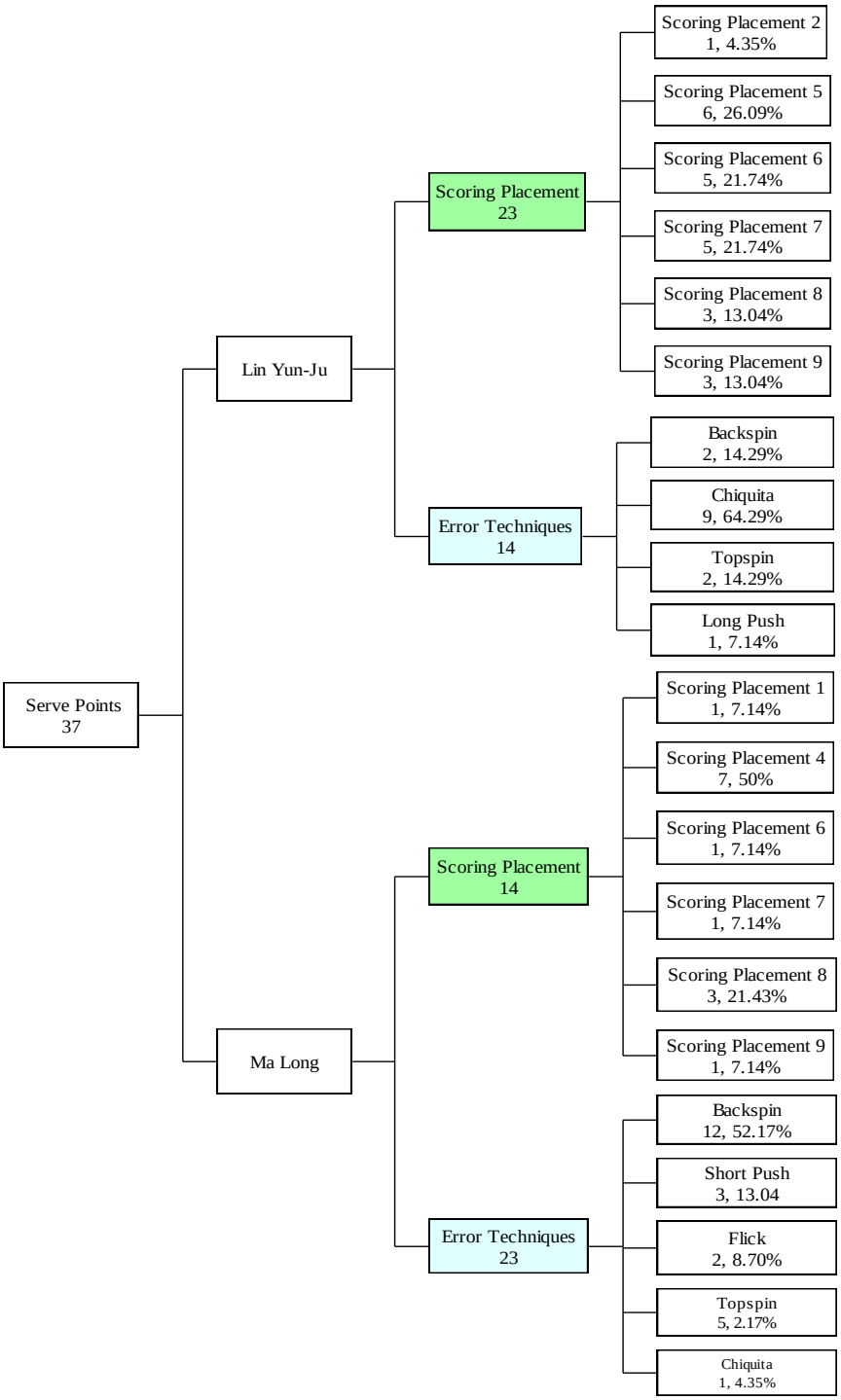


Fig. 4. Serve scoring locations of Lin Yun-Ju and Ma Long

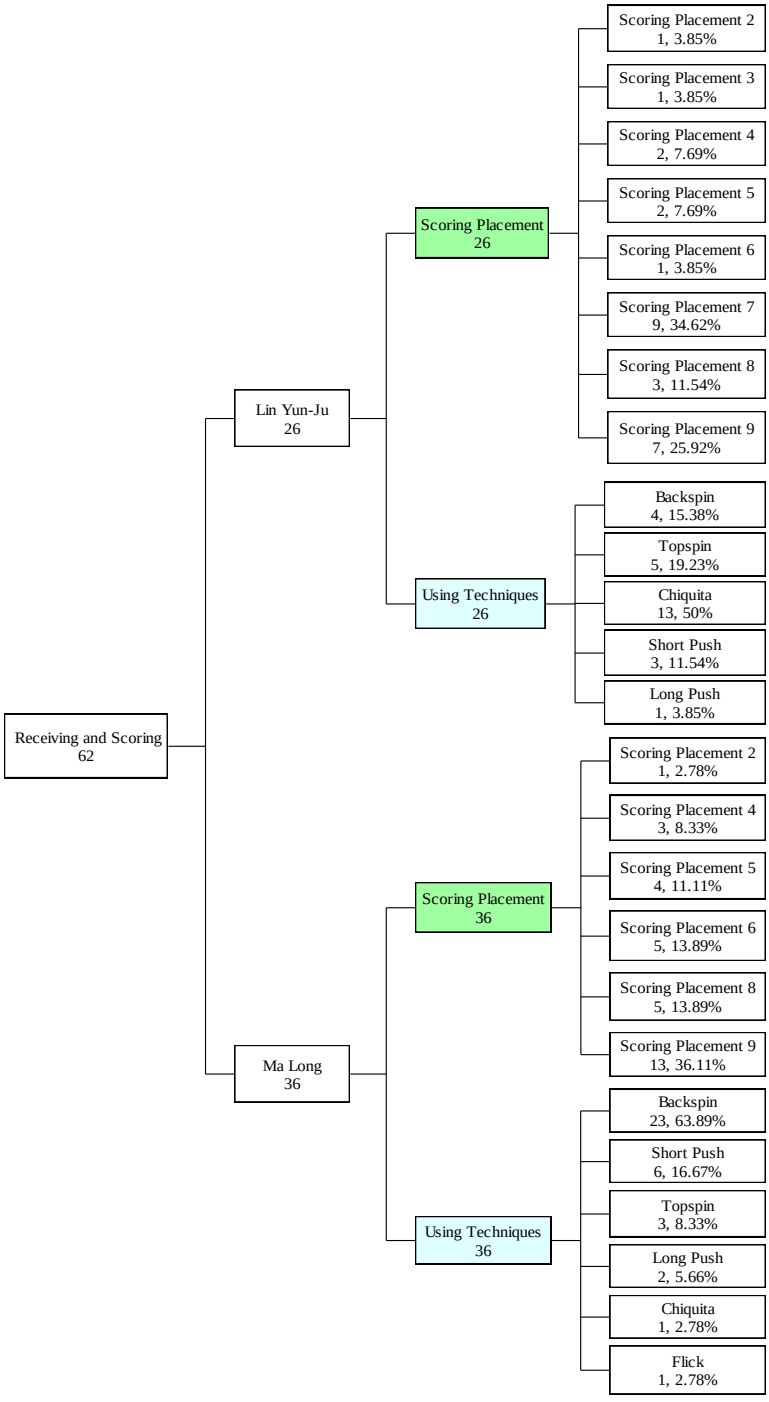


Fig. 5. Relationship between serve-return scoring locations and techniques used by Lin Yun-Ju and Ma Long

From the receiving and scoring data, it is evident that Ma Long scored more points while receiving services than Lin Yun-Ju. However, Lin Yun-Ju's backhand was more prominent when receiving serves, with chiquitas being his primary scoring method. His scoring placements were mainly concentrated at Placements 7 and 9. However, Ma Long's forehand and backhand were relatively balanced when receiving serves, with backspin being his main scoring technique. Additionally, his scoring placements were primarily concentrated in Placement 9. Therefore, if Lin Yun-Ju can use more chiquitas to target Placements 7 and 9 when receiving serves and focus on Ma Long's backspin targeting Placement 9 after serving, it will likely help him score more efficiently. Conversely, if Ma Long can use more backspin to attack Lin Yun-Ju's Placement 9 when receiving serves and defend against Lin Yun-Ju's chiquitas targeting Placements 7 and 9 after serving, it will likely help him score more points.

In terms of third-shot scoring, 49 records were available for both players (see Fig. 6). Lin Yun-Ju scored 25 points, with 16 forehand and 9 backhand shots. Regarding the points scored based on the techniques used, 9 were from backspin, 1 from topspin, 7 from counter, 4 from chiquitas, 1 from flicks, 1 from fast drives, and 2 points from defensive shots. Regarding the placements, 1 point was scored at Placement 4, 15 at Placement 7, 1 at Placement 8, and 8 at Placement 9. Ma Long scored 24 points from the third-shot scoring, with 16 forehand and 8 backhand shots. Regarding the techniques used, 8 points were scored from backspin, 2 from topspin, 7 from counter, 1 from short pushes, 2 from flicks, and 4 from fast drives. Regarding the placement, 1 point was scored at Placement 4, 2 at Placement 5, 2 at Placement 6, 8 at Placement 7, 6 at Placement 8, and 5 at Placement 9.

In terms of third-shot scoring, the two players had relatively balanced scores and primarily used forehand responses. The scoring techniques were predominantly backspin and counter. Lin Yun-Ju's scoring was concentrated at Placements 7 and 9, whereas Ma Long's were distributed across Placements 7, 8, and 9. Therefore, both players should create opportunities for forehand backspin and countering during the third shot after serving. To achieve the most effective scoring strategy, Lin Yun-Ju should focus on attacking Ma Long's Placements 7 and 9, whereas Ma Long should distribute his attacks across Lin Yun-Ju's Placements 7, 8, and 9.

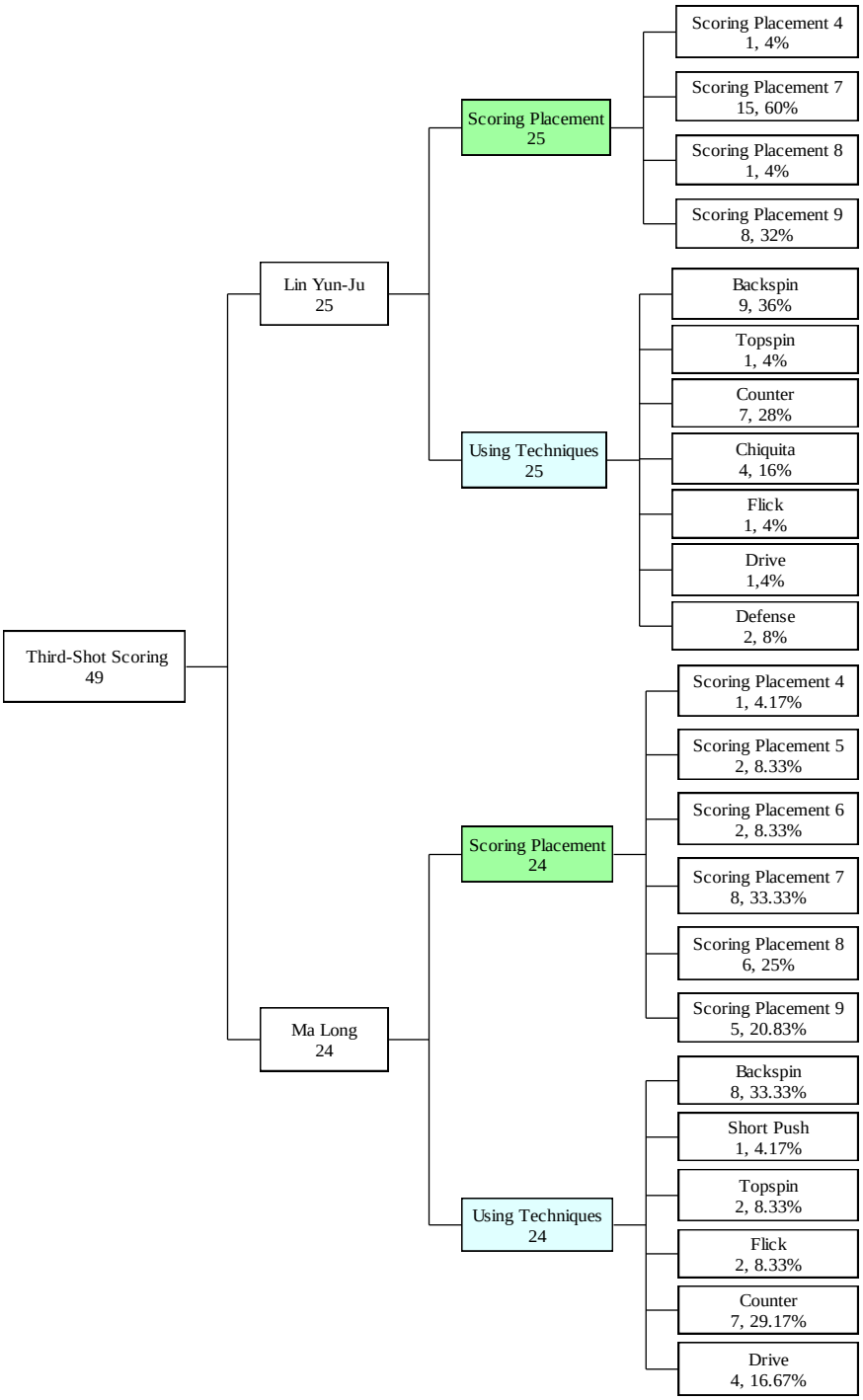


Fig. 6. Relationship between scoring placements and techniques used by Lin Yun-Ju and Ma Long in the third shot

5. Conclusions and Limitations

This study developed an innovative artificial intelligence system for analyzing table tennis matches, integrating notational analysis, deep learning techniques, and decision tree algorithms. By analyzing match videos of Lin Yun-Ju and Ma Long, the system automatically extracted key information, including player techniques, hitting positions, and scoring outcomes. Additionally, a predictive model was constructed to identify optimal serve positions and estimate the probability of scoring within the first three strokes. The main contributions of this study are as follows:

(1) Advancement of Scientific Table Tennis Training: The proposed system not only analyzes the technical characteristics and tactical preferences of two elite players, such as Lin Yun-Ju's preference for backhand strokes and Ma Long's inclination toward forehand strokes, but more importantly, it serves as an objective and quantitative analytical tool. Thus, it can help coaches and players gain a deeper understanding of match dynamics, facilitating the development of evidence-based training plans and more effective competitive strategies.

(2) Facilitation of Personalized Training Regimens: Based on the analytical output of the developed model, coaches can design personalized training regimens tailored to individual players' specific strengths and weaknesses. For example, targeted training can address areas identified for improvement, such as Lin Yun-Ju's receiving skills at Positions 4 and 8, and Ma Long's backhand returns from mid- to long-distance Positions 5, 6, and 7, ultimately enhancing technical proficiency and competitive performance.

(3) Enhancement of Match Outcome Prediction Accuracy: Through deep learning and decision tree algorithms, the system can predict the development trend of a match more accurately, helping players make more informed decisions during matches.

5.1 Future Research Directions

(1) Expanding the research sample: Future research should include more players and match data to improve the generalizability and predictive accuracy of the proposed model.

(2) Analyzing the entire match: In addition to the first three strokes, technical and tactical changes throughout the entire match can be analyzed to gain a more comprehensive understanding of player performance.

(3) Incorporating opponent information: Future research should integrate data on the technical characteristics and tactical strategies of opponents. This inclusion will enable a more accurate assessment of player performance within the context of specific match-ups and facilitate the development of targeted competitive strategies.

(4) Extending the Model to Different Genders and Age Groups: Future studies should aim to extend the model’s applicability to encompass players of different genders and age groups. This can facilitate the identification of gender- and age-specific technical and tactical variations, enabling the development of tailored training and competition strategies for diverse player demographics.

The DLDDTEA table tennis match analysis system developed in this study

represents a novel approach for table tennis training and performance analysis. By integrating deep learning and decision tree algorithms, it effectively analyzes match data, providing evidence-based insights for coaches and players to enhance to enhance player development and match performance.

5.2 Limitations

Despite the successful development and preliminary validation of the integrated deep-learning and decision tree-based table tennis match analysis system using data from the matches of Lin Yun-Ju and Ma Long, this study had certain limitations:

(1) Limited Sample Size: The analysis was conducted on a limited sample of three matches involving Lin Yun-Ju and Ma Long. This restricted sample size may limit the generalizability of the findings, making it difficult to extend the results to other players or diverse match scenarios.
(2) Restricted Analytical Scope: The study primarily focused on techniques and tactics employed within the first three exchanges of a match, thus not fully capturing the complex dynamics of complete matches. Given the dynamic nature of table tennis, focusing solely on the initial exchanges may not fully represent players' overall performance and strategic adaptability.
(3) Lack of Opponent Information: Although this study offered an in-depth analysis of the techniques and scoring patterns of Lin Yun-Ju and Ma Long, it did not sufficiently account for the influence of their opponents' technical characteristics and tactical approaches. As opponent strategies can significantly influence player performance, the absence of this information may have limited the comprehensiveness of the analysis.
(4) Validation of Model Generalizability: The developed model was trained on data specific to Lin Yun-Ju and Ma Long. Therefore, its generalizability to other players requires further rigorous validation. Its accuracy and reliability may vary when applied to players with different playing styles and skill sets.

These limitations highlight important avenues for future research. Subsequent studies should prioritize increasing the sample size, conducting more comprehensive analyses of full matches, incorporating opponent-specific information, and validating the model across a broader range of players to enhance its generalizability and practical applicability.

Acknowledgments. This research was funded by the MOE Teaching Practice Research Program, grant number PBM1110042.

References

1. Wang, J.: Comparison of Table Tennis Serve and Return Characteristics in the London and the Rio Olympics. International Journal of Performance Analysis in Sport. Vol. 19, No. 5, 683–697. (2019)
2. Yu, J., Gao, P.: Interactive Three-Phase Structure for Table Tennis Performance Analysis: Application to Elite Men's Singles Matches. Journal of Human Kinetics. Vol. 81, 177–188. (2022)

3. Yin, H., Chen, X., Zhou, Y., Xu, J., Huang, D: Contribution Quality Evaluation of Table Tennis Match by Using TOPSIS-RSR Method - An Empirical Study. BMC Sports Science, Medicine and Rehabilitation, Vol. 15, No. 1, e132. (2023)
4. Pradas de la Fuente, F., Ortega-Zayas, M.Á., Toro-Román, V., Moreno-Azze, A.: Analysis of Technical–Tactical Actions in High-Level Table Tennis Players: Differences between Sexes. Sports. Vol. 11, No. 11, 225. (2023)
5. Pradas, F., Toro-Román, V., Castellar, C., Carrasco, L.: Analysis of The Spatial Distribution of the Serve and the Type of Serve-Return in Elite Table Tennis. Sex differences. Frontiers in Psychology. Vol. 14, 1243135. (2023)
6. Gumilar, A., Negara, J.D.K., Nuryadi, N., Firmansyah, H., Mudjihartono, M., Hambali, B., Purnomo, E.: Development of Digital-Based Return Board Table Tennis Learning Media. Jurnal Patriot. Vol. 6, No. 1, 13–20. (2024)
7. Santosh, R.S.: The SPORTS CLASS THINKING Towards Business Success: Unique Ideas from Sports-field to Win in Business Management. Notion Press, Chennai, India. (2021)
8. Grycan, J., Kołodziej, M., Bańkosz, Z.: Technical and Tactical Actions of the World's Leading Male Table Tennis Players Between 1970 and 2021. Journal of Sports Science and Medicine. Vol. 22, No. 4, 667–680. (2023)
9. Malagoli Lanzoni, I., Di Michele, R., Merni, F.: A Notational Analysis of Shot Characteristics in Top-Level Table Tennis Players. European Journal of Sport Science. Vol. 14, No. 4, 309–317. (2013)
10. Djokić, Z., Malagoli Lanzoni, I., Katsikadelis, M., Straub, G.: Serve Analyses of Elite European Table Tennis Matches. International Journal of Racket Sports Science. Vol. 2, No. 1. (2020)
11. Zhou, X.: Explanation and Verification of The Rules of Attack in Table Tennis Tactics. BMC Sports Science, Medicine and Rehabilitation. Vol. 14, No. 1, 6. (2022)
12. Guarnieri, A., Presta, V., Gobbi, G., Ramazzina, I., Condello, G., Malagoli Lanzoni, I.: Notational Analysis of Wheelchair Paralympic Table Tennis Matches. International Journal of Environmental Research and Public Health. Vol. 20, No. 5, 3779. (2023)
13. Huang, J. C.: Analysis of Players' Hitting Techniques on Tennis Courts Made of Different Materials, Physical Education Journal. Vol. 12, 225-240. (1990)
14. Jiang, Z. G.: Research on Men's Tennis Singles Skills and Winning and Losing Factors in Taiwan, Physical Education Journal. 34, 79-92. (2003)
15. Gambhir, M.: "Match Analysis" Using Notational Analysis and Data Analytics in Table Tennis with Interactive Visualization, In Proceedings Book of the 16th ITTF Sports Science Congress. International Table Tennis Federation, Budapest, Hungary, 286–299. (2019)
16. Malagoli Lanzoni, I., Cortesi, M., Russo, G., Bankosz, Z., Winiarski, S., Bartolomei, S.: Playing Style of Women and Men Elite Table Tennis Players. International Journal of Performance Analysis in Sport. Vol. 24, No. 5, 495–509. (2024)
17. Folgado, H., Bravo, J., Pereira, P., Sampaio, J.: Towards the Use of Multidimensional Performance Indicators in Football Small-Sided Games: The Effects of Pitch Orientation. Journal of Sports Sciences. Vol. 37, No. 9, 1064–1071. (2019)
18. McGuckian, T. B., Cole, M. H., Chalkley, D., Jordet, G., Pepping, G. J.: Constraints on Visual Exploration of Youth Football Players During 11v11 Match-Play: The Influence of Playing Role, Pitch Position and Phase of Play. Journal of Sports Sciences. Vol. 38, No. 6, 658–668. (2020)
19. Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., Meyer, T.: Machine Learning in Men's Professional Football: Current Applications and Future Directions for Improving Attacking Play. International Journal of Sports Science & Coaching. Vol. 14, No. 6, 798–817. (2019)
20. Sigari, M. H., Sureshjani, S. A., Soltanian-Zadeh, H.: Sport Video Classification Using an Ensemble Classifier. In 2011 7th Iranian Conference on Machine Vision and Image Processing, Tehran, Iran. 1–4. (2011)

21. Kostuk, K. J., Willoughby, K. A.: A Decision Support System for Scheduling the Canadian Football League. Interfaces. Vol. 42, No. 3, 286–295. (2012)
22. Pai, P. F., ChangLiao, L. H., Lin, K. P.: Analyzing Basketball Games by a Support Vector Machines with Decision Tree Model. Neural Computing & Applications. Vol. 28, No. 12, 4159–4167. (2017)
23. Mumcu, C., Mahoney, K.: Use of Decision Tree Model in Sport Management. Case Studies in Sport Management. Vol. 7, No. 1, 1–3. (2018)
24. Çene, E., Parim, C., Özkan, B.: Comparing the Performance of Basketball Players with Decision Trees and TOPSIS. International Journal of Data Science and Applications. Vol. 1, No. 1, 21–28. (2018)
25. Yıldız, B.F.: Applying Decision Tree Techniques to Classify European Football Teams. Journal of Soft Computing and Artificial Intelligence. Vol. 1, No. 2, 86–91. (2020)
26. Gu, Z., He, C.: Application of Fuzzy Decision Tree Algorithm Based on Mobile Computing in Sports Fitness Member Management. Wireless Communications and Mobile Computing. Vol. No. 1, 4632722. (2021)
27. Tsai, Y. H., Wu, S. K., Yu, S. S., Tsai, M. H.: Analyzing Brain Waves of Table Tennis Players with Machine Learning for Stress Classification. Applied Sciences. Vol. 12, No. 16, 8052. (2022)
28. Ghosh, S., Sadhu, S., Biswas, S., Sarkar, D., Sarkar, P.P.: A Comparison Between Different Classifiers for Tennis Match Result Prediction. Malaysian Journal of Computer Science. Vol. 32, No. 2, 97–111. (2019)
29. Chiang, H. H., Hsieh, C. H., Xiao, S. H., Lin, C. Y., Tsai, M. H.: Analysis of Swimming Strokes of 200 Meters Individual Medley for Japanese Swimmers Using a Decision Tree, Sports & Exercise Research, 21(1), 17-29. (2019)
30. Madinabeitia, I., Pérez, B., Gomez-Ruano, M.Á., Cárdenas, D.: Determination of Basketball Players' High-Performance Profiles in The Spanish League. International Journal of Performance Analysis in Sport. Vol. 23, No. 2, 83–96. (2023)
31. Zuccolotto, P., Sandri, M., Manisera, M.: Spatial Performance Analysis in Basketball with CART, Random Forest and Extremely Randomized Trees. Annals of Operations Research. Vol. 325, No. 1, 495–519. (2023)
32. Papageorgiou, G., Sarlis, V., Tjortjis, C.: Evaluating the Effectiveness of Machine Learning Models for Performance Forecasting in Basketball: A Comparative Study. Knowledge and Information Systems. Vol. 66, No. 7, 4333–4375. (2024)
33. Mat Sanusi, K.A., Mitri, D.D., Limbu, B., Klemke, R.: Table Tennis Tutor: Forehand Strokes Classification Based on Multimodal Data and Neural Networks. Sensors. Vol. 21, No. 9, 3121. (2021)
34. Liu, Q., Ding, H.: Application of Table Tennis Ball Trajectory and Rotation-Oriented Prediction Algorithm Using Artificial Intelligence. Frontiers in Neurorobotics. Vol. 16, 820028. (2022)
35. Qiao, F.: Application of Deep Learning in Automatic Detection of Technical and Tactical Indicators of Table Tennis. PLOS ONE. Vol. 16, No. 3, e0245259. (2021)
36. Song, H., Li, Y., Zou, X., Hu, P., Liu, T.: Elite Male Table Tennis Matches Diagnosis Using SHAP and a Hybrid LSTM–BPNN Algorithm. Scientific Reports. Vol. 13, No. 1, 11533. (2023)
37. Liu, J. W., Hsu, M. H., Lai, C. L., Wu, S. K.: Using Video Analysis and Artificial Neural Network to Explore Association Rules and Influence Scenarios in Elite Table Tennis Matches. The Journal of Supercomputing. Vol. 80, No. 4, 5472–5489. (2024)
38. Li, H., Ali, S.G., Zhang, J., Sheng, B., Li, P., Jung, Y., Wang, J., Yang, P., Lu, P., Muhammad, K., Mao, L.: Video-based Table Tennis Tracking and Trajectory Prediction Using Convolutional Neural Networks. Fractals. Vol. 30, No. 05, 2240156. (2022)
39. Yanan, P., Jilong, Y., Heng, Z.: Using Artificial Intelligence to Achieve Auxiliary Training of Table Tennis Based on Inertial Perception Data. Sensors. Vol. 21, No. 19, 6685. (2021)

40. Chang, C. W., Qiu, Y. R.: Constructing a Gaming Model for Professional Tennis Players Using the C5.0 Algorithm. Applied Sciences. Vol. 12, No. 16, 8222. (2022)
41. Chiu, C. H., Ke, S. W., Tsai, C. F., Lin, W. C., Huang, M. W., Ko, Y. H.: Deep Learning Based Decision Tree Ensembles for Incomplete Medical Datasets. Technology and Health Care, Vol. 32, No. 1, 75–87. (2024)
42. Wang, L., Zhou, Z., Zou, Q.: Analysis System for Table Tennis Techniques and Tactics Using Data Mining. Soft Computing. Vol. 27, No. 19, 14269–14284. (2023)
43. Holsti, O.R.: Content Analysis for the Social Sciences and Humanities. Addison-Wesley Pub. Co, Reading, Massachusetts, USA. (1969).
44. Riffe, D., Lacy, S., Fico, F., Watson, B.: Analyzing Media Messages: Using Quantitative Content Analysis in Research. Routledge, New York, USA. (2019)

Che-Wei Chang is a professor of Department of Recreational Sport, National Taiwan University of Sport. He specializes in artificial intelligence, decision science, information technology application, big data, data mining, software evaluation, system evaluation, and grey systems. His paper appeared in International Journal of Advanced Manufacturing Technology, International Journal of Manufacturing Technology and Management, Information and Software Technology, Quality and Quantity, Robotics and Computer Integrated Manufacturing, Production Planning & Control, Computers & Industrial Engineering, Expert Systems with Applications, Information Sciences, Knowledge Management Research & Practice, Mathematics, Applied Sciences and others.

Sheng-Hsiang Chen is currently an Associate Professor with the Department of Sport Information and Communication, National Taiwan University of Sport. His research interests include sociology of sport, analysis of techniques and tactics in sports, sport policy analysis, sport management, and sport marketing. His articles mainly appeared in Physical Education journal, the International Journal of Contemporary Hospitality Management, IEEE Access, Soft Computing, and so on.

Peng-Yu Chen is a Master's degree candidate at the Department of Recreational Sport, National Taiwan University of Sport. She started playing table tennis at the age of 8. She was selected for the Taiwan table tennis national team at the ages of 12, 15, and 18. She achieved first place in the team event at the Taipei Open, second place in the team event at the World Middle School Games, and third place in the singles event. After turning 20, she transitioned to coaching, specializing in the analysis of table tennis techniques and tactics.

Jing-Wei Liu is an associate professor of Department of Sport Information & Communication, National Taiwan University of Sport. He specializes in Sport Science, Artificial Intelligence, Big Data, Data Mining, Soft Computing, and Fuzzy Time Series. His paper appeared in IEEE Access, International Journal of Interactive Multimedia and Artificial Intelligence, Journal of Supercomputing, Soft Computing, Journal of Ambient Intelligence and Humanized Computing, International Journal of Information Technology & Decision Making, Journal of Systems and Software, Journal of Computer Information Systems, Computers and Industrial Engineering, Computers & Education, Computers and Mathematics with Applications, Economic Modelling, Journal of Computer Information Systems, International Journal of Information and Management

Sciences, Plant Systematics and Evolution, Expert Systems with Applications, Advanced Materials Research, Open Journal of Social Sciences, and others.

Received: October 30, 2024; Accepted: January 24, 2025.

Computer Science and Information Systems 22(3):1081–1104 https://doi.org/10.2298/CSIS241018041W

Exploring Factors Affecting User Intention to Accept Explainable Artificial Intelligence

Yu-Min Wang[1] and Chei-Chang Chiou[2,*]

[1] Department of Information Management, National Chi Nan University, Puli 545301, Taiwan
ymwang@ncnu.edu.tw

[2] Department of Accounting, National Changhua University of Education, Changhua 500, Taiwan
ccchiou@cc.ncue.edu.tw

Abstract. Explainable Artificial intelligence (XAI) represents a pivotal innovation aimed at addressing the "black box" problem in AI, thereby enhancing users' understanding of AI reasoning processes and outcomes. The implementation of XAI is not merely a technological endeavor but also involves various individual factors. As XAI remains in its early developmental stages and exhibits unique characteristics, identifying and understanding the factors influencing users' intention to adopt XAI is essential for its long-term success. This study develops a research model grounded in the characteristics of XAI and prior technology acceptance studies that consider individual factors. The model was evaluated using data collected from 252 potential XAI users. The validated model exhibits strong explanatory power, accounting for 45% of the variance in users' intention to use XAI. Findings indicate that perceived value and perceived need are key determinants of users' intention to adopt XAI. These results provide empirical evidence and deepen the understanding of user perceptions and intentions regarding XAI adoption.

Keywords: explainable artificial intelligence, artificial intelligence, user acceptance, individual differences, intention to use.

1. Introduction

Advancements in computing capabilities and algorithms have driven the rapid progress and widespread adoption of Artificial Intelligence (AI) [1],[2],[3]. AI encompasses a wide array of techniques, algorithms, machines, and software capable of learning, reasoning, self-correcting, and executing instructions or actions [4],[5]. As AI performance and applications expand, it increasingly integrates into daily life, replacing human roles across various professional fields such as healthcare, public safety, inspections, and finance.

Historically, AI development prioritized effectiveness and performance, often overlooking the transparency of reasoning processes and the interpretability of results.

* Corresponding author

Users have typically aware only of the inputs and outputs, lacking insight into the underlying decision-making processes of AI systems. This opacity has led to AI being referred to as a "black box" [6],[7]. The growing importance of AI applications has heightened concerns about privacy, fairness, ethics, and the potential for deception [8],[9]. The lack of transparency in AI systems raises questions regarding the fairness and accountability of AI-generated decisions, as well as potential biases and unintended consequences.

Explainable AI (XAI) seeks to address these concerns by prioritizing interpretable and transparent AI models. This approach has garnered increasing attention and expectations from various fields [10]. XAI involves three key elements: a new machine learning process, an explainable model, and an interactive interface [11]. Unlike traditional AI, XAI emphasizes performance, reasoning transparency, and interpretability equally. This dual focus presents significant development challenges, necessitating innovative technologies and resources. Furthermore, implementing XAI extends beyond technological considerations, encompassing individual user factors. As XAI remains in its early stages and exhibits unique characteristics, understanding the factors influencing users' intention to adopt XAI is crucial for its long-term success. By exploring these factors, researchers and practitioners can develop user-centric XAI systems that address users' needs and expectations while ensuring transparency, trustworthiness, and effective decision-making.

Drawing from the unique attributes of XAI and prior technology acceptance studies, this study develops a model to investigate the factors influencing users' intention to adopt XAI. The findings aim to provide empirical evidence and deepen understanding of user perceptions and intentions regarding XAI adoption.

The remainder of this paper is structured as follows: Section 2 introduces XAI and outlines its theoretical foundations. Section 3 presents the research model and hypotheses derived from the literature. Section 4 describes the research methods, including the data collection and measurement of constructs. Section 5 details the data analysis techniques and findings. Finally, Sections 6 and 7 discuss the implications of the results and offer conclusions based on the study's findings.

2. Literature Review

2.1. Explainable AI

The widespread application of Artificial Intelligence (AI) in various aspects of human life, both commercially and industrially, has become increasingly prevalent. Over the past decade, significant advancements in AI technology, particularly in machine learning, have facilitated the integration of AI into critical decision-making processes, including credit scoring, criminal justice, job recruitment, and teaching evaluation [10]. The demand for AI arises from the human need for effective decision-making. Bucincai et al. [12] emphasized that decision-making is a fundamental cognitive process through which individuals select a choice or action plan from a range of alternatives. While

humans often rely on mental shortcuts or heuristic methods to make decisions, these methods, although efficient, can sometimes lead to systematic errors. To support reliable and sound decision-making, various fields, including management and medicine, have employed computer-based decision support systems [13],[14]. With recent advancements in AI, these systems have achieved high levels of precision, leading to the adoption of AI in an increasing number of domains for decision support purposes [12].

However, Hind et al. [10] highlight the growing concern about the trustworthiness of AI systems' decision-making processes in society. As AI use becomes more widespread, there is a strong demand for AI systems to provide explanations for their decisions. Paradoxically, as AI systems becomes more effective, they also grow increasingly complex, making it difficult to understand their inner workings. Hind et al. [10] specifically note that certain technologies, such as deep neural networks and large random forests, are challenging to explain even for experts, resulting in AI models functioning as “black boxes”. This lack of transparency introduces significant risks. Liao et al. [15] further support this assertion, highlighting the adoption of machine learning technology, particularly those utilizing opaque deep neural networks, across various practical fields. This trend has sparked significant interest in XAI within both academic and practical communities, emphasizing the need for greater transparency and interpretability in AI systems.

Bucinca et al. [12] emphasize the importance of evaluating interpretable systems within XAI-driven decision support systems. They argue for user-centered methods and interdisciplinary research to align technological advancements with user needs. Liao et al. [15] support this perspective, noting the diversity and their ability to incorporate multiple styles of interpretation to enhance user experience. Hoffman et al. [6] stress that XAI is a dynamic process, requiring continuous user experience evaluation to foster trust and dependence. Doshi-Velez and Kim [16] propose a taxonomy for evaluating XAI systems, focusing on domain experts, non-professionals, and agency tasks, and stress the importance of selecting appropriate evaluation indicators [17].

Hoffman et al. [18] suggest subjective evaluation measures, such as user trust and satisfaction, as key indicators for interpretable systems. However, Lakkaraju and Bastani [19] caution that subjective measures may fail to reliably predict user performance, potentially leading to biases or dependence on flawed interpretations. Therefore, it is necessary to consider multiple dimensions and comprehensive evaluation indicators when assessing XAI systems. This approach goes beyond subjective measures and aims to provide a more complete understanding of the system's performance and effectiveness [12].

Casimir Wierzynski [7] and Hani Hagras [20] were pioneers in developing a comprehensive evaluation index for XAI needs from a user perspective. Wierzynski [7] emphasized that interpretability is a subject of great scientific fascination and societal significance, as it resides at the convergence of various actively researched domains in machine learning and AI. The key areas of focus encompass the following elements [7],[20]:

- Bias: Ensure the AI system avoids biases from training data, models, or objective functions. Curate diverse and representative data, use techniques like augmentation and balancing, and regularly evaluate performance on different subgroups.

- Fairness: Verify fairness in AI-based decisions. Define fairness and identify whom it should be fair to. Assess decision-making processes for bias and ensure equal treatment.
- Transparency: Users should have the right to understand how AI affects decision-making. Seek explanations in understandable terms, formats, and language. Establish grounds for appealing decisions.
- Security: A lack of explanation may undermine confidence in AI reliability. Relate this to generalization in statistical learning. Address the challenge of tying errors to unseen data.
- Causality: Learn from data and obtain accurate inferences and explanations for underlying phenomena. Seek a mechanical understanding from the learned model.
- • Engineering: Debug incorrect output from trained models. Identify and rectify errors through thorough analysis and troubleshooting.

Addressing the technical aspects of XAI methods, the recent introduction of Shapley Additive Explanations (SHAP) by Lundberg and Lee [21] has gained widespread adoption in AI research applications [22-29]. SHAP addresses a critical challenge in interpreting tree-based and ensemble models by directly uncovering the contribution of features to predictions [27]. It provides a significant advantage over traditional feature importance analysis by accurately reflecting the impact of features on individual samples, capturing both positive and negative influences [27]. Antwarg et al. [22] and Jabeur et al. [23] emphasize SHAP's superiority over other statistical methods in interpreting machine learning model outputs. Rizk-Allah et al. [30] proposed a model utilizing Local Interpretable Model-Agonistic Explanation (LIME), based on XAI, to identify the critical factors influencing the accuracy of the power generation forecasts in smart solar systems. Similarly, Rajabi and Etminani [31] conducted a systematic review of knowledge graphs (KGs) in XAI systems. Their findings revealed that KGs are primarily used in pre-model XAI for feature and relationship extraction and in post-model XAI for reasoning and inference. The review also highlighted several studies employing KGs to explain XAI models in the healthcare domain.

In exploring the subjective and behavioral aspects of XAI, Chinu and Bansal [32] conducted a literature review that identified several key issues with AI systems, including unfair or biased decisions, poor accuracy, insufficient reliability, and the absence of evaluation metrics for assessing the effectiveness of explanations and data security. These findings underscore the challenges, and opportunities in advancing the field of XAI. In practical applications, Wang, Bian, and Chen [33] proposed and validated the use of XAI to address the interpretability challenges of deep learning models in classroom dialogue analysis. Their results indicated that XAI enhances teachers' trust in and acceptance of AI models for classroom dialogue analysis without increasing cognitive load. Additionally, Sano, Shi, and Kawabata [34] employed gradient-weighted class activation mapping, an XAI technique, to extract key features for each impression based on facial images and impression evaluation results. Their findings indicated that this computational method using XAI could independently identify the determinants of facial impressions without relying on visual attention captured by eye-tracking devices. Ebermann, Selisky, and Weibelzahl [35] explored the impact on user acceptance when the decisions made by an AI system and their associated explanations contradict the user's decisions. They found that in decision

scenarios with cognitive misfit, users are significantly more likely to experience negative emotions and provide unfavorable evaluations of the AI system's support.

2.2. Theoretical Bases

The Technology Acceptance Model (TAM) is widely utilized in research on technology adoption. TAM proposes a causal chain involving external factors (stimulus) $\rightarrow$ beliefs (cognitive response) $\rightarrow$ intention $\rightarrow$ behavior [36]. Building on TAM, Agarwal and Prasad [37] incorporated five individual differences as external factors, arguing that these differences influence the behavioral intention to adopt new information technology through beliefs. Additionally, Agarwal and Prasad [37] suggested that cross-sectional research, where beliefs and intention are measured simultaneously, might exclude usage as a research variable.

Hong et al. [38] made a significant contribution by emphasizing the impact of individual differences on technology adoption, particularly in the context of digital libraries. They introduced a model examining the factors influencing user acceptance of digital libraries, incorporating system characteristics as a variable and proposing causal relationships: individual differences and system characteristics influence beliefs, which in turn influence intention. Furthermore, Hong et al. [38] highlighted the significance of computer anxiety and computer self-efficacy as potential individual differences. They suggested that future studies should include these constructs in research models to further explore their influence on technology adoption.

Drawing from the aforementioned studies and existing literature on IT acceptance, Wang and Wang [39] developed a research model examining causal relationships in technology adoption. In their model, they posited that individual differences and system characteristics influence beliefs, which subsequently impact intention. Wang and Wang [39] departed from conventional approaches by introducing perceived playfulness as the variable representing beliefs, differing from the commonly used constructs in previous studies. Additionally, they incorporated computer anxiety and computer self-efficacy as individual difference variables in their research model.

Wang et al. [40] built on the theoretical model proposed by Wang and Wang [39] to investigate the acceptance of hedonic information systems. They utilized the same model and variables, including computer anxiety and computer self-efficacy as individual differences, and perceived playfulness as the variable representing beliefs. Using this model, they analyzed the factors influencing the acceptance of hedonic information systems.

Wang et al. [41] provided further support for the arguments by Hong et al. [38] and Wang and Wang [39], emphasizing the significance of individual differences in shaping behavioral intentions through beliefs about IT usage. They introduced a research model positing causal relationships: individual differences influence beliefs, which subsequently influence intention. In their model, Wang et al. [40] incorporated perceived enjoyment as the variable representing beliefs. They categorized individual differences into two groups: personality traits and computer skills-related individual differences, including computer self-efficacy and personal innovativeness. By considering these variables, they examined how individual differences impact the

formation of beliefs and subsequently influence behavioral intention in the context of technology adoption.

Based on the literature review, four key insights emerge. First, the causal chain linking external factors (individual differences and system characteristics) to beliefs and then to intention is a powerful framework for analyzing technological innovations. This chain provides a solid foundation for constructing the research model in this study. Second, as XAI represents an innovative category rather than a specific system, system characteristics are excluded in the research model of this study. Third, individual differences can be categorized as AI-related and AI-unrelated. Anxiety and self-efficacy are important AI, whereas personality traits are significant individual differences that are independent of AI. Lastly, belief variables should align with the characteristics of the technology under study. In the context of XAI adoption, this research model includes perceived value and perceived need as belief variables. Perceived value, proposed by Kim et al. [42], has been highlighted in previous technology adoption research [3],[43],[44]. It reflects users' preferences and evaluations of whether innovation attributes can meet their needs [45]. Considering that XAI is developed based on people's perceptions of AI's shortcomings and deficiencies, perceived need for XAI is included in the research model. This construct reflects the level of demand potential users have for XAI.

3. Research Model and Hypotheses

The research model is depicted in Fig. 1, illustrating the proposed framework. The dependent variable in this study is the "intention to use." Drawing from the characteristics of XAI and previous IT acceptance research, two belief constructs—perceived value and perceived need—are incorporated as antecedents of intention. Furthermore, the interrelationships between these constructs are integrated into the research model. Additionally, three individual differences—personality, AI anxiety, and AI self-efficacy—are identified as significant external factors that influence intention, mediated by the two beliefs.

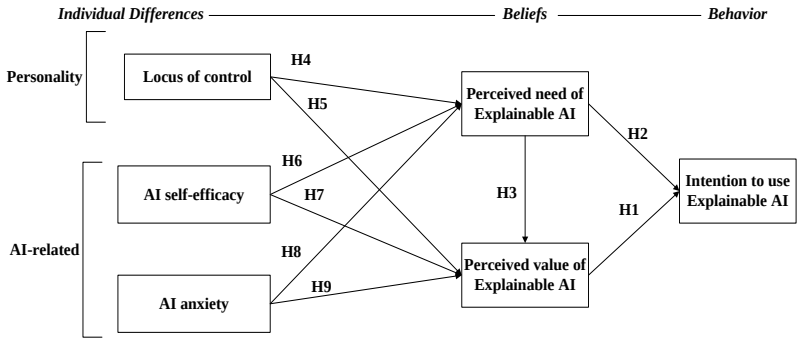


Fig. 1. The research model

3.1. Perceived Value, Perceived Need, and Intention to Use

Perceived value, as defined by Zeithaml [46], refers to the overall assessment made by potential users regarding the utility of an innovative product or service. Numerous studies have consistently demonstrated that perceived value significantly and positively influences adoption intention or behaviors. Empirical evidence supporting the impact of perceived value has been observed across various domains of innovative technologies and applications. For instance, perceived value play a crucial role contexts such as of mobile commerce [27], online gaming [47], Internet protocol television [48], online content services [49], mobile GPS applications [44], mobile catering applications [50], AI technology [51], and XAI [52]. In the XAI environment, when potential users perceive XAI as valuable, they are more likely to exhibit a greater willingness to adopt and utilize it. Therefore, the following hypothesis is proposed:

H1: Perceived value has a positive effect on the intention to use XAI.

Perceived need refers to an individual's personal assessment of the necessity or benefits associated with a particular innovation or change [53],[54],[55]. When potential users perceive a strong need for a specific innovation, they are more likely to attribute higher perceived value to it and demonstrate greater eagerness to adopt the innovation. Numerous studies support the positive impact of perceived need on perceived value and emphasize its significance as a facilitator of behavioral intention [56],[57],[58]. In light of the above, the following hypotheses are proposed:

H2: Perceived need has a positive effect on the intention to use XAI.

H3: Perceived need has a positive effect on perceived value.

3.2. Individual Differences—Personality

Personality is recognized as a significant individual difference influencing innovations adoption through beliefs [37],[41]. Among various personality traits, locus of control has garnered considerable attention and is commonly employed in IT acceptance analyses [59],[60]. Locus of control refers to the extent to which individuals believe they can control events that affect them [61]. Individuals who perceive events as within their control are referred to as having an internal locus of control (internals), while those who attribute events to external factors are characterized as having an external locus of control (externals) [62].

Individuals with a high internal locus of control are more inclined to adopt innovative technologies due to their greater confidence in controlling outcomes compared to individuals with a high external locus of control [63]. Internals, being predisposed to exert control and mastery over their environment, are more likely to perceive the needs and value of XAI, which offers a comprehensible and self-controlled usage environment. Thus, the following hypotheses are proposed:

H4: Internal locus of control has a positive effect on the perceived need of XAI.

H5: Internal locus of control has a positive effect on the perceived value of XAI.

3.3. Individual differences—Self-efficacy and Anxiety

Numerous studies [39],[64],[65] indicate that self-efficacy and anxiety related to specific technology or innovations are crucial individual differences influencing beliefs about using technologies. Self-efficacy refers to an individual's confidence in their ability to execute a specific task or master a new technology [66],[67]. This construct significantly affects perceptions, needs, and the desirability of an innovation or technology [67],[68],[69].

Individuals with high levels of AI self-efficacy are more likely to feel confident and willing to use AI. Consequently, they tend to perceive higher levels of value and need for explainable AI (XAI) because it is viewed as clearer and easier to operate compared to traditional AI. Based on this reasoning, the following hypotheses are proposed:

H6: AI self-efficacy has a positive effect on the perceived need of XAI.

H7: AI self-efficacy has a positive effect on the perceived value of XAI.

Anxiety is another crucial individual difference that has negatively influences IT adoption [39]. Researchers such as Hong et al. [70] emphasize the importance of investigating anxiety in the context of IT adoption. AI anxiety specifically refers to feelings of fear or agitation about AI being out of control [71]. Wang and Wang [72] define AI anxiety as an overall affective response of fear or discomfort that hinders individuals from engaging with AI.

XAI, designed to be more transparent and understandable than traditional AI, can mitigate concerns among individuals with AI anxiety. The enhanced transparency and interpretability of XAI can help alleviate anxiety, making such individuals more likely to perceive XAI as valuable and necessary. Based on this rationale, the following hypotheses are proposed:

H8: AI anxiety has a positive effect on the perceived need of XAI.

H9: AI anxiety has a positive effect on the perceived value of XAI.

4. Research Methodology

4.1. Construct Measures

To ensure the content validity of the construct measures in this study, initial items were developed based on existing instruments from the fields of IT/innovation adoption, AI anxiety, and XAI. These items were subsequently revised and adapted to fit the specific context of XAI. The wording, completeness, and appropriateness of the items were reviewed and confirmed by five experts specializing in information management and AI. Ultimately, a total of 29 items were used to measure the six constructs outlined in the research model. All measurement items were evaluated using a five-point Likert scale ranging from "1 - strongly disagree" to "5 - strongly agree." The specific measurement items and their corresponding references for each construct are summarized in Table 1.

Table 1. Measurement items used in the study

Construct	Items	References
Locus of Control (LC)	LOC1. People's misfortunes result from the mistakes they make. LOC2. In the long run, people get the respect they deserve in this world. LOC3. Capable people who fail to become leaders have not taken advantage of their opportunities. LOC4. Becoming a success is a matter of hard work; luck has little or nothing to do with it. LOC5. What happens to me is my own doing. LOC6. When I make plans, I am almost certain that I can make them work. LOC7. In my case, getting what I want has little or nothing to do with luck. LOC8. Getting people to do the right thing depends upon ability; luck has little or nothing to do with it. LOC9. There is really no such thing as "luck." LOC10. Most misfortunes are the result of lack of ability, ignorance, laziness, or all three. LOC11. It is impossible for me to believe that chance or luck plays an important role in my life.	[59]
AI Self-Efficacy (ASE)	ASE1. I am confident in my ability to effectively utilize an AI technique/product. ASE2. I have confidence in my capacity to proficiently use an AI technique/product independently. ASE3. Based on my knowledge and skills, I am confident that I can readily employ an AI technique/product.	[73] ,[74]
AI Anxiety (AIA)	AIA1. Learning how an AI technique/product works makes me anxious. AIA2. I am afraid that an AI technique/product may replace humans. AIA3. I am afraid that an AI technique/product may get out of control and malfunction. AIA4. I find humanoid AI techniques/products (e.g., humanoid robots) scary.	[72]
Perceived Need of XAI (PN)	PN1. Transparency: XAI is capable of providing me with explanations that I can comprehend if it makes a decision that affects me. PN2. Causality: XAI not only offers accurate inferences but also provides me with explanations when it learns a model from data. PN3. Bias: It is essential for an AI technique/product to ensure that forecasts and recommendations are based on objective and comprehensive data rather than problematic, defective, or biased data. PN4. Fairness: XAI should guarantee that decisions made by an AI technique/product that impact me are conducted in a fair manner. PN5. Safety: Even without an explanation of how it reaches conclusions, I can have confidence in the reliability of an AI technique/product. PN6. Engineering: I possess the capability to identify and rectify incorrect outputs generated by an AI technique/product.	[7], [35]
Perceived Value of XAI (PV)	PV1. I think the development towards XAI is worthwhile. PV2. I think the development towards XAI is important. PV3. I think the development towards XAI is valuable.	[42],[75],[76]
Behavioral Intention (BI)	BI1. I am willing to use XAI products/services in the future. BI2. I expect I will use XAI products/services in the future	[77]

4.2. Data Collection and Sample Characteristics

The survey methodology was chosen for this study because it allows for the generalization of results [78]. To collect empirical data and validate the research model, a web-based survey platform was developed. A total of 265 responses were obtained for this study. Of these, thirteen were excluded because the respondents either did not complete the questionnaire in its entirety or reported no prior knowledge of AI. Consequently, 252 valid responses were considered for subsequent analysis. The demographic characteristics of the sample are presented in Table 2. Among the respondents, 95 (37.7%) were male and 157 (62.3%) were female. The distribution of respondents' ages was as follows: 20 years or younger: (8.3%), 21-30 years: (28.6%), 31-40 years: (20.2%), 41-50 years: (23.0%), and 51 years or older: (19.9%). Regarding educational attainment, approximately 45.6% of respondents had completed a college education, while 44.8% held a master's degree or higher, reflecting a high level of education among the majority of participants. The sample demonstrated considerable diversity, as evidenced by the wide range of ages and income levels represented.

Table 2. Respondent profiles

Demographics	Frequency
Gender	
Male	37.7%
Female	62.3%
Age	
≦ 20	8.3%
21-30	28.6%
31-40	20.2%
41-50	23.0%
≧ 51	19.9%
Education	
Senior high school	9.6%
College	45.6%
Graduate school or above	44.8%
Monthly income (NT$)	
Less than 10,000	23.4%
10,001-30,000	9.9%
30,001-60,000	36.1%
60,001-100,000	16.7%
Over 100,000	13.8%

5. Data Analysis and Results

The SmartPLS software, utilizing the partial least squares-structural equation modeling (PLS-SEM) approach, was chosen for data analysis in this study. PLS-SEM was selected for its strengths in exploratory research and its ability to handle non-normal data distributions. Following the guidelines of Hair et al. [79], data analysis was conducted in two stages. In the first stage, the measurement model was assessed to evaluate the relationships between constructs and their corresponding measurement items. This

included an examination of the reliability and validity of the measurement items, as well as the overall fit of the measurement model. The second stage involved assessing the structural model, focusing on the hypothesized relationships between constructs. This stage aimed to evaluate the significance and strength of these relationships. By adopting this two-stage approach, the study ensured a comprehensive evaluation of both the measurement and structural models to derive insights into the relationships among the constructs under investigation.

5.1. Measurement Model

The measurement model was assessed using four criteria: indicator reliability, internal consistency reliability, convergent validity, and discriminant validity.

Indicator Reliability. Indicator reliability was evaluated by analyzing the outer loadings of the measurement items. Items with outer loadings below 0.6 and insufficient content validity were considered for removal to enhance the model's robustness. Consequently, five items (LC1, LC2, LC6, PN5, and PN6) were excluded. Table 3 demonstrates that most items have outer loadings above 0.7, indicating strong reliability. For items with outer loadings exceeding 0.4—the minimum threshold for exploratory research—all constructs displayed satisfactory indicator reliability.

Internal Consistency Reliability. Internal consistency reliability was assessed using the rho_A and Cronbach's Alpha coefficients, as recommended by Wong [80]. Table 3 shows that all rho_A and Cronbach's Alpha values surpassed the threshold of 0.7, indicating strong internal consistency reliability for each construct. This indicates that the constructs were measured consistently across items.

Convergent Validity. Convergent validity was assessed using the Average Variance Extracted (AVE). Table 3 reveals that all constructs, except locus of control, have AVE values exceeding 0.5, supporting their convergent validity. As suggested by Cheung and Wang [81], Fornell and Larcker [82], and Lam [83], convergent validity can still be acceptable if the AVE is below 0.5, provided the composite reliability is above the recommended level and all factor loadings are greater than 0.5. For the locus of control construct, all outer loadings exceeded 0.5, and its composite reliability was 0.85. Therefore, despite its AVE being below 0.5, the convergent validity of this construct was deemed adequate. Overall, the measurement model demonstrated satisfactory convergent validity based on AVE values and additional criteria.

Discriminant Validity. Discriminant validity was assessed using the heterotrait-monotrait ratio (HTMT), which compares the average correlation between items across different constructs to the average correlation between items within the same construct [84]. A threshold value of 0.85 is typically used to indicate discriminant validity. As shown in Table 4, all HTMT ratios were below this threshold, confirming discriminant validity. This indicates that the constructs were sufficiently distinct, as inter-construct correlations were lower than intra-construct correlations.

In summary, the measurement model demonstrated satisfactory reliability and validity, as evidenced by the assessment of indicator reliability, internal consistency reliability, convergent validity, and discriminant validity.

Table 3. Construct reliabilities and validities

Constructs	Items	Outer Loading	rho_A	Cronbach's Alpha	AVE
LC	LC3	0.66	0.86	0.83	0.41
	LC4	0.59			
	LC5	0.76			
	LC7	0.70			
	LC8	0.71			
	LC9	0.53			
	LC10	0.63			
	LC11	0.54			
ASE	ASE1	0.93	0.96	0.89	0.81
	ASE2	0.88			
	ASE3	0.89			
AIA	AIA1	0.81	0.81	0.77	0.55
	AIA2	0.80			
	AIA3	0.78			
	AIA4	0.57			
PN	PN1	0.84	0.91	0.90	0.76
	PN2	0.91			
	PN3	0.86			
	PN4	0.89			
PV	PV1	0.92	0.90	0.90	0.84
	PV2	0.94			
	PV3	0.88			
BI	BI1	0.96	0.92	0.91	0.92
	BI4	0.96			

Table 4. Heterotrait-Monotrait ratio (HTMT)

	LC	PBC	AIA	PN	PV
PBC	0.15				
AIA	0.19	0.53			
PN	0.32	0.25	0.20		
PV	0.39	0.34	0.22	0.75	
BI	0.27	0.34	0.41	0.64	0.69

5.2. Structural Model (Hypotheses Testing)

The structural model was analyzed using the bootstrapping technique with 5,000 resamples to evaluate the significance and predictive power of the hypothesized relationships within the research model. Table 5 presents the path coefficients (β), t-values, p-values, f-square, variance inflation factor (VIF), and coefficients of determination (R^2) for each dependent variable, while Table 6 summarizes the total effects of each independent variable on the dependent variables.

The coefficient of determination (R^2) measures the proportion of variance in a dependent variable that is explained by the independent variables in the research model [84]. It is a critical metric for assessing the model's predictive power [85]. According to Falk and Miller [86] and Weidich and Bastiaens [87], an R^2 value exceeding 0.1 is considered indicative of an adequate level of explanation for the dependent variables.

The path coefficients (β) reflect the strength, direction, and significance of the relationships between independent and dependent variables, indicating the magnitude of the effects within the research model. The f-square values indicate the effect size of the

independent variables on the dependent variables, while the VIF values assess multicollinearity issues among the independent variables.

Table 5. The results of the structural model

Dependent variable	Independent variable	Path coefficient	t-value	p-value	f-square	VIF	R^2
BI	PN	0.30	2.83	0.005*	0.089	1.887	0.45
	PV	0.43	4.05	0.000*	0.567	1.887	
PN	LC	0.37	6.57	0.000*	0.153	1.026	0.22
	ASE	0.11	1.63	0.102	0.013	1.283	
	AIA	0.16	2.17	0.030*	0.022	1.254	
PV	PN	0.57	10.87	0.000*	0.166	1.238	0.54
	LC	0.19	3.72	0.000*	0.062	1.182	
	ASE	0.15	2.84	0.005*	0.036	1.299	
	AIA	0.02	0.27	0.788	0.000	1.282	

* $p < 0.05$

Table 6. The results of total effect

Dependent variable	Independent variable	Total effect	t-value	p-value
BI	PN	0.55	9.82	0.000*
	PV	0.43	4.12	0.000*
	LC	0.28	6.93	0.000*
	ASE	0.12	2.88	0.004*
	AIA	0.10	2.21	0.027*
PN	LC	0.37	6.87	0.000*
	ASE	0.11	1.63	0.102
	AIA	0.16	2.17	0.030*
PV	PN	0.57	10.80	0.000*
	LC	0.40	7.57	0.000*
	ASE	0.21	3.52	0.000*
	AIA	0.11	1.70	0.089

* $p < 0.05$

The results of the hypotheses testing are summarized in Fig 2, indicating the relationships between the variables and whether they are supported or not. The model explains a significant amount of variance in behavioral intention (45%), perceived need (22%), and perceived value (54%). Regarding the effects on behavioral intention, both perceived need ($\beta = 0.30$) and perceived value ($\beta = 0.43$) have positive and significant influences, supporting Hypotheses 1 and 2. In terms of perceived need, locus of control ($\beta = 0.37$) and AI anxiety ($\beta = 0.16$) are significant determinants, supporting Hypotheses 4 and 8. However, AI self-efficacy does not have a significant influence on perceived need, contradicting Hypothesis 6. For perceived value, perceived need ($\beta = 0.57$), locus of control ($\beta = 0.19$), and AI self-efficacy ($\beta = 0.15$) are significant predictors, supporting Hypotheses 3, 5, and 7. However, AI anxiety does not show a significant influence on perceived value, not supporting Hypothesis 9.

Overall, the results suggest that perceived need and perceived value are important factors influencing behavioral intention to use XAI. Locus of control and AI anxiety also play significant roles in shaping perceived need, while locus of control and AI self-efficacy contribute to perceived value. Furthermore, from Table 6, it can be observed that the strongest factor influencing perceived need is locus of control, while the strongest factor influencing perceived value is perceived need. Finally, the strongest factor affecting behavioral intention to use XAI is perceived need.

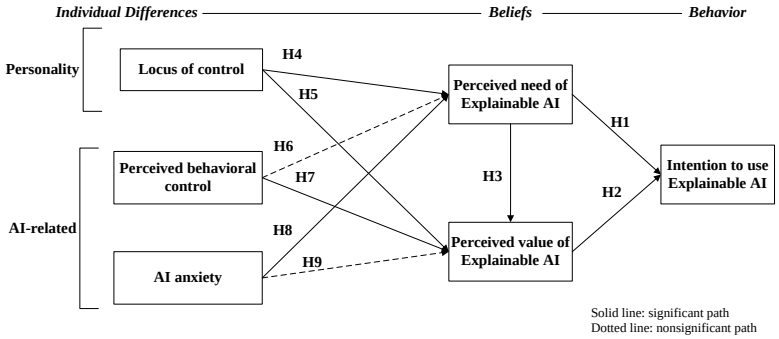


Fig. 2. Summary of hypotheses testing results

6. Discussion

6.1. The Influences of Perceived Needs and Perceived Value

The empirical findings of this study support Hypotheses 1 and 2, indicating that perceived value and perceived need for XAI positively influence the intention to use XAI, with perceived needs having the greatest total effect. Perceived need is recognized as a crucial determinant of user behavior and the intention to adopt innovations or change in the literature.

Perceived need refers to an individual's judgment regarding the necessity or benefits of a specific innovation or change. When individuals perceive a need for the innovation, they understand its potential benefits and view it as essential for themselves, which increases their intention to adopt it. This notion is supported by previous studies. Coulton and Frost [53], King and Teo [54], and Mukred and Singh [55] have emphasized the significance of perceived need as a driver of behavioral intention. In the context of XAI, perceived need addresses key concerns surrounding the opaque, "black-box" nature of AI algorithms, as well as the importance of transparency and interpretability. When users recognize the necessity of XAI in providing explanations and insights into AI decision-making processes, their intention to use XAI increases. Several studies further substantiate this perspective: Jeong et al. [56], Lee and Han [57], and Wang et al. [58] have all highlighted perceived need as critical to influencing user behavior. Similarly, Lin [88] demonstrated that users' intention to adopt mobile communication software is heightened when they perceive a need for it.

The findings of this study resonate with these insights, suggesting that users who appreciate the benefits and necessity of XAI—particularly in addressing transparency and interpretability concerns—exhibit a stronger intention to adopt it. This underscores the significant role perceived needs play in promoting the use of XAI as a solution to the challenges posed by complex AI systems.

Perceived value, as defined by Zeithaml [46], reflects the overall evaluation of the practicality and usefulness of an innovation from the user's perspective. When users perceive XAI as practically valuable and beneficial to them, it naturally increases their intention to use it. Dodds et al. [89] define perceived value as the ratio of perceived benefits to perceived sacrifices. If users perceive that the benefits of using XAI outweigh the sacrifices or costs associated with it, their perceived value of XAI increases. The findings align with the research of Liu et al. [52], which highlights that users with higher perceived value for XAI are more likely to demonstrate a stronger intention to adopt it. These results indicate that users' perceptions of the practicality, usefulness, and cost-benefit ratio of XAI significantly influence their behavioral intentions. When XAI is perceived as offering substantial benefits and a favorable cost-benefit ratio, users are more motivated to adopt it.

The empirical results of this study support Hypothesis 3, which suggests that perceived need for XAI positively impacts the perceived value of XAI, with the total effect being the greatest. While there is limited research specifically examining the relationship between perceived need and perceived value of innovations, existing empirical literature [56],[57],[58] suggests that perceived need does indeed positively influence perceived value. When potential users perceive a high need for a particular innovative product or service, they also tend to perceive it as having greater value and are more likely to adopt it. In the context of AI and algorithmic decision-making, the black-box nature and lack of transparency have been subjects of criticism and concern [91]. Users inherently desire to understand why and how AI systems or algorithms make decisions [92]. As AI systems and algorithms become more complex, they are often seen as "black boxes," which increases decision risks and requires expertise to comprehend their decisions or performance [90],[91].

This lack of transparency in complex AI systems hampers understanding and diminishes trust [91]. When trust in the outcomes of AI decreases, the need to understand AI decision-making or performance becomes more pronounced, thereby increasing the perceived value of XAI. In other words, when users demand XAI to trust the outcomes of AI, they perceive XAI as valuable. Therefore, the perceived need for XAI strongly and positively influences the perceived value of XAI. The perceived value, in turn, has a positive effect on the intention to use XAI, as evidenced by previous studies [42],[44],[50],[52],[91],[93].

6.2. The Influences of Locus of Control

The empirical results of this study provide support for Hypotheses 4 and 5, which propose that locus of control positively influences the perceived need and perceived value of XAI, with the greatest total effect on perceived need. Locus of control refers to individuals' beliefs about the extent to which they can control events in their lives [61],[94]. Individuals with a high degree of internal locus of control believe that they have control over events in their lives and can influence their surrounding environment [95]. When individuals with an internal locus of control consider whether to use AI, they are more inclined to seek an understanding of how AI makes decisions. They believe that they have the ability to comprehend and influence the outcomes, leading to a higher perceived need for XAI. These individuals perceive XAI as valuable because it

aligns with their desire for control and understanding. Conversely, individuals with an external locus of control feel that they have little control their lives and are less likely to seek explanations for the decision-making process of AI. They may simply accept the use of AI without questioning or desiring explanations. As a result, their need for XAI is lower, and they are less likely to perceive XAI as valuable.

In summary, individuals with an internal locus of control have a higher need for explanations of AI results and perceive greater value in XAI compared to those with an external locus of control.

6.3. The Influences of AI Self-efficacy

The findings support for Hypothesis 7, indicating that AI self-efficacy has a significant positive impact on the perceived value of XAI. However, the results do not support Hypothesis 6, suggesting that AI self-efficacy does not significantly impact the perceived needs of XAI.

AI self-efficacy refers to an individual's belief and perception of their capability to perform specific tasks or master new technologies [66],[67],[96]. Individuals with high self-efficacy in AI have confidence in their abilities and resources related to AI manipulation and its outcomes. As a result, their demand for XAI may not be significant, as they possess the necessary knowledge and skills to navigate AI effectively. Nevertheless, these individuals place significant value on understanding how AI makes decisions. They appreciate the importance of explainability and transparency, as these features enable them to comprehend AI systems more effectively. Transparency, a core characteristic of XAI, facilitates user understanding by providing clear explanations of AI decision-making processes [91],[97].

The findings indicate that individuals with high AI self-efficacy perceive XAI as valuable because it aligns with their desire for understanding and control. Transparency offered by XAI further enhances their appreciation of AI systems, reinforcing their positive perception of XAI’s value.

In summary, while individuals with high AI self-efficacy may not express a heightened need for XAI, they recognize its value in fostering understanding and transparency, which supports their confidence in engaging with AI technologies.

6.4. The Influences of AI Anxiety

The empirical results confirm Hypothesis 8, suggesting that AI anxiety increases perceived needs for XAI. However, the findings do not support Hypothesis 9, indicating that AI anxiety does not significantly influence the perceived value of XAI.

AI anxiety refers to the fear or agitation individuals experience regarding the control or lack thereof over AI [71]. It can lead people to reduce or avoid using AI due to their emotional response of anxiety or fear, which hinders their interaction with AI [72]. However, XAI addresses the black-box paradox of AI by providing transparency and explainability. XAI allows users to understand the inner workings of machine learning algorithms, even with limited technical knowledge [98].

The study aligns with prior research [98], demonstrating that XAI can enhance transparency, trust, and user adoption by addressing concerns about AI decision-making processes. Users with higher AI anxiety and lower trust in AI are more likely to perceive a need for XAI, as it provides the transparency they require to build confidence in AI systems. However, individuals with AI anxiety often view AI as a potential threat, associating it with fears of replacement or loss of control. These concerns may hinder their ability to recognize the value of XAI, as they perceive it as part of the broader AI ecosystem that evokes their apprehension. To address this, it is crucial to emphasize the benefits and transparency offered by XAI, helping users understand its role in reducing risks and increasing trust in AI systems.

In summary, individuals with AI anxiety perceive a strong need for XAI due to their desire for transparency and understanding. However, their concerns about AI as a whole may obscure their recognition of XAI's value. Efforts to highlight the advantages of XAI and address their apprehensions are essential to encourage its adoption and use.

7. Conclusions

7.1. Conclusions

Understanding the factors that influence potential users' intention to adopt explainable AI (XAI) is essential for promoting its development and widespread acceptance. By identifying these factors, developers and researchers can design XAI systems that better meet user needs and preferences, thereby enhancing their acceptance and utilization.

This study developed a research model grounded in the characteristics of XAI and prior studies on technology acceptance, with a particular focus on individual factors. Using data from 252 potential XAI users, the model demonstrated strong explanatory power, accounting for 45% of the variance in users' intention to adopt XAI. Key findings include:

Technology Acceptance Model (TAM) Validation. The study confirms the causal chain proposed by TAM, demonstrating that external stimuli influence beliefs, which in turn shape users' intention to adopt technology. For XAI, perceived value and perceived need emerged as critical determinants of adoption intentions. Additionally, the study extended the TAM framework by incorporating individual difference variables, including locus of control, AI self-efficacy, and AI anxiety.

Perceived Needs as a Driving Force. Among the determinants, perceived needs exerted the strongest influence on perceived value. When users recognize a significant need for XAI, their perception of its value is enhanced, ultimately driving their intention to adopt it.

Role of Internal Locus of Control. Individuals with an internal locus of control—those who believe they can influence events in their lives—exhibited higher perceived needs and perceived value for XAI. This suggests that their sense of agency fosters a stronger alignment with XAI's transparency and explainability.

Differentiated Effects of AI Self-Efficacy. While AI self-efficacy positively influenced perceived value, it did not significantly affect perceived needs. This indicates that confidence in one's ability to use AI enhances appreciation for its value but does not directly increase the recognition of XAI's necessity.

Contrasting Effects of AI Anxiety. AI anxiety positively influenced perceived need but had no significant effect on perceived value. Individuals experiencing anxiety about AI recognize a need for XAI to mitigate their concerns but may struggle to appreciate its broader value due to apprehension about AI technology.

In summary, perceived value and perceived need are pivotal in driving users' intention to adopt XAI. Individual factors such as locus of control, AI self-efficacy, and AI anxiety play significant roles in shaping these perceptions, providing valuable insights for XAI development and promotion.

7.2. Implications

The findings of this study have important implications for both academics and practitioners in the XAI field. Here are some key implications based on the provided information.

Academic Implications. This study reinforces the causal chain proposed by the Technology Acceptance Model (TAM), which posits that external factors influence beliefs, which subsequently shape the intention to use a technology. By validating this framework, the study provides robust support for the relevance and applicability of TAM in understanding user acceptance of XAI. Future research can build on this foundation to examine the acceptance of other innovative technologies and further explore the adaptability of TAM across different contexts.

Moreover, the study highlights the importance of individual differences, such as personality traits and human-technology-related characteristics, in shaping users' perceptions of technology. It underscores that beliefs such as perceived needs and perceived value are critical determinants of users' intentions to adopt XAI. These findings suggest that future research should place greater emphasis on individual differences when examining the adoption of innovative technologies, as these factors play a pivotal role in influencing users' decision-making processes.

Practical Implications. The findings of this study offer valuable insights for practitioners aiming to address the primary concerns of potential XAI users and enhance their acceptance intentions. By understanding the critical roles of perceived need and perceived value, practitioners can formulate strategies to strengthen user acceptance. These strategies may include designing user-friendly interfaces, providing clear and transparent explanations of AI reasoning, and emphasizing the tangible benefits and value that XAI delivers.

Additionally, the study highlights the importance of internal locus of control. Individuals with a strong internal locus of control—those who believe they can influence events affecting them—exhibit higher perceived need and perceived value for XAI. These individuals are more likely to recognize the benefits of XAI and demonstrate a greater willingness to use it. Practitioners can leverage this insight by involving such individuals in the design, evaluation, and promotion of XAI. Their feedback can help ensure that XAI systems align with user preferences for control and transparency.

In summary, this study provides actionable insights for both practitioners and academics in the development and promotion of XAI. It emphasizes the significance of internal locus of control in shaping users' acceptance intentions and confirms the applicability of TAM's causal chain in understanding technology adoption. Furthermore, it underscores the influence of individual differences on user perceptions, highlighting the need for continued research in this area to refine strategies and improve user-centered designs.

7.3. Limitations

The study acknowledges several limitations that should be taken into consideration:

Limited sample size and generalizability. The research was conducted with a relatively small sample of potential AI users in Taiwan. This limits the generalizability of the results to other populations and contexts. To validate and expand upon these findings, future studies should include larger, more diverse samples from various countries and cultural backgrounds.

Early stage of XAI development. Explainable AI (XAI) applications are still in the early stages of their development and adoption. The general population's limited familiarity and understanding of XAI may have influenced participants' intentions to adopt it. As XAI technology matures and gains broader recognition, future research should investigate adoption dynamics at different stages of XAI development to account for these evolving perspectives.

Lack of differentiation among XAI applications. This study examined XAI adoption as a general concept, without differentiating between specific types of XAI applications. However, user concerns and acceptance factors may vary significantly depending on the application's context and purpose. Future research should explore adoption intentions across various XAI applications, identifying similarities and differences to provide more tailored insights.

These limitations underscore the need for further research to enhance the understanding of XAI adoption. By incorporating larger, more diverse samples, considering the evolving nature of XAI, and examining application-specific factors, future studies can strengthen the validity and applicability of findings, enabling more informed decision-making and practical implementations.

References

1. Johnson, D. G., Verdicchio, M.: Reframing AI Discourse. Minds and Machines, Vol. 27, No.4, 575-590. (2017)
2. Sohn, K., Kwon. O.: Technology Acceptance Theories and Factors Influencing Artificial intelligence-based Intelligent Products. Telematics and Informatics, Vol.47, 101324. (2020)
3. H.Jarrahi. M.: Artificial Intelligence and The Future of Work: Human-AI Symbiosis in Organizational Decision Making. Business Horizons, Vol.61, No.4, 577-586. (2018)
4. Oracle. Restaurant 2025: Emerging Technologies Destined to Reshape Our Business. Available Online:

https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc36/Restaurant-2025-Oracle-Hospitality.pdf. (accessed on 19 January 2021).
5. Hoffman, R.R., Klein, G., Mueller, S.T.: Explaining Explanation for Explainable AI. Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting.
6. Wierzynski, C.: The Challenges and Opportunities of Explainable AI. Available Online: https://ai.intel.com/the-challenges-and-opportunities-of-explainable-ai/ (accessed on 12 January 2018)
7. Holzinger, A., Carrington, A., Müller, H.: Measuring the Quality of Explanations: The System Causability Scale (SCS). KI-Künstliche Intelligenz, 1-6. (2020)
8. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., Müller, K. R. (Eds.): Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer Nature, 11700. (2019)
9. Hind, M., Wei, D., Campbell, M., Codella, N. C., Dhurandhar, A., Mojsilović, A., Varshney, K. R.: TED: Teaching AI to Explain Its Decisions. In Proceedings of The 2019 AAAI/ACM Conference on AI, Ethics, and Society, 123-129. (2019)
10. Gunning, D., Aha, D.: DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, Vol.40, No.2, 44-58. (2019)
11. Bucinca, Z., Phoebe, L., Krzysztof Z. Gajos, Elena Glassman, L.: Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating Explainable AI Systems. In IUI'20: ACM Proceedings of the 25th Conference on Intelligent User Interfaces, March pages, 17–20. Cagliari, Italy. ACM, New York, NY, USA, 11. (2020)
12. Gorry, G. A., Morton, M. S. S.: A Framework for Management Information Systems. Sloan Management Review, Vol.13, 55-70. (1971)
13. Johnston, M.E., Langton, K.B., Brian Haynes, R., Mathieu, A.: Effects of Computer-Based Clinical Decision Support Systems on Clinician Performance and Patient Outcome: A Critical Appraisal of Research. Annals of internal medicine, Vol.120, No.2, 135–142. (1994)
14. Vera Liao, Q., Singh, M., Yunfeng Zhang, and K.E. Bellamy, R.: Introduction to Explainable AI. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April, 25-30. (2020)
15. Doshi-Velez, F., Kim, B.: Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv, Vol.1702, 08608. (2017)
16. Arnold, C. K., Chaunce, K., Krzysztof, Z. Gajos.: Predictive Text Encourages Predictable Writing. In Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20). ACM, New York, NY, USA. (2020)
17. Hoffman, R.R. Shane T Mueller, Klein, G., Litman, J.: Metrics for Explainable AI: Challenges and Prospects. arXiv preprint arXiv, Vol.1812, 04608. (2018)
18. Lakkaraju H., Bastani, O.: How Do I Fool You: Manipulating User Trust via Misleading Black Box Explanations. arXiv preprint arXiv, Vol.1911, 06473. (2011)
19. Hagras, H.: Toward Human-Understandable. Explainable AI. Computer, Vol.51, No.9,28-36. (2018)
20. Lundberg, S. M., Lee, S. I.: A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, Vol.30. (2017)
21. Antwarg, L., Miller, R. M., Shapira, B., Rokach, L.: Explaining Anomalies Detected by Autoencoders Using Shapley Additive Explanations. Expert Systems with Applications, Vol.186, 115736. (2021)
22. Jabeur, S. B., Mefteh-Wali, S., Viviani, J. L.: Forecasting Gold Price with The XGBoost Algorithm and SHAP Interaction Values. Annals of Operations Research, Vol.334, 679–699. (2024)
23. Kaur, B. P., Singh, H., Hans, R., Sharma, S. K., Sharma, C., Hassan, M. M.: A Genetic Algorithm Aided Hyper Parameter Optimization Based Ensemble Model for Respiratory Disease Prediction With Explainable AI. PLoS ONE, Vol. 19, No. 12, e0308015. (2024)

24. Tan, B., Gan, Z., Wu, Y.: The Measurement and Early Warning of Daily Financial Stability Index Based on XGBoost and SHAP: Evidence from China. Expert Systems with Applications, Vol. 227, 120375. (2023)
25. Nylén-Forthun, E., Møller, M., Abrahamsen, N. G. B.: Financial Distress Prediction Using Machine Learning and XAI: Developing An Early Warning Model for Listed Nordic Corporations (Master's thesis, NTNU). (2022)
26. Wu, C. F., Zhang, K., Lin, M. C., Chiou, C. C.: Predicting Consumer Electronics E-Commerce: Technology Acceptance Model and Logistics Service Quality. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 8, No. 7, 66-85. (2024)
27. Yang, C., Chen, M., Yuan, Q.: The Application of XGBoost and SHAP to Examining the Factors in Freight Truck-Related Crashes: An Exploratory Analysis. Accident Analysis & Prevention, Vol. 158, 106153. (2021)a
28. Yang, H., Li, E., Cai, Y. F., Li, J., Yuan, G. X.: The Extraction of Early Warning Features for Predicting Financial Distress Based on XGBoost Model and Shap Framework. International Journal of Financial Engineering, Vol. 8, No. 3, 2141004. (2021)b
29. Rizk-Allah, R. M., Abouelmagd, L. M., Darwish, A., Snasel, V., Hassanien, A. E.: Explainable AI and Optimized Solar Power Generation Forecasting Model Based on Environmental Conditions. PLoS ONE, Vol. 19, No. 10, 1-33. (2024)
30. Rajabi, E., Etminani, K.: Knowledge-Graph-Based Explainable AI: A Systematic Review. Journal of Information Science, Vol. 50, No. 4, 1019-1029. (2024)
31. Chinu, C. S., Bansal, U.: Explainable AI: To Reveal the Logic of Black-Box Models. New Generation Computing, Vol. 42, No. 1, 53-87. (2024)
32. Wang, D., Bian, C., Chen, G.: Using Explainable AI to Unravel Classroom Dialogue Analysis: Effects of Explanations on Teachers' Trust, Technology Acceptance and Cognitive Load. British Journal of Educational Technology, Vol. 55, No. 6, 2530-2556. (2024)
33. Sano, T., Shi, J., Kawabata, H.: The Differences in Essential Facial Areas for Impressions between Humans and Deep Learning Models: An Eye-Tracking and Explainable AI Approach. British Journal of Psychology, 1-26. doi: 10.1111/bjop.12744. (2024)
34. Ebermann, C., Selisky, M., Weibelzahl, S.: Explainable AI: The Effect of Contradictory Decisions and Explanations on Users' Acceptance of AI Systems. International Journal of Human-Computer Interaction, Vol. 39, No. 9, 1807-1826. (2023)
35. Davis, F. D., Venkatesh, V.: A Critical Assessment of Potential Measurement Biases in the Technology Acceptance Model: Three Experiments. International journal of human-computer studies, Vol. 45, No. 1: 19-45. (1996)
36. Agarwal, R., Prasad, J.: Are Individual Differences Germane to the Acceptance of New Information Technologies? Decision sciences, Vol. 30, No. 2, 361-391. (1999)
37. Hong, W., Thong, J. Y., Wong, W. M., Tam, K. Y.: Determinants of User Acceptance of Digital Libraries: An Empirical Examination of Individual Differences and System Characteristics. Journal of management information systems, Vol. 18, No. 3, 97-124. (2002)
38. Wang, C., Wang, S.: Study on Some Key Problems Related to Distributed Generation Systems. Automation of Electric Power Systems, Vol. 20, No. 32, 1-4. (2008)
39. Wang, Y. S., Wang, H. Y., Lin, H. H.: Investigating the Mediating Role of Perceived Playfulness in the Acceptance of Hedonic Information Systems. In 2009 Proceedings of the 13th WSEAS International Conference on SYSTEMS, Stevens Point, Wisconsin, United States, 322-327. (2009)
40. Wang, Y. S., Lin, H. H., Liao, Y. W.: Investigating the Individual Difference Antecedents of Perceived Enjoyment in Students' Use of Blogging. British Journal of educational technology, Vol. 43, No. 1, 139-152. (2012)
41. Kim, H. W., Chan, H. C., Gupta, S.: Value-Based Adoption of Mobile Internet: An Empirical Investigation. Decision support systems, Vol. 43, No. 1, 111-126. (2007)
42. Chung, N., Koo, C.: The Use of Social Media in Travel Information Search. Telematics and Informatics, Vol. 32, No. 2, 215-229. (2015)

43. Wang, Y. Y., Lin, H. H., Wang, Y. S., Shih, Y. W., Wang, S. T.: What Drives Users' Intentions to Purchase a GPS Navigation App: The Moderating Role of Perceived Availability of Free Substitutes. Internet Research, Vol. 28, No. 1, 251-274. (2018)
44. Lim, W. M., Yong, J. L. S., Suryadi, K.: Consumers' Perceived Value and Willingness to Purchase Organic Food. Journal of Global Marketing, Vol. 27, No. 5, 298-307. (2014)
45. Zeithaml, V. A.: Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. Journal of marketing, Vol. 52, No. 3, 2-22. (1988)
46. Koo, D. M.: The Moderating Role of Locus of Control on the Links between Experiential Motives and Intention to Play Online Games. Computers in Human Behavior, Vol, 25, No. 2, 466-474. (2009)
47. Lin, T. C., Wu, , S., Hsu, J. S. C., Chou, Y. C.: The Integration of Value-Based Adoption and Expectation–Confirmation Models: An Example of IPTV Continuance Intention. Decision Support Systems, Vol. 54, No.1, 63-75. (2012)
48. Wang, Y. S., Yeh, C. H., Liao, Y. W.: What Drives Purchase Intention in the Context of Online Content Services? The Moderating Role of Ethical Self-Efficacy for Online Piracy. International Journal of Information Management, Vol. 33, No. 1, 199-208. (2013)
49. Wang, Y. S., Tseng, T. H., Wang, W. T., Shih, Y. W., Chan, P. Y.: Developing and Validating a Mobile Catering App Success Model. International Journal of Hospitality Management, Vol. 77, 19-30. (2019)
50. Yin, J., Qiu, X.: AI Technology and Online Purchase Intention: Structural Equation Model Based on Perceived Value. Sustainability, Vol. 13, No. 10, 5671. (2021)
51. Liu, C. F., Chen, Z. C., Kuo, S. C., Lin, T. C.: Does AI Explainability Affect Physicians' Intention to Use AI? International Journal of Medical Informatics, Vol. 168, 104884. (2022)
52. Coulton, C., Frost, A. K.: Use of Social and Health Services by the Elderly. Journal of Health and social Behavior, Vol. 23, No. 4, 330-339. (1982)
53. King, W. R., Teo, T. S.: Facilitators and Inhibitors for the Strategic Use of Information Technology. Information and Management, Vol. 27, No. 2, 71-87. (1994)
54. Mukred, A., Singh, D., Safie, N.: Investigating the Impact of Information Culture on the Adoption of Information System in Public Health Sector of Developing Countries. International Journal of Business Information Systems, Vol. 24, No. 3, 261-284. (2017)
55. Jeong, N., Yoo, Y., Heo, T. Y.: Moderating Effect of Personal Innovativeness on Mobile-RFID Services: Based on Warshaw's Purchase Intention Model. Technological Forecasting and Social Change, Vol. 76, No. 1, 154-164. (2009)
56. Lee, E., Han. S.: Determinants of Adoption of Mobile Health Services. Online Information Review, Vol. 39, No. 4, 556-573. (2015)
57. Wang, Y. Y., Wang, Y. S., Lin, H. H., Tsai, T. H.: Developing and Validating a Model for Assessing Paid Mobile Learning App Success. Interactive Learning Environments, Vol. 27, No. 4, 458-477. (2018)
58. Hsia. C. T.: The Classic Chinese Novel: A Critical Introduction. The Chinese University of Hong Kong Press. (2016)
59. Wang, Y. D., Hsieh, H. H.: Toward a Better Understanding of the Link between Ethical Climate and Job Satisfaction: A Multilevel Analysis. Journal of business ethics, Vol. 105, No. 4, 535-545. (2012)
60. Singh, J., Dubey, A. K., Singh. R. P.: Antarctic Terrestrial Ecosystem and Role of Pigments in Enhanced UV-B Radiations. Reviews in Environmental Science and Bio/Technology, Vol. 10, No. 1, 63-77. (2011)
61. Spector, P. E.: Behavior in Organizations as a Function of Employee's Locus of Control. Psychological bulletin, Vol 91, No. 3, 482-497. (1982)
62. Hoffman, R. L., Norris, B. J., Wager, J. F.: ZnO-Based Transparent Thin-Film Transistors. Applied Physics Letters, Vol. 82, No. 5, 733-735. (2003)

63. Albashrawi, M., Alashoor, T.: Entrepreneurial Intention: The Impact of General Computer Self-Efficacy and Computer Anxiety. Interacting with Computers, Vol. 32, No. 2, 118-131. (2020)
64. Huang, H. M., Liaw, S. S.: Exploring Users' Attitudes and Intentions Toward the Web as a Survey Tool. Computers in human behavior, Vol. 21, No. 5, 729-743. (2005)
65. Compeau, D. R., Higgins. C. A.: Computer Self-Efficacy: Development of a Measure and Initial Test. MIS quarterly, Vol. 19, No. 2, 189-211. (1995)
66. Holden, H., Rada, R.: Understanding the Influence of Perceived Usability and Technology Self-Efficacy on Teachers' Technology Acceptance. Journal of Research on Technology in Education, Vol. 43, No. 4, 343-367. (2011)
67. Chi, M., Henning, C., Khanna, S. K.: Factors Associated with School Teachers' Perceived Needs and Level of Adoption of HIV Prevention Education in Lusaka, Zambia. International Electronic Journal of Health Education, Vol. 14, 1-15. (2011)
68. Zhang, Y., Espinoza, S.: Relationships among Computer Self-Efficacy, Attitudes toward Computers, and Desirability of Learning Computing Skills. Journal of research on Computing in Education, Vol. 30, No. 4, 420-436. (1998)
69. Hong, Z., Ueguchi-Tanaka, M., Shimizu-Sato, S., Inukai, Y., Fujioka, S., Shimada, Y., Matsuoka, M.: Loss-of-Function of a Rice Brassinosteroid Biosynthetic Enzyme, C-6 Oxidase, Prevents the Organized Arrangement and Polar Elongation of Cells in the Leaves and Stem. The Plant Journal, Vol. 32, No. 4, 495-508. (2002)
70. Johnson, D. G., Verdicchio, M.: AI Anxiety. Journal of the Association for Information Science and Technology, Vol. 68, No. 9, 2267-2270. (2017)
71. Wang, J., Wang, S.: Preparation, Modification and Environmental Application of Biochar: a Review. Journal of Cleaner Production, Vol. 227, 1002-1022. (2019)
72. Seol, S., Lee, H., Zo, H.: Exploring Factors Affecting the Adoption of Mobile Office in Business: an Integration of TPB with Perceived Value. International Journal of Mobile Communications, Vol. 14, No. 1, 1-25. (2016)
73. Taylor, S., Todd, P.: Assessing IT Usage: The Role of Prior Experience. MIS quarterly, Vol. 19, No. 4, 561-570. (1995)
74. Lin, T. T., Bautista J,. R.: Content-Related Factors Influence Perceived Value of Location-Based Mobile Advertising. Journal of Computer Information Systems, Vol. 60, No. 2, 184-193. (2018)
75. Tsou, W. L., Huang, Y. H.: The Effect of Explicit Instruction in Formulaic Sequences on Academic Speech Fluency. Taiwan International ESP Journal, Vol. 4, No. 2, 57-80. (2012)
76. Untaru, E. N., Ispas, A., Candrea, A. N., Luca, M., Epuran, G.: Predictors of Individuals' Intention to Conserve Water in a Lodging Context: The Application of An Extended Theory of Reasoned Action. International Journal of Hospitality Management, Vol. 59, 50-59. (2016)
77. Dooley, D.: Social Research Methods. Prentice Hall: Upper Saddle River, NJ. (2001)
78. Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Thiele, K. O.: Mirror, Mirror on the Wall: a Comparative Evaluation of Composite-Based Structural Equation Modeling Methods. Journal of the Academy of Marketing Science, Vol. 45, No. 5, 616-632. (2019)
79. Wong, K. K. K.: Mastering Partial Least Squares Structural Equation Modeling (PLS-Sem) with Smartpls in 38 Hours. IUniverse. (2019)
80. Cheung, G. W., Wang, C.: Current Approaches for Assessing Convergent and Discriminant Validity with SEM: Issues and Solutions. In Academy of management proceedings, 2017(1):12706. Briarcliff Manor, NY 10510: Academy of Management. (2027)_
81. Fornell, C., Larcker, D. F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of marketing research, Vol. 18, No. 1, 39-50. (1981)
82. Lam. L. W.: Impact of Competitiveness on Salespeople's Commitment and Performance. Journal of Business Research, Vol. 65, No. 9, 1328-1334. (2012)

83. Hair Jr, J. F., Sarstedt, M., Ringle, C. M., Gudergan, S. P.: Advanced Issues in Partial Least Squares Structural Equation Modeling. SAGE Publications. (2017)
84. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E.: Multivariate Data Analysis (8th ed.). Boston: Cengage. (2019)
85. Falk, M., Miller, A. G.: Infrared Spectrum of Carbon Dioxide in Aqueous Solution. Vibrational spectroscopy, Vol. 4, No. 1, 105-108. (1992)
86. Weidlich, J., Bastiaens, T. J.: Explaining Social Presence and the Quality of Online Learning with the SIPS Model. Computers in Human Behavior, Vol. 72, 479-487. (2017)
87. Lin, K. Y.: User Communication Behavior in Mobile Communication Software. Online Information Review, Vol. 40, No. 7, 1071-1089. (2016)
88. Dodds, W. B., Monroe, K. B., Grewal, D.: Effects of Price, Brand, and Store Information on Buyers' Product Evaluations. Journal of Marketing Research, Vol. 28, No. 3, 307-319. (1991)
89. Castelvecchi, D.: Can We Open the Black Box of AI? Nature, Vol. 538, 20-23. (2016)
90. Shin, D.: The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI. International Journal Human–Computer Studies, Vol. 146, 102551. (2021)
91. Shin, D., Park, Y.: Role of Fairness, Accountability, and Transparency in Algorithmic Affordance. Computers in Human Behavior, Vol. 98, 277-284. (2019)
92. Shin, D., Zhong, B., Biocca, F.: Beyond User Experience: What Constitutes Algorithmic Experiences. International Journal Information Management, Vol. 52, 1-11. (2020)
93. Nykänen, M., Salmela-Aro, K., Tolvanen, A., Vuori, J.: Safety Self-Efficacy and Internal Locus of Control as Mediators of Safety Motivation - Randomized Controlled Trial (RCT) Study. Safety Science, Vol. 117, 330-338. (2019)
94. Sharan, N. N., Romano, D. M.: The Effects of Personality and Locus of Control on Trust in Humans versus Artificial Intelligence. Heliyon, Vol. 6, No. 8, e04572. (2020)
95. Smith, E. C., Starratt, G. K., McCrink, C. L., Whitford, H.: Teacher Evaluation Feedback and Instructional Practice Self-Efficacy in Secondary School Teachers. Educational Administration Quarterly, Vol. 56, No. 4, 671-701. (2020)
96. Haque, A. K. M. B., Islam, A. K. M. N., Mikalef, P.: Explainable Artificial Intelligence (XAI) from a User Perspective: A Synthesis of Prior Literature and Problematizing Avenues for Future Research. Technological Forecasting and Social Change, Vol. 186, 122120. (2023)
97. Bernardo, E. l., Seva, R. R.: Exploration of Explainable AI for Trust Development on Human-AI Interaction. In 2023 6th Artificial Intelligence and Cloud Computing Conference (AICCC) (AICCC 2023), Kyoto, Japan. ACM, New York, NY, USA. https://doi.org/10.1145/3639592.363962 (2023)

Yu-Min Wang is a Distinguished Professor in the Department of Information Management at National Chi Nan University, Taiwan. His research interests include e-learning, business education, and information management. He is a former Chairman for the Applied Science Education Discipline in the Ministry of Science and Technology of Taiwan.

Chei-Chang Chiou is a Distinguished Professor in the Department of Accounting at National Changhua University of Education, Taiwan. His research interests include educational technology, teaching strategy, concept mapping strategy, multimedia instruction, capital market, and empirical accounting study.

Received: October 18, 2024; Accepted: January 13, 2025.

Computer Science and Information Systems 22(3):1105–1119 https://doi.org/10.2298/CSIS241103035H

Development of an Explainable AI-Based Disaster Casualty Triage System

Po-Hsuan Hsiao[1], Ming-Yen Chen[2], Hsien-Cheng Liao[3], Ching-Cheng Lo[4], and Hsin-Te Wu[4]

[1] National Taiwan University,
106319 Taipei, Taiwan (R.O.C.)
d08944006@csie.ntu.edu.tw
[2] Industrial Technology Research Institute,
310401 Hsinchu, Taiwan (R.O.C.)
benjamin@itri.org.tw
[3] Institute for Information Industry,
106 Taipei, Taiwan (R.O.C.)
hcliao@iii.org.tw
[4] Dept. Computer Science and Information Engineering, National Taitung University,
950309 Taitung, Taiwan (R.O.C.)
roger20011009@gmail.com
wuhsinte@nttu.edu.tw

Abstract. Disaster response and management are critical components of rescue team training in countries worldwide. In addition to conducting various disaster drills, rescue teams are trained to perform casualty triage in simulated scenarios, allowing medical personnel to provide optimal treatment based on triage classifications. Due to the necessity of adapting disaster scenarios to enable responders to handle diverse disaster sites, each scenario must be interactive, helping rescuers understand how to perform triage effectively during disaster response. To enrich the variety of scenarios, AI can now be utilized for scenario design. However, for more rational script creation, generative AI must be grounded in Explainable AI (XAI) to make the generation process transparent, thus enhancing the scenario's realism.

This paper proposes an XAI-based disaster casualty triage scenario system. The system generates scenarios through generative AI, utilizing XAI to ensure data transparency. The primary output is a simulation training game focused on disaster scenarios, developed on the Unity platform to build realistic accident scenes. The goal is to provide frontline firefighters with immersive training to strengthen their on-site response and emergency handling skills. The game incorporates a triage mechanism that guides users to categorize injuries based on symptoms and apply appropriate medical actions, aiming to minimize casualties during disasters. From an educational perspective, this game provides the general public with an understanding of how firefighters perform triage based on injury symptoms in emergencies, ensuring that each casualty receives necessary medical support within the golden rescue window. Through simulation and decision-making training in the game, users enhance their judgment and responsiveness, further improving their rapid reaction and handling skills in disaster scenarios.

Keywords: Explainable AI, Virtual Reality, Disaster Casualty Triage System, Generative AI.

1. Introduction

Frequent natural disasters pose significant threats to human life and property, making the assessment of building damage a crucial indicator for determining post-disaster emergency response and resource allocation. This indicator enables rescue teams to formulate action plans and allocate resources with greater precision, ensuring maximum rescue efficiency and reducing the long-term impacts of disasters on social stability and the economy [14]. In recent years, the frequency and intensity of global natural disasters have markedly increased, presenting unprecedented challenges to human societies. Between 1980 and 2020, there was an average of seven natural disasters each year that caused over $1 billion in economic losses—a figure of considerable significance. However, in the past five years, this average has sharply risen to 16.2 incidents per year, highlighting the impact of climate change and environmental degradation on the frequency of natural disasters [13]. This trend poses serious threats to the economy, infrastructure, public health, and social stability, further complicating global emergency management and disaster prevention. Governments and international organizations are under greater pressure to not only enhance disaster early warning systems but also to allocate more resources to disaster prevention and mitigation measures to lessen the long-term effects of disasters. As the threat of natural disasters continues to grow, increasing global collaboration and response capabilities, as well as pursuing climate change mitigation strategies, has become a critical priority. Housing reconstruction following catastrophic events is a complex and challenging task. Severe weather events, particularly in coastal areas, pose significant threats to local residents. These storms often cause widespread damage to local structures, displacing large numbers of families and subjecting them to trauma, suffering, and psychological distress [10]. The housing recovery process requires not only substantial financial and resource investment but also psychological and social support to help affected residents rebuild stable living environments. Studies have conducted in-depth analyses of natural disaster chains, including earthquake chains, geological disaster chains, typhoon chains, and snow and rain disaster chains, examining the evolution and characteristics of specific disaster chains.[9][24] These studies reveal the interactions and cascading effects between various disasters, helping to provide a more comprehensive understanding of the complexities involved. On the other hand, [21] developed six major types of marine disaster chains, summarizing the impact characteristics of marine disasters. This paper underscores the research motivation derived from the increasing frequency and severity of global natural disasters, highlighting the urgent need for innovative training methods to improve disaster response efficiency. It also showcases the distinctive features of the proposed system, including the integration of Generative AI and Explainable AI for transparent and realistic disaster scenario creation, the use of immersive digital games and virtual reality for hands-on training, and the deployment of a triage mechanism with real-time feedback to enhance decision-making accuracy and engagement among emergency responders and the general public.

The term ”triage” is derived from the French verb ”trier,” which means to categorize, rank, or select. Originally, in the 17th and 18th centuries, it was primarily applied in agriculture and commerce, particularly in the quality grading of goods such as wool and coffee. Over time, this concept gradually expanded into the medical field and became essential in casualty care, eventually evolving into a systematic method for medical triage. With advancements in medical technology and shifting needs, triage expanded beyond the

military and was adopted in public healthcare services, becoming a critical component of emergency medical systems. In modern healthcare systems, emergency triage protocols are widely used in hospitals, disaster sites, and other medical settings, classifying patients into different priority levels based on the severity and urgency of their condition. For instance, in emergency medicine, triage personnel typically categorize patients into five levels: Level 1 is for critically ill patients needing immediate treatment; Level 2 is for patients who require prompt care but can tolerate a slight delay; Level 3 is for patients needing treatment that is not immediately urgent; Level 4 is for those with minor conditions; and Level 5 is for cases with no urgent medical needs. This tiered system facilitates the efficient allocation of medical resources, ensuring that critical patients receive timely medical assistance, thereby enhancing overall treatment efficiency. In large-scale disasters or public health crises, the triage system plays an even more vital role. When medical resources are rapidly depleted due to high demand, triage systems help rescue teams quickly assess the situation, prioritize patients for immediate care, and allocate resources according to actual needs. This system not only improves the emergency response capability of medical services but also reduces the burden on rescue personnel, allowing them to perform rescue operations more swiftly. Against a backdrop of globalization, with continuous refinement of medical triage systems and international cooperation in rescue efforts, the concept and applications of triage will become increasingly diversified and systematic.

The application of digital games in disaster education has opened up new teaching methods, immersing learners in highly realistic virtual environments that help concretize abstract knowledge. Unlike traditional book-based learning, digital games allow for flexible adjustments to the scope of the interactive environment, providing firefighters with a diverse learning platform that enhances their innovation and practical application abilities. With the rapid advancement of technology and the increasing maturity of hardware and software, the proposed digital games have gradually entered the field of education, particularly in medical and rescue simulation training. The integration of virtual reality (VR) and digital games significantly enhances trainees' practical skills and response speeds.

This paper proposes an innovative teaching approach based on digital games, which includes the following steps: First, design digital game scenarios and develop instructional modules that align with real-world situations. For example, in medical and rescue simulations, game scenarios can simulate emergency situations such as cardiopulmonary resuscitation (CPR) or trauma care, allowing trainees to repeatedly practice in-game and strengthen their ability to respond in real-life situations. Second, create an immersive learning environment through VR technology, allowing trainees to practice in a risk-free virtual space and familiarize themselves with procedures under pressure. Finally, incorporate a real-time feedback mechanism within the digital game, enabling trainees to receive immediate feedback after each operation, clearly identifying their strengths and areas for improvement. This real-time feedback also provides instructors with objective data to evaluate trainees' learning outcomes and further develop personalized improvement strategies.

2. Related Work

Han et al. [11] proposed a Multi-Level Damage Assessment (MLDA) framework for accurately assessing building damage in multi-hazard environments. Addressing the challenge of traditional methods struggling to differentiate levels of building damage across various hazards, the MLDA framework includes a Global Spatial Feature Guidance (GSFG) module and a Difference Change Feature Attention (DCFA) module. GSFG uses a non-local attention mechanism to enhance focus on distant spatial deformations, extracting common features across multiple hazards. DCFA improves the recognition accuracy of fine-grained building damage by integrating multi-scale feature fusion with dual attention mechanisms, emphasizing subtle differences between pre- and post-disaster images. Diaz et al. [7] introduced a simulation-based disaster management framework to analyze housing recovery in the Hampton Roads area, USA. This framework quantifies potential losses to residential areas through simulations and generates predictive scenarios to support decision-making during the reconstruction process. It comprises four main stages: disaster simulation, pre-disaster planning, immediate post-disaster response, and long-term recovery. The simulation results offer guidance for reconstruction, enabling local governments to prepare in advance and enhance disaster resilience. Ye et. al. [22] developed an audio data mining framework to identify human-induced disasters. The framework employs unsupervised learning and data-driven classification to automatically construct a hierarchical structure of disaster audio. This method involves three main stages: initially, a dictionary learning algorithm extracts robust acoustic features to effectively identify disaster events in noisy environments; then, audio event classification is generated based on probabilistic distances between categories; finally, this classification structure is embedded in a hierarchical classifier to enhance event identification performance.

Suf et al. [18] proposed an AI-based disaster monitoring framework that uses geolocation and sentiment analysis from social media posts to monitor disasters. By applying techniques such as Named Entity Recognition (NER), sentiment analysis, regression analysis, and anomaly detection, the framework can automatically extract disaster-related information from multilingual Twitter data worldwide. Dwarakanath et al. [8] reviewed machine learning methods for post-disaster emergency coordination using social media data. With the increasing role of social media in crisis communication and coordination, this study aims to analyze how various machine learning techniques can automatically extract useful information to aid rescue efforts. The study encompasses multi-level classifications, including early warning and event detection at the onset of disasters, post-disaster coordination and response, and damage assessment.

Saleem et al. [17] proposed a lightweight deep transfer model framework called DeLTran15, specifically designed for multi-class classification of disaster-related posts on the X (formerly Twitter) platform to support humanitarian actions in disaster management. The framework utilizes the OSEMN methodology (comprising data acquisition, cleaning, exploration, modeling, and interpretation) to enhance the comprehensiveness and reliability of the classification process. Chamola et al. [5] conducted a comprehensive review of the applications of machine learning technologies in disaster and pandemic management. The study points out that, with the development of IoT, drones, 5G, satellites, and other technologies, machine learning can effectively handle the vast and multi-dimensional data involved in disaster and pandemic management. This capability aids in disaster prediction, crowd evacuation route analysis, social media monitoring, and post-disaster management.

Zhao et al. [23] proposed a disaster chain evolution analysis and simulation method based on Fuzzy Petri Nets (FPN), using marine oil spill disasters as an example to demonstrate its application. The study indicates that a single disaster can trigger a series of secondary disasters, forming complex disaster chain effects that complicate emergency management. To describe and simulate the evolution of these disaster chains, the study developed an improved Fuzzy Petri Net model (DCFPN), which uses dynamic observational data to infer disaster chain evolution and identify the most risk-prone evolutionary paths. Talley [19] explored the application of digital technologies in disaster management, noting that the frequency and cost of natural disasters are continuously increasing, especially in disaster-prone regions like the United States. With rising urban and coastal population densities, disaster management challenges are becoming increasingly complex. The study emphasizes that disaster management is a ”big data problem” requiring coordinated solutions from public and private sectors, with digital technology playing a critical role. Li et al. [12] proposed a scenario-driven hybrid network model (SHN) for the dynamic simulation of disaster propagation in engineering systems. This method divides the disaster chain into a series of related scenarios and models the cascading effects among these scenarios through risk, propagation, and outcome processes. Bae et al. [3] proposed an agent-based disaster response system assessment framework that integrates geospatial and medical details to address mass casualty incidents (MCI). This framework simulates the entire emergency response process from the disaster site to hospitals, covering rescue, triage, transportation, and further analyzing the impact of medical resource distribution and road networks. A case study in the Gangnam area of Seoul, South Korea, demonstrated that the number of emergency physicians and operating rooms significantly affects disaster response efficiency.

Bala et al. [4] conducted an in-depth case study on the response actions of the Veterans Health Administration (VHA) during Hurricane Katrina and proposed five strategies to develop and utilize Information Technology (IT) to enhance healthcare disaster response capabilities. These strategies include: 1) establishing an integrated IT architecture to facilitate data interoperability, 2) developing a universal database for storing and accessing patient data, 3) creating a web-based disaster communication and coordination system, 4) developing an IT-supported disaster management system, and 5) standardizing and integrating IT disaster response processes. Ray et al. [15] provided a comprehensive review of the application of Internet of Things (IoT) technology in disaster management, exploring how IoT can enhance disaster response efficiency through early warning, data analysis, remote monitoring, and real-time analysis. The study introduces IoT-supported protocols and available market solutions, covering applications in natural disasters (such as earthquakes, floods, and wildfires) and human-induced disasters. It also emphasizes the role of IoT in post-disaster management, such as victim localization and support. Oscar Rodriguez-Espindola [16] proposed a multi-period dynamic model for managing simultaneous multi-regional disasters. This model combines bi-objective dynamic stochastic programming, covering supplier selection, facility location, resource allocation, inventory management, and distribution of disaster relief supplies. Chou et al. [6] proposed an ontology-based approach to develop a web design framework for natural disaster management, aiming to improve the functionality of web information systems in response to natural disasters. Using grounded theory, the researchers identified 2,094 web elements from 6,032 pages across 100 disaster management websites and organized them into an

ontology structure covering the five major phases of disaster management: general preparedness, specific disaster preparedness, disaster occurrence, post-disaster recovery, and learning. Agarwal et al. [1] proposed a Procedural Content Generation (PCG) framework using Reinforcement Learning (RL) to support disaster evacuation training in virtual 3D environments. This system provides a safe and cost-effective disaster response training alternative to traditional physical drills. The study designed a three-tier PCG architecture to create dynamic and realistic disaster scenarios and used an RL-PCG algorithm to develop a prototype for fire evacuation training. The references and discussions are summarized in Table 1.

In [2], a real-time semi-automated staff assignment system based on machine learning and text mining was proposed to enable efficient task allocation in multi-project management environments. The core objective of this system is to address the inefficiency and lack of flexibility in traditional task assignment methods that rely on historical data. By analyzing task descriptions in real time, the system quickly and accurately assigns tasks to the most suitable staff. The system employs text mining techniques, including tag removal, stop-word filtering, and stemming, as part of its preprocessing operations, and utilizes various vectorization methods such as TF-IDF, Word2Vec, and Doc2Vec to represent text data. Combining these methods with machine learning-based demand prediction, the system applies cosine similarity to precisely match task requirements with staff qualifications. Additionally, the system features dynamic updates to staff profiles, automatically adjusting their qualifications based on completed tasks, thereby ensuring continuous optimization. Tested in a real-world consulting company, the system achieved an 80% accuracy rate in task assignment, demonstrating comparable or superior performance to traditional systems reliant on historical data. In [20], the design and performance evaluation of MOS-Net and its enhanced version, MOS-Net 2.0, for moving object segmentation are thoroughly discussed. Moving object segmentation plays a crucial role in applications such as video surveillance, traffic monitoring, and autonomous navigation. However, traditional methods often rely on handcrafted feature designs, which are limited in handling complex scenarios. MOS-Net, based on a U-Net-like architecture, integrates the flux tensor algorithm and 3D Convolutional Neural Networks (3D CNNs) to capture both spatial and temporal features effectively. The flux tensor efficiently extracts motion information, while the 3D CNN processes temporal sequences to enhance segmentation accuracy. Building upon this foundation, MOS-Net 2.0 incorporates ConvLSTM layers to capture long-term temporal dependencies, making it more robust in dynamic backgrounds and complex scenes. Evaluations on the CDNet2014 dataset demonstrate that both models achieve superior performance on unseen data, particularly in F1 scores, surpassing traditional and contemporary methods such as BSUV-Net and FgSegNet. MOS-Net and its enhanced version exhibit exceptional capabilities in dynamic and unseen scenarios, offering effective solutions for applications requiring precise motion analysis and segmentation. Future work includes optimizing network architecture and integrating additional temporal features.

Integrating triage classification into digital games not only lowers the learning barrier for the general public but also turns the game into an educational resource accessible to everyone. Using a game format can create a more realistic disaster scenario than typical games, increasing player engagement and fostering a heightened level of preparedness among citizens when disasters occur.

Table 1. The references and discussions are summarized

Related Work	Content
Han et al. (2024)	Disaster Management Frameworks: For instance, Han et al. (2024) proposed a multi-level damage assessment framework utilizing spatial and attention mechanisms for building damage analysis. While effective for hazard recognition, it lacks a training component for responders. In contrast, our system bridges the gap by providing a hands-on, immersive training environment.
Agarwal et al. (2023)	Simulation-Based Disaster Training: Agarwal et al. (2023) introduced a procedural content generation framework for virtual fire evacuation drills. While their approach excels in scenario variety, it does not leverage XAI for transparency, which is a key feature of our system to ensure the interpretability of triage decisions.
Bae et al. (2018)	AI in Triage Systems: Bae et al. (2018) presented an agent-based disaster response system focusing on logistics and resource distribution. Unlike our system, theirs does not emphasize educational aspects or individual decision-making training.

3. Methodology

3.1. System Model

The purpose of this system is to use digital games to simulate disaster scenarios, thereby enhancing the triage skills of emergency personnel in critical situations. The system architecture comprises four main modules: the Scene Simulation Module, Triage Classification Module, Interactive Control Module, and Props System Module. Each module is responsible for different functions and exchanges data interactively to ensure the completeness and realism of the triage training. The system architecture diagram illustrates the interactions and data flow among the four modules, clearly depicting the roles of each module and how they collaborate to accomplish the triage simulation, as shown in Figure 1. This module leverages Generative Adversarial Networks (GANs) to produce diverse and realistic disaster scenarios, such as varying levels of structural damage and patient distribution. SHAP (SHapley Additive exPlanations) is employed to compute the impact weights of key features, ensuring the scenarios are both rational and diverse.

3.2. Generative AI Scenario System

In this system, the Scene Simulation Module uses Generative Adversarial Networks (GAN) to generate diverse disaster scenarios, including the extent of damage to compartments and the distribution of patients. To enhance the system's interpretability, an XAI method based on SHAP (SHapley Additive exPlanations) is employed to calculate the impact weight of each feature within each generated scenario, ensuring the rationality and diversity of scenario generation. The impact of these features helps in understanding how to allocate rescue resources in different scenarios. The formula for calculating the feature value S_i of the generated scenario is as follows, where S_i represents the weight of each feature in the scenario generation:

$$S_i = \sum_{j=1}^{n} w_j \cdot x_{i,j} \tag{1}$$

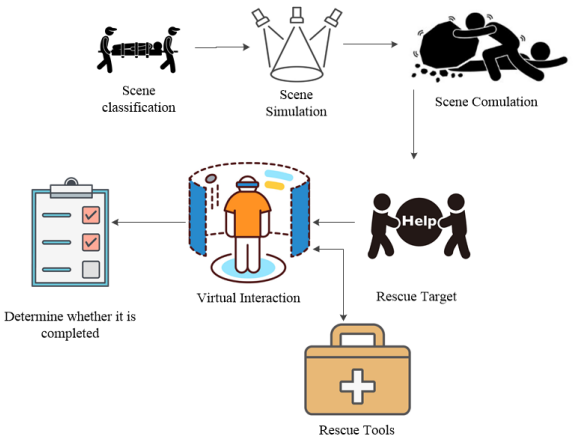


Fig. 1. System Architecture Diagram

Here, w_j represents the importance weight of feature j, $x_{i,j}$ represents the standardized value of feature, and n is the total number of features.

To help rescue personnel understand how the triage classification system categorizes patients based on symptoms, this system uses LIME (Local Interpretable Model-agnostic Explanations) to explain the triage classification process. Whenever the AI system automatically assigns a patient to a specific level (red, yellow, green, or black), LIME displays key symptom indicators, such as respiratory rate and heart rate, assisting rescue personnel in understanding the reasoning behind the classification decisions. In the system, the priority level P of a patient is a weighted result based on symptom indicators, where the weights of respiratory rate R, heart rate H, and capillary refill time T_c are w_1, w_2, and w_3, respectively:

$$P = w_1 \cdot R + w_2 \cdot H + w_3 \cdot T_c \tag{2}$$

Based on the explanation results from LIME, users can view the contribution of each symptom indicator to the final triage level, thus gaining a better understanding of the rationality and importance of the classification. XAI technology can also automate the generation of multi-scenario assessments, setting optimal parameters for different situations. Through this technology, the training system can dynamically adjust the rescue priority needed for each scenario. For example, if the damage level in a scenario is high, the system will automatically increase the emphasis on urgency in the triage classification, allowing high-priority patients to be classified earlier.

Assuming the damage index of the scenario is D, the adjustment formula for calculating patient priority is as follows:

$$P' = P \times (1 + \alpha \cdot D) \tag{3}$$

In this context, α represents the impact factor of scenario damage. When the D value is high, the system raises the priority classification of patients to meet the urgent demands of the disaster site. The feedback functionality of XAI allows the system to provide tailored recommendations for each trainee, showing the rationale behind each classification

decision, thereby enhancing the trainee's understanding of triage principles. Additionally, the system's self-learning capability optimizes parameters for the next round of scenario generation upon receiving new classification data, making the simulations more closely aligned with real disaster scenarios.

3.3. Virtual Reality Scenario System

The Scene Simulation Module focuses primarily on scene generation and scene scaling. This module uses the Unity 3D engine to construct an overturned compartment scenario, simulating damaged structures inside the compartment, scattered patients, and broken objects. Unity's terrain generation and physics engine operate within the disaster scenario to provide realistic visual effects and physical behaviors. The scene is scaled based on the size of the scenario and the distribution of patients to ensure the accurate representation of scene details. The calculation formula is as follows:

$$S_C = \frac{R_s}{S_s} \tag{4}$$

Here, S_C represents the scene scaling factor, R_s is the real-world scene length, and S_s is the real-world scene width. By controlling the scaling, the compartment's length and width are adjusted to achieve a realistic simulation effect. This scaling reduces the system resource load, enabling a smoother simulation process.

Patients are classified based on their condition (e.g., breathing, heartbeat, consciousness) into four categories: red, yellow, green, and black. Each color represents a different level of urgency. First, check the patient's breathing status to determine if they are breathing. If there is no breathing, the patient is classified as black, calculated as follows:

$$R = \frac{B_n}{M} \tag{5}$$

Here, R represents the respiration rate, B_n is the number of breaths, and M is the number of minutes. When $R = 0$, the patient is classified as black, indicating no signs of life. If the capillary refill time (TC) exceeds 2 seconds, the patient is assigned a red classification. Based on the patient's heart rate H, if $H < 60$ or $H > 120$, the patient is also classified as red. Using these formulas, the system quickly classifies patients and automatically assigns the corresponding color label.

The Interactive Control Module is designed for player movement and interaction control. Using the keyboard (W, A, S, D keys), players can move their character, and a first-person perspective is used to simulate the intensity of the scene. When players approach a patient, symptom information automatically pops up, allowing them to select the appropriate triage level. In the simulation scenario, users can interact with different patients and choose the correct triage level based on symptoms. The system provides various emergency items (such as first aid bandages, pain relievers, etc.), enabling players to select and use resources in the simulation. Players can pick up and use items to respond to specific patient conditions.

The system process is designed as follows:

1. Enter the simulation scenario
   Once the player enters the scene, the system presents a damaged compartment, scattered patients, and emergency supplies. The player can freely explore the scene and search for patients.

2. Triage Operation
   When the player approaches a patient, the system automatically displays the patient's basic symptoms (such as breathing status, consciousness, etc.). Based on the observed symptoms, the player determines the triage level (red, yellow, green, or black) and can make a selection using mouse controls.
3. Item Usage
   The player selects from available first aid items in the scene and applies them to the patient. Items are automatically matched to symptom needs, assisting the player in performing effective first aid.
4. Result Evaluation
   Once triage is completed for all patients, the system automatically generates a score and feedback, evaluating the accuracy and efficiency of the player's classifications and providing suggestions for improvement.

4. Experimental Results

This system was developed on the Unity platform using C# as the programming language, simulating a train derailment disaster scenario with the aim of training emergency responders in triage decision-making when faced with a large number of casualties. The experiment included 10 patients with varying symptoms, and 15 participants were invited to test the system to evaluate its effectiveness and the learning outcomes of the participants.

4.1. Experimental Design and Procedure

In the experiment, participants were asked to play the role of emergency responders arriving at the scene, required to quickly assess patients' injuries and classify them (red, yellow, green, or black level) to appropriately address varying degrees of injury. After completing the triage for each patient, that patient would disappear from the scene, indicating successful treatment. Patients who were not successfully classified would remain in the scene, and participants would need to re-evaluate them until a correct classification decision was made.

For all 15 participants, the system recorded their accuracy in judgment across different symptoms, average completion time, and response speed. The results showed that the system effectively improved participants' accuracy and speed in triage classification. Specific data are as follows:

1. Accuracy Analysis
   In the first attempt, the average triage classification accuracy for the 15 participants was 83%. Most participants were able to correctly distinguish between red and green level patients, but there was some deviation in judging yellow and black levels. After three repeated tests, the participants' average accuracy improved to 94%.
2. Completion Time
   In the initial test, participants took an average of 48 seconds to complete the classification for each patient. As they became more familiar with the triage process, the average completion time was reduced to 32 seconds, indicating that practice helped participants respond more quickly.

3. Retest Scores and Error Rate
   For patients initially misclassified, participants were able to correctly classify them in 90% of cases on the second attempt. The most common errors were in distinguishing between black and yellow levels; some participants found it challenging to decide when to classify a patient as black (no signs of life) or yellow (secondary priority).
4. Sense of Achievement and Feedback
   After classifying all patients, the system provided participants with an achievement score based on the number of successfully classified patients. Most participants reported that the achievement score increased their motivation to learn and helped them understand triage standards. They noted that repeated testing on unclassified patients allowed them to clearly understand the reasons for misjudgments and make targeted improvements in future decisions.

4.2. Effectiveness Evaluation

The experimental results of this system indicate that conducting triage training in a simulated environment helps improve responders' reaction speed and triage accuracy. Through repeated testing and the system's real-time feedback, participants were able to more effectively grasp classification principles, particularly in making rapid judgments for severely injured patients. Additionally, the achievement system and feedback mechanism motivate learners, allowing them to gain deeper learning experiences from mistakes.

Overall, this system has demonstrated good effectiveness in enhancing triage skills. In the future, the complexity of simulated scenarios could be expanded for field application to further improve training outcomes. Figures 2 to 5 illustrate the system interface during scenario development.

Fig. 2. Train Interior Scene

Fig. 3. Displaying Symptoms upon Patient Contact

Fig. 4. Patient

Fig. 5. Triage Classification Options

5. Conclusion

This paper leverages digital game simulation technology to create a realistic accident scene training environment, providing emergency responders with an immersive platform for triage training. Using Unity 3D, the system places players in the role of first responders at a disaster site, where they perform triage to enhance their response capabilities and decision-making efficiency in emergencies. Experimental results demonstrate that the system significantly improves the accuracy and speed of patient classification, while also enhancing learners' sense of achievement and engagement. By combining the interactivity of digital games with the realism of 3D simulations, this training model surpasses traditional approaches, offering both emergency personnel and the general public a safe and effective way to experience and learn the importance of triage.

Moreover, the digital game simulation model extends beyond professional responders to serve as a valuable disaster response training tool for the general public. With its user-friendly interface and real-time feedback, the system enables individuals to learn basic triage principles, understand injury classifications, and contribute to casualty reduction during real disasters. This study highlights the significant potential of digital game simulations to revolutionize disaster training, presenting an academically innovative approach that enhances practical emergency response skills. Future research could explore integrating virtual reality technology to further heighten the immersive experience of simulated scenarios. Additionally, expanding the system to address a broader range of disaster contexts would enhance its versatility, establishing it as a widely applicable training platform for achieving more effective rescue outcomes.

References

1. Agarwal, J., Shridevi, S.: Procedural content generation using reinforcement learning for disaster evacuation training in a virtual 3d environment. IEEE Access 11, 98607–98617 (2023)
2. Arslan, H., Işik, Y.E., Görmez, Y., Temiz, M.: Machine learning and text mining based real-time semi-autonomous staff assignment system. Computer Science and Information Systems 21(1), 75–94 (2024)
3. Bae, J.W., Shin, K., Lee, H.R., Lee, H.J., Lee, T., Kim, C.H., Cha, W.C., Kim, G.W., Moon, I.C.: Evaluation of disaster response system using agent-based model with geospatial and medical details. IEEE Transactions on Systems, Man, and Cybernetics: Systems 48(9), 1454–1469 (2018)
4. Bala, H., Venkatesh, V., Venkatraman, S., Bates, J.: If the worst happens: Five strategies for developing and leveraging information technology-enabled disaster response in healthcare. IEEE Journal of Biomedical and Health Informatics 20(6), 1545–1551 (2016)
5. Chamola, V., Hassija, V., Gupta, S., Goyal, A., Guizani, M., Sikdar, B.: Disaster and pandemic management using machine learning: A survey. IEEE Internet of Things Journal 8(21), 16047–16071 (2021)
6. Chou, C.H., Zahedi, F.M., Zhao, H.: Ontology for developing web sites for natural disaster management: Methodology and implementation. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 41(1), 50–62 (2011)
7. Diaz, R., Behr, J.G., Acero, B., Giles, B.D., Yusuf, J.E.W.: A simulation-based disaster management framework to analyze housing recovery: The case of hampton roads, usa. IEEE Transactions on Engineering Management 71, 4587–4601 (2024)

8. Dwarakanath, L., Kamsin, A., Rasheed, R.A., Anandhan, A., Shuib, L.: Automated machine learning approaches for emergency response and coordination via social media in the aftermath of a disaster: A review. IEEE Access 9, 68917–68931 (2021)
9. Haeberli, W., Whiteman, C.: Snow and Ice-Related Hazards, Risks, and Disasters, pp. 1–34 (12 2015)
10. Hallegatte, S.: The indirect cost of natural disasters and an economic definition of macroeconomic resilience. Policy Research Working Paper Series 7357, The World Bank (Jul 2015), https://ideas.repec.org/p/wbk/wbrwps/7357.html
11. Han, D., Yang, G., Xie, R., Lu, W., Huang, M., Liu, S.: A multilevel damage assessment framework for mixed-hazard buildings with global spatial feature guidance module and change feature attention in vhr remote sensing images. IEEE Geoscience and Remote Sensing Letters 21, 1–5 (2024)
12. Li, C., Ding, L., Fang, Q.: Dynamic simulation of the probable propagation of a disaster in an engineering system using a scenario-based hybrid network model. IEEE Transactions on Engineering Management 71, 1490–1503 (2024)
13. NOAA., N.: National centers for environmental information (ncei). u.s. billion-dollar weather and climate disasters. National Centers for Environmental Information (2021), [Online]. Available: https://www.ncdc.noaa.gov/billions/
14. Qiao, W., Shen, L., Wang, J., Yang, X., Li, Z.: A weakly supervised semantic segmentation approach for damaged building extraction from postearthquake high-resolution remote-sensing images. IEEE Geoscience and Remote Sensing Letters 20, 1–5 (2023)
15. Ray, P.P., Mukherjee, M., Shu, L.: Internet of things for disaster management: State-of-the-art and prospects. IEEE Access 5, 18818–18835 (2017)
16. Rodriguez-Espindola, O.: Multiperiod model for disaster management in simultaneous disasters. IEEE Transactions on Engineering Management 71, 4971–4984 (2024)
17. Saleem, S., Hasan, N., Khattar, A., Jain, P.R., Gupta, T.K., Mehrotra, M.: Deltran15: A deep lightweight transformer-based framework for multiclass classification of disaster posts on x. IEEE Access 12, 153676–153693 (2024)
18. Sufi, F.K., Khalil, I.: Automated disaster monitoring from social media posts using ai-based location intelligence and sentiment analysis. IEEE Transactions on Computational Social Systems 11(4), 4614–4624 (2024)
19. Talley, J.W.: Disaster management in the digital age. IBM Journal of Research and Development 64(1), 1–5 (2020)
20. Turker, A., Eksioglu, E.M.: 3d convolutional long short-term encoder-decoder network for moving object segmentation. Computer Science and Information Systems 21(1), 363–378 (2024)
21. Xinxin Zheng, Fei Wang, W.J.X.Z.Z.W.X.Q.Q.M., Chen, Q.: Construction and spatio-temporal derivation of hazardous chemical leakage disaster chain. International Journal of Image and Data Fusion 12(4), 335–348 (2021)
22. Ye, J., Kobayashi, T., Wang, X., Tsuda, H., Murakawa, M.: Audio data mining for anthropogenic disaster identification: An automatic taxonomy approach. IEEE Transactions on Emerging Topics in Computing 8(1), 126–136 (2020)
23. Zhao, Q., Wang, J.: Disaster chain scenarios evolutionary analysis and simulation based on fuzzy petri net: A case study on marine oil spill disaster. IEEE Access 7, 183010–183023 (2019)
24. Zhou, H., Wang, X., Yuan, Y.: Risk assessment of disaster chain: Experience from wenchuan earthquake-induced landslides in china. Journal of Mountain Science 12, 1169–1180 (09 2015)

Po-Hsuan Hsiao is a Ph.D. candidate in the Department of Computer Science and Information Engineering at National Taiwan University. His research interests focus on

computer vision, including image processing and recognition, as well as large language models (LLMs). Through his doctoral studies, he has developed expertise in applying deep learning techniques to solve complex visual computing challenges and exploring the capabilities of advanced language models.

Ming-Yen Chen is a Ph.D. in the Department of Computer Science and Information Engineering at National Cheng Kung University, while concurrently serving as a Department Manager at the Information and Communications Research Laboratories, Industrial Technology Research Institute (ITRI). His research interests span multiple domains of artificial intelligence, including audio signal processing and analysis, physiological measurement and analysis, information retrieval, semantic recognition, computer vision, and generative AI. Through his dual role in academia and industry, he focuses on bridging cutting-edge research with practical applications, contributing to technological innovation and advancement.

Hsien-Cheng Liao, Dr. is the Deputy Director of the Connected Mobility Technology Center at the Institute for Information Industry. He is mainly responsible for technology projects related to artificial intelligence, human-computer interaction, and their industrial applications. His research interests include human-computer interaction, artificial intelligence, and smart spaces.

Lo Ching-Cheng is a master's student in the Department of Computer Science and Information Engineering at National Taitung University, Taiwan. His expertise focuses on blockchain, AR, and machine learning. During his undergraduate studies, his research project explored the integration of AR with triage classification, and he has published a paper at Symposium on Digital Life Technolongies.

Received: November 3, 2024; Accepted: December 10, 2024.

Computer Science and Information Systems 22(3):1121–1138 https://doi.org/10.2298/CSIS241119036C

The Integration of Artificial Intelligence and Ethnic Music Cultural Inheritance under Deep Learning

Wenbo Chang

Performing Arts and Culture, The Catholic University of Korea, 43 Jibong-ro, Wonmi-gu, Bucheon-si, Gyeonggi-do 14662, South Korea
changwenbo2023@163.com

Abstract. The traditional music education system faces numerous challenges in inheriting ethnic music culture. Especially in the modern educational environment, the protection and dissemination of ethnic music encounter many difficulties. This work aims to utilize advanced technologies such as deep learning (DL) to explore methods for optimizing the inheritance of ethnic music culture. By summarizing the current situation of ethnic music cultural inheritance, and analyzing its background and content, this work proposes innovative solutions that integrate artificial intelligence (AI) technology. Leveraging a newly constructed model, it performs multi-level and comprehensive analyses of ethnic music elements, uncovering the internal emotional expression mechanism of ethnic music. The experimental results of timbre emotion recognition are presented and compared. The findings reveal that the unsupervised training method improves the feature accuracy by 1.96% compared to the Mel-Frequency Cepstral Coefficients (MFCC), while the supervised training method achieves a 3.46% improvement. In addition, the timbre recognition rate is compared between the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) and the Deep Neural Network-Hidden Markov Model (DNN-HMM). The result shows that the DNN-HMM is better. These findings highlight the significant advantages of applying DL methods in preserving and transmissing ethnic music cultural inheritance. This work can effectively enhance the accuracy of music emotion recognition, thus providing new technical support for the protection and inheritance of ethnic music.

Keywords: Deep Learning, Artificial Intelligence, Ethnic Music, Cultural Inheritance, Performance Evaluation, GMM-HMM, DNN-HMM.

1. Introduction

1.1. Research Background and Motivations

With the rapid progress of artificial intelligence (AI) technology, particularly significant breakthroughs in Deep Learning (DL), machines are now better equipped to understand and process complex, unstructured data, such as music audio signals. Thanks to their outstanding feature learning and pattern recognition abilities, DL models offer new

possibilities for the exploration, analysis, and preservation of ethnic music culture data [1]. Traditional music education systems face several challenges in transmitting ethnic music culture, including uneven distribution of educational resources, difficulties in implementing personalized teaching, and a shortage of qualified teaching staff [2]. Hence, optimizing music teaching methods using advanced technologies like DL, especially in the context of ethnic music inheritance, becomes crucial.

In the era of globalization, local and minority ethnic music cultures are at risk of marginalization or even loss. AI technologies, especially DL, offer the potential to digitize, classify, intelligently retrieve, and reproduce ethnic music resources, thereby providing robust support for protecting and disseminating ethnic music cultural inheritance [3-5]. The education sector universally seeks innovation and keeps pace with the times, aiming to apply advanced scientific technologies to educational practices to enhance teaching quality and efficiency. The fusion of DL and ethnic music cultural inheritance research reflects this trend, aiming to modernize ethnic music education processes through intelligent means [6].

The motivation behind this work stems from two main aspects. One is to recognize the powerful capabilities of DL in handling complex music information, uncovering underlying patterns, and implementing personalized teaching. Thus, new opportunities can be provided to optimize music education methods and enhance teaching effectiveness. The other is to deeply care about the issues of protecting and preserving ethnic music culture both in China and globally. Faced with the limitations of traditional educational approaches in promoting and popularizing ethnic music culture, such as unequal distribution of resources and monotonous teaching methods, AI technology, especially DL, is expected to offer innovative solutions. Therefore, this work explores how advanced AI technologies like DL can be applied to the education of ethnic music inheritance. By leveraging intelligent analysis, personalized recommendations, and interactive learning methods, this work seeks to enhance students' understanding and appreciation of ethnic music culture and promote the modernization of music aesthetic education. At the same time, it effectively fosters the inheritance and development of ethnic music cultures.

1.2. Research Objectives

First, this work explores how DL algorithms can be applied to accurately extract features from ethnic music and recognize patterns, including but not limited to melody, rhythm, harmony, and emotional expression across multiple dimensions. The goal is to achieve intelligent management and utilization of ethnic music cultural resources. Second, personalized music aesthetic education strategies are designed and implemented by integrating AI technology. These strategies can cater to students of different ages and backgrounds, providing learning content and pathways tailored to their characteristics. This approach aims to enhance the effectiveness of ethnic music education and foster greater student engagement. Third, by deeply integrating DL with ethnic music education, the work seeks to overcome geographical and temporal limitations, expanding the dissemination scope of ethnic music culture. This approach is intended to protect and inherit China's rich and diverse ethnic music inheritance,

contribute to the modernization of music education, and provide new technological support for international exchanges in ethnic music culture.

2. Literature Review

In the rapidly evolving information technology era, the development of AI and DL has brought unprecedented innovative opportunities for cultural inheritance [7,8]. Particularly within the realm of ethnic music cultural inheritance, the challenge of effectively combining cutting-edge technology with traditional culture to achieve mutual enhancement has become a widely discussed focus in the academic community [9]. This section provides an in-depth analysis of existing literature, highlighting the importance and urgency of integrating advanced technology with ethnic music cultural inheritance. It also delineates the achievements and challenges encountered by scholars in this field, offering a solid theoretical foundation and practical guidance for future research and practice.

2.1. Digital Preservation of Ethnic Music

In exploring the methods and technologies for the digital collection, organization, and storage of ethnic music resources, this work integrates audio signal processing technology with DL algorithms for feature extraction and classification. This provides a solid foundation for music education, academic research, and cultural dissemination. In addition, the technical and practical challenges faced during the digital preservation process, such as sound quality restoration, and long-term data storage stability, are discussed. Huang et al. were dedicated to digitizing, organizing, and storing traditional ethnic music resources, and constructing an ethnic music database. They employed audio signal processing techniques and DL algorithms for feature extraction and classification of ethnic music, laying the groundwork for subsequent music education, academic research, and cultural dissemination [10]. Lei et al. (2024) argued that digital technologies, particularly digital audio and video technologies, could accurately record and restore traditional music's sound characteristics and cultural background. Thus, the limitations of traditional recording and manual preservation methods could be overcome. Furthermore, digital preservation enabled the online sharing of ethnomusicological resources, breaking down geographical and cultural barriers, and allowing future generations to access and understand these traditional arts [11]. Zhang et al. (2023) pointed out that digital preservation was not only a technical issue but also involved respecting and protecting indigenous music's cultural context and intellectual property. They suggested that collaboration with indigenous communities should be considered during the digitization process, ensuring that the preservation methods aligned with local cultural and social practices, to avoid cultural appropriation or misuse [12]. Zhao (2024) highlighted that traditional ethnic music often featured complex melodies and rhythms that were difficult to accurately record using standardized audio formats. Therefore, customized recording equipment and analytical tools were necessary to capture the details of these notes and rhythms [13].

2.2. AI-Driven Personalized Ethnic Music Education

The design and implementation of an AI-based ethnic music teaching platform provide students with a customized learning experience. The system tailors content to students' learning habits, ability levels, and interest preferences, delivering relevant ethnic music materials. Simultaneously, it enhances students' understanding and appreciation of ethnic music through real-time feedback and interaction. Zhang and Romainoor developed such a platform, enabling intelligent recommendation, personalized teaching, and online learning of ethnic music. The system tailored suitable ethnic music content based on students' learning habits, proficiency levels, and interests, thus fostering a deeper understanding and appreciation of ethnic music through dynamic feedback and engagement [14]. Jing et al. (2024) emphasized that AI could personalize teaching content and facilitate the dynamic dissemination of ethnic music, attracting more young people to engage with and appreciate traditional culture [15]. Li et al. (2024) stressed that AI-driven emotional teaching could more comprehensively achieve the educational goals of ethnic music. This enabled students to master performance techniques and to deeply understand the cultural essence behind ethnic music [16]. Qin et al. (2022) highlighted that this personalized teaching approach could significantly enhance students' learning efficiency and engagement, while also aiding in the protection and dissemination of diverse ethnic music cultures [17].

2.3. DL and the Emotional Inheritance of Ethnic Music

This work explores using DL models to analyze and identify the emotional information embedded in ethnic music. It reveals the emotional expressions and aesthetic values in different ethnic music, enhancing students' perception and understanding of the deeper connotations of ethnic music through emotional education. In addition, the work discusses the importance of emotional inheritance in transmissing ethnic music culture and its role in fostering cultural identity. Reshma et al. utilized DL models to analyze and recognize emotional information in ethnic music, exploring the inherent emotional expressions and aesthetic values across different music genres. Their approach aimed to enhance students' perception and understanding of the profound connotations of ethnic music, thus facilitating emotional inheritance in ethnic music culture [18]. Catalina et al. (2023) highlighted the advantages of the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM), particularly in the speech recognition field. GMM could describe the probabilistic distribution of speech signals, while HMM was employed to capture the temporal characteristics of speech signals [19]. Hai et al. (2022) mentioned that HMM provided a natural framework for modeling state transitions in sequential data. In contrast, GMM helped improve the observation probability model of HMM, enabling the system to better fit continuous-valued observation data, particularly in the processing of audio signals or other time-series data [20].

2.4. Comparison between DL and Traditional Methods

The application of traditional ethnic music inheritance methods is compared with that of modern DL technology in the inheritance process. For endangered ethnic music heritage, Ning investigated the use of AI technology for recording and restoring such music, developing scientifically reasonable conservation strategies. How to pass on these music inheritances to the next generation through modern technological means was studied, ensuring the enduring vitality of ethnic music culture [21]. Zhou et al. (2024) argued that DL technologies could spread ethnic music globally by constructing large-scale ethnic music databases and recommendation systems. Through music generation techniques based on Variational Autoencoders (VAE), new variations of ethnic music could be created and pushed to a wider audience via online platforms [22]. Wang (2024) mentioned that modern DL technologies, such as emotion recognition and generation models, had the potential in conveying the emotional aspects of ethnic music. Emotion analysis tools based on DL could extract emotional features from audio data and generate emotionally consistent music segments through models. Although technological methods could not fully replace the emotional depth of live teaching, they could record and reproduce emotional details on a large scale, offering new possibilities for the preservation of ethnic music culture [23]. Yuan (2023) believed that DL technologies could record and analyze the audio features of ethnic music (such as melody, rhythm, and timbre) on a large scale, thus preserving traditional music as a digital resource. By using Generative Adversarial Networks and Convolutional Neural Networks (CNNs), music samples close to real performances could be synthesized, helping more learners access this music. Meanwhile, it could overcome the dependence of traditional inheritance on teachers and scenes [24].

In summary, the integration of AI with ethnic music cultural inheritance from the perspective of DL demonstrates broad application prospects and profound social value. Scholars have tackled numerous challenges in traditional music education by exploring the digital construction of ethnic music, developing intelligent music education systems, emotional recognition and education, heritage protection and inheritance strategies, and global dissemination. These efforts have paved the way for innovative approaches to preserving and promoting ethnic music culture. However, further in-depth research and continuous exploration are required to address emerging challenges in integrating cutting-edge technology and cultural heritage. These challenges may include enhancing DL models' understanding and expression of ethnic music's uniqueness and ensuring that technological means do not weaken cultural essence during the inheritance process. This work continues to evolve and optimize AI technology to propose more mature and comprehensive solutions, effectively advancing the preservation and inheritance of ethnic music culture to new heights.

3. Research Methodology

3.1. The Use of AI in Inheriting Ethnic Music Culture

Chinese excellent traditional culture boasts a long history and profound richness. It is the unshirkable responsibility of cultural workers in the new era to inherit and innovate this cultural heritage, ensuring its enduring vitality and preserving its essential role in society [25-27]. Applying AI to this endeavor can open up vast space for the inheritance and innovation of Chinese culture. The specific functions are reflected in four aspects. First, leveraging computer vision technology can enhance learning efficiency and the acquisition of traditional techniques, thereby contributing to the widespread inheritance of traditional culture. Second, integrating augmented reality (AR) technology with cultural scenic spots can promote cultural tourism and facilitate the digital development of traditional culture. Third, employing DL technology can foster secondary creation, enrich cultural forms, and endow the innovation of traditional culture with more personalized features. Fourth, big data technology can achieve precise promotion, gain insights into user needs, and ensure the accurate inheritance of traditional culture. Figure 1 illustrates these roles in inheriting and innovating traditional culture [28].

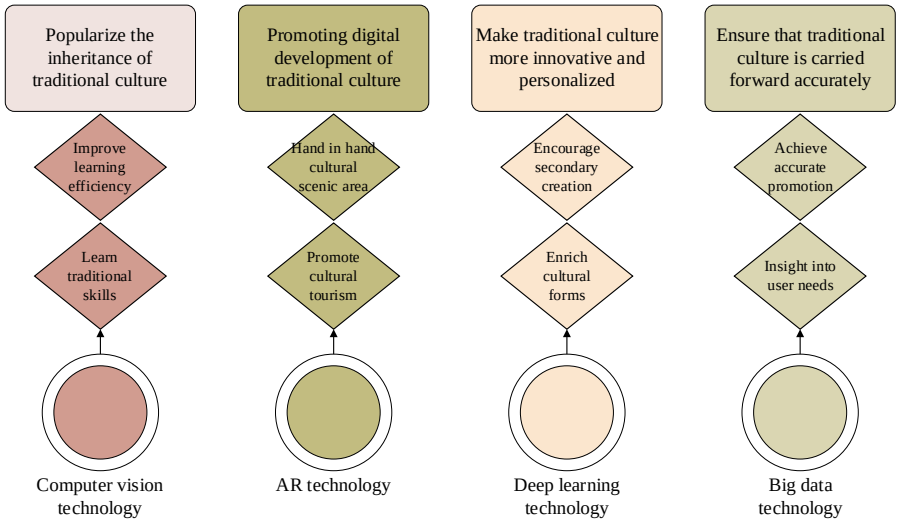


Fig.1. The Role of AI in Inheriting and Innovating Traditional Culture

Based on this, AI technology also plays a crucial role in inheriting ethnic music culture. It utilizes advanced techniques such as DL, big data analysis, and natural language processing (NLP) to digitize and intelligently analyze ethnic music [29]. On the one hand, AI facilitates the efficient collection, organization, classification, and storage of various ethnic music resources, establishing rich databases for easy retrieval

and dissemination. On the other hand, intelligent algorithms based on DL can accurately identify and extract features such as melody, rhythm, and mode of ethnic music. These can help people deeply understand and appreciate the unique music styles and emotional expressions of different ethnicities. Additionally, AI can be applied in intelligent music education to design personalized learning paths and provide engaging interactive teaching experiences that engage students more effectively. This fosters greater interest in and engagement with ethnic music among younger generations, ensuring the continued inheritance and promotion of the musical treasures of various ethnic cultures. Moreover, with the assistance of AI technology, ethnic music culture can transcend geographical restrictions, disseminating in a more modernized and internationalized manner to achieve global sharing and exchange [30-32]. Table 1 demonstrates its specific manifestations [33].

Table 1. Elements of AI Used in Inheriting Ethnic Music Culture

Number	Element	Implementation Means
1	Digitization Collection	Utilizing professional equipment and technical means to preserve ethnic music culture in digital form
2	Data Analysis and Processing	Employing big data analysis techniques to organize, annotate, and store collected audio data
3	Feature Extraction and Recognition	Using DL algorithms to extract features from music data and identify unique elements and styles of ethnic music
4	Intelligent Education Platform	Constructing an online music education platform, providing ethnic music-related courses and educational resources
5	Personalized Recommendation	Offering customized learning content based on users' learning records and preferences through intelligent recommendation algorithms
6	Dissemination	Promoting ethnic music culture globally through various channels such as the internet, social media, and mobile applications, facilitating cultural exchange and inheritance

3.2. The Use of DL in Acoustic Models

DL is a machine learning (ML) technology whose fundamental concept is rooted in research on artificial neural networks, especially the design of multi-layer nonlinear network structures. DL models consist of multiple interconnected layers, where each layer progressively extracts more abstract and complex feature representations from input data [34,35]. During the training process, DL algorithms adjust the network's weight parameters via backpropagation, enabling the model to automatically learn from raw input data and extract useful features. This process allows DL to perform complex tasks such as image recognition, speech recognition, NLP, and computer vision [36]. Compared to traditional ML methods, DL excels particularly in solving problems with

high-dimensional, unstructured data. With the support of large labeled datasets and substantial computational resources, DL can achieve performance beyond the human level, making revolutionary advancements in various AI fields. Figure 2 compares the structure of acoustic models based on the deep neural network (DNN) [37]. The right side of Figure 2 shows the structure of DNNs for an acoustic model. A DNN consists of an input layer, multiple hidden layers, and an output layer. The input layer receives acoustic features, such as Mel-frequency cepstral coefficients (MFCCs). Neurons in the hidden layers are connected by weights and introduce non-linearity through activation functions, enabling the network to learn complex patterns. The output of each hidden layer serves as the input for the subsequent layer, forming a feed-forward network structure. Model states (e.g., h1, h2, h3, h4, h5) represent the feature representations at different levels in the network. The depth of the DNN is determined by the number of hidden layers, which affects its learning ability and training difficulty. This structure is widely used in acoustic models for tasks such as speech recognition, speech synthesis, and other audio processing applications, as it is highly effective at capturing the intricate features and patterns of input data.

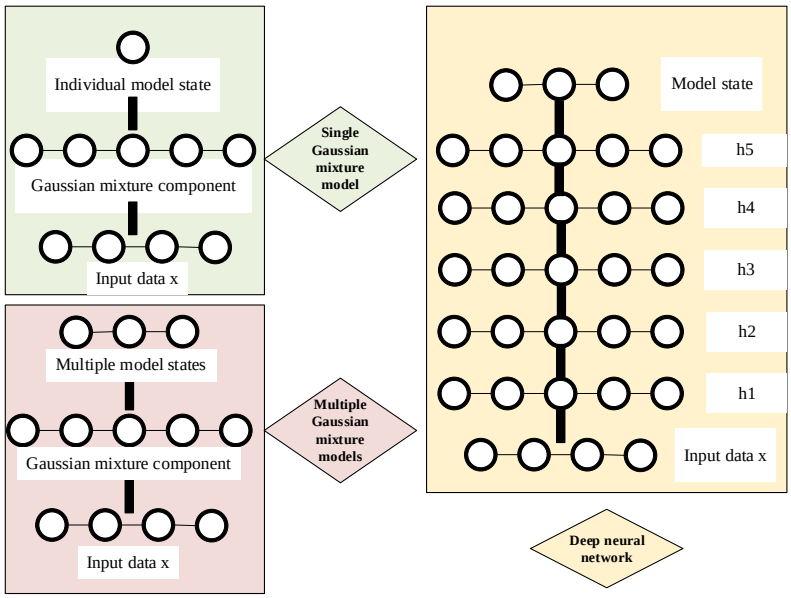


Fig. 2. Model Structure Comparison

In comparison, the structure of the GMM is similar to that of a DNN but features only a single hidden layer. Each node in this layer represents a Gaussian mixture component of the model. The output layer and nodes constitute the model's state vector, derived from the hidden layer nodes, thereby obtaining the posterior probability of the feature state vector from the input layer [38].

Next, the basic process of acoustic modeling is introduced. It is assumed that the feature observation vector at time t is y_{ut} and the activation probability of the output

layer state vector is P. Then, the state probability of the DNN-based acoustic model can be computed using the Softmax function as follows [39]:

$$y_{ut}(s) = P(s|O_{vt}) = \frac{exp\{a_{ut}(s)\}}{\sum_{s'} exp\{a_{ut}(s')\}} \quad (1)$$

$a_{ut}(s)$ represents the activation probability of the output layer node state s, that is, the effective output value, and its expression reads:

$$\log p(O_{vt}|s) = \log y_{ut}(s) - \log P(s) \quad (2)$$

The standard mean squared error backpropagation algorithm trains the DNN with the specified optimization objective function. Cross-entropy is chosen as the objective control function for the DNN system, while the optimization algorithm adopts the stochastic gradient descent algorithm [40-42]. Given that acoustic speech recognition is a multi-state classification application, the DNN selects the logarithmic function as the objective function, which can be written as:

$$F_{CE} = -\sum_{u=1}^{U}\sum_{t=1}^{T_u} \log y_{ut}(s_{ut}) \quad (3)$$

s_{ut} refers to the state of the system at the time t; F_{CE} represents the cross-entropy between the reference state vector and the predicted state vector. The gradient calculation equation between the system's objective function and the output layer's node state vector [43] is expressed as:

$$\frac{\partial F_{CE}}{\partial a_{ut}(s)} = -\frac{\partial \log y_{ut}(s_{ut})}{\partial a_{ut}(s)} = y_{ut}(s) - \partial s, s_{ut} \quad (4)$$

$\partial s, s_{ut}$ represents the Kronecker delta function.

3.3. Music Emotion Classification Based on Transfer Learning from a DL Perspective

Ethnic music culture embodies the essence of music art accumulated over the history of various ethnic groups, carrying profound ethnic emotions and cultural memories. Music emotion classification involves analyzing elements such as melody, rhythm, harmony, and timbre to identify and understand the emotional states and emotional connotations conveyed by the music. The inheritance of ethnic music culture plays a critical role in the research and application of music emotion classification. Each ethnic music genre has developed a diverse range of emotional categories, with its unique expressions of emotions and aesthetic concepts [44,45]. Employing modern technologies like DL to classify and interpret the emotional content of ethnic music can help people better understand the emotional characteristics of different ethnic music genres. Moreover, it can effectively preserve and inherit ethnic music culture, allowing listeners to experience and resonate with it at a deeper level, thereby promoting the continuation and development of ethnic music culture. Furthermore, accurate emotion classification can provide valuable insights for music education, assisting teachers and students in better grasping and imparting the intrinsic emotional appeal of ethnic music [46].

Using the trained CNN model for music emotion as a feature extractor, a new target-domain classifier is established to classify the music feature vectors. Figure 3 illustrates the basic process of model transfer [47]. It demonstrates the basic process of transfer learning. Here, Ds represents the source-domain classifier, which is used to pre-train the source-domain classifier. Then, the learned knowledge is applied to the target-domain classifier Dt through model transfer. Finally, feature extraction and fine-tuning are

performed on the target-domain data to improve the performance of the target-domain classifier.

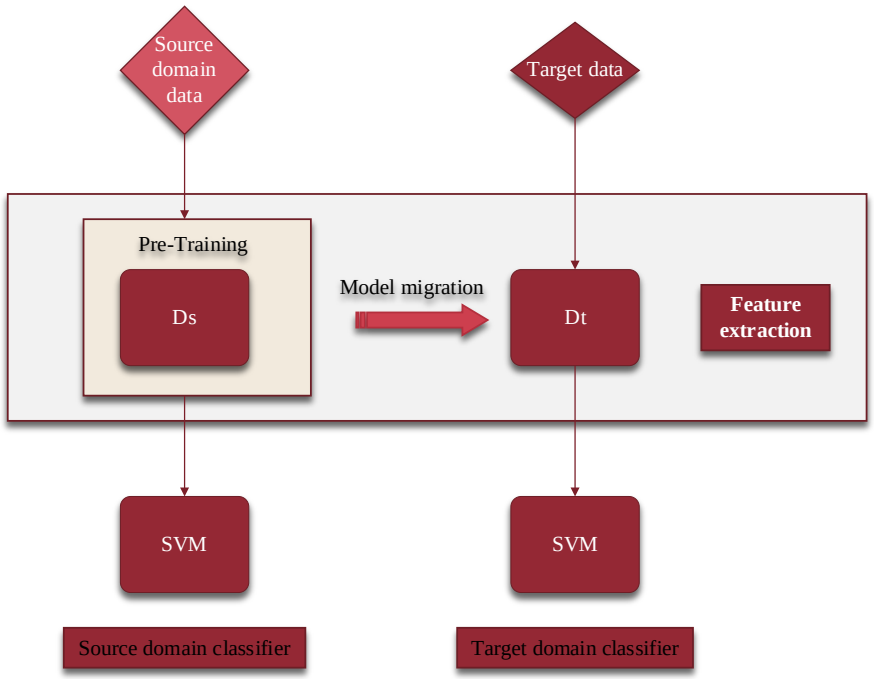


Fig. 3. Model Transfer Process

From a neural network perspective, similar spectral information, such as tones and rhythms, is mainly reflected in the structure, with this structural information concentrated in the layers preceding the convolutional layer. If the low-level information is directly transferred to the high-level, its expressive capacity may be diminished. Therefore, this project intends to merge the middle-level features with the high-level information, thus enriching low-level features with higher-level data. It aims to enhance the fusion ability of features and tasks, reduce the dependency of simple CNNs on high-level information, and improve the network's generalization ability [48,49].

Through music emotion classification, people can analyze the emotional characteristics such as joy, sadness, calmness, and excitement, embedded in different ethnic music works. These characteristics are often closely related to their unique cultural and historical backgrounds. By leveraging AI technologies like DL, emotional features can be extracted from various dimensions of ethnic music such as melody, rhythm, harmony, and timbre. The music can then be classified according to its emotional content [50]. This process helps to reveal and preserve the deep cultural and emotional meanings embedded in ethnic music. Moreover, it provides precise teaching resources and scientific teaching methods for music education. Enabling learners to

understand and perceive the emotions of different ethnic music is essential for them to inherit and promote ethnic music culture. Furthermore, it can cultivate students' aesthetic literacy and emotional engagement with ethnic music, thus facilitating the inheritance of ethnic music culture [51].

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

The dataset used for the music emotion classification experiments is based on the MIREX-like dataset, consisting of 709 songs. Each song is 30 seconds in duration and categorized into five fairly balanced groups. Table 2 presents the emotion labels for each category.

Table 2. Dataset Categories

Category	Number of music (pieces)	Emotion Labels
Category 1	127	Affectionate, Pleasant, Amusing, Joyful, Sweet
Category 2	127	Intense, Confident, Passionate, Lively, Noisy
Category 3	152	Restless, Humorous, Foolish, Eccentric, Witty, Distorted
Category 4	152	Sad, Thoughtful, Troubled, Graceful, Bitter, Reluctant
Category 5	151	Strong-willed, Intense, Tense, Anxious, Vulgar, Capricious

4.2. Experimental Environment

In this experiment, the open-source DL framework TensorFlow is used. Training for the support vector machine (SVM) is conducted using the SVM package from Scikit-learn. The MIREX-like music emotion dataset is utilized, and the Librosa software package is used for audio processing. The audio signals are sampled at 22050Hz, with a frame size of 1024 samples per second and a frameshift of 512 samples per second for Fourier transformation. This results in the extraction of the time-domain spectrogram of the audio signal. Logarithmic amplitude spectrograms are then obtained through logarithmic operations. Using a frequency of every 3 seconds, the processed spectrograms are fed into a mobile-based model as128×128 input data. Accuracy remains the metric for all experimental comparisons [52].

4.3. Parameters Setting

The established transfer model includes initializing model parameters. In the experiment, a very small random number (0.001* randn ()) is used to allocate

connectivity weights, and the bias of each node is set to 0. For each training set, the learning rate is set to 0.001, and termination terminates after 30 repetitions. After training, the final hidden layer (h1) and the state vectors of each node are retained as the input vectors for the risk-based monitoring (RBM) model. The termination condition for the experiment's learning process is either a maximum of 500 iterations or a change in the mean square error of 0.001 per iteration [53].

Table 3 presents the parameter settings for the experimental validation platform:

Table 3. Parameter Settings

Model Type	Parameters
DL Model	Model Architecture
	Hidden Layer (h1)
	Activation Function
	Weight Initialization Method
	Bias Initialization
	Learning Rate
	Number of Training Epochs
SVM	Kernel Function
	Penalty Coefficient (C)
	Kernel Function Parameters
	Maximum Number of Iterations
	Tolerance (tol)

4.4. Performance Evaluation

The performance of the three methods is first compared based on word precision, word accuracy, and sentence accuracy. Among them, word precision and word accuracy refer to the system's performance in word recognition. The word precision is the ratio of correctly recognized words to the total number of words. The word accuracy means the ratio of the number of correctly recognized words, excluding insertion errors. Sentence accuracy is similarly calculated by the ratio of correctly recognized words, but it excludes any additional inserted words from the total count. Figure 4 presents the experimental results.

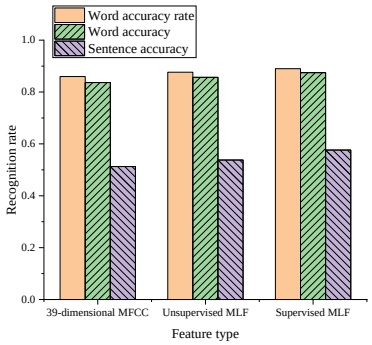


Fig. 4. Comparison of Recognition Rates

The experimental simulation results show that, compared with the traditional MFCC feature extraction method, the speech feature extraction method based on a deep auto-encoder mainly transcribes speech data and outputs the results or the model. It means the Master Label File (MLF), which can effectively improve speech recognition performance. Considering word recognition accuracy, the unsupervised training method's a feature accuracy reaches 1.96%, exceeding that of MFCC, while the supervised training method performs 3.46% higher than MFCC. However, for sentence accuracy, the performance of all three methods is relatively suboptimal, mainly due to the lack of a complete acoustic model.

The parameters of the DNN in the experiment are as follows. The input and output layer vectors are 143-dimensional, while the hidden layer consists of 1024×5 dimensions. The input layer adopts 11 frames of MFCC superframe features. Numerical experiments are conducted on the sentence error rate (SER), word error rate (WER), and model training time to evaluate the basic performance of the model. Table 4 displays the experimental results, comparing the performance of the GMM-HMM and the DNN-HMM in the task of timbre recognition.

Table 4. Comparison of Timbre Recognition Rates

Model	GMM-HMM	DNN-HMM
SER	30.4%	23.1%
WER	5.0%	3.6%
Training Time	12h	47h

The experiments indicate that AI-based DL algorithms outperform GMM algorithms, as GMM has a lower hierarchy and may not adequately simulate the brain's basic requirements for external environments. Considering the practical application context, there is a demand for intelligent speech recognition technology that enables the brain to process external sounds. With the assistance of DNNs, multi-level nonlinear mappings can abstract and simplify complex speech data, thereby obtaining features that meet practical requirements.

4.5. Discussion

Madzík et al. focused on leveraging DL technology to extract and analyze the tonal, rhythmic, and structural features of ethnic music. They innovatively applied this technology to the specific practice of ethnic music cultural inheritance. They investigated how DL models could accurately capture and interpret the emotions and cultural connotations embedded in ethnic music. Also, they designed and implemented a comprehensive DL-driven system for ethnic music inheritance aimed at music education. However, their research did not directly address the practical application of these findings in music education and cultural inheritance [54]. Bai et al. explored the transformative impact of AI on music education, particularly in personalized teaching and resource recommendation. However, they did not specifically examine ethnic music cultural inheritance, nor did they delve deeply into the role of DL in this context [55]. Unlike previous research that focused solely on technical feature extraction or was

limited to general music education applications, this work tightly integrates DL technology with ethnic music cultural inheritance. It explores specific strategies and methodologies to facilitate personalized teaching, emotional education, and cultural inheritance on intelligent music education platforms. Additionally, this work emphasizes the use of AI technology to promote the global popularity and recognition of ethnic music while simultaneously protecting and disseminating the diversity of ethnic music cultures. The goal is to foster a deep integration of technology and humanities in the realms of music education and cultural inheritance.

5. Conclusion

5.1. Research Contribution

This work, through the multi-level and comprehensive analysis of ethnic music elements, integrates DL into timbre and emotion recognition. This effectively uncovers the internal mechanism of emotional expression in ethnic music while deepening the understanding of its uniqueness and diversity. Empirical research has verified the effectiveness and feasibility of the proposed AI and DL model in the educational inheritance of ethnic music culture. The experimental results show that, compared with traditional teaching methods, DL-driven intelligent music education significantly improves students' cognitive understanding and appreciation of ethnic music. Consequently, it effectively promotes the inheritance and development of ethnic music culture.

The main contribution of this work lies in systematically introducing DL technologies into the field of ethnic music culture inheritance, demonstrating how DL can be utilized for precise feature recognition, emotion analysis, and personalized teaching. By establishing a multi-level and three-dimensional analytical model, this work deeply explores subtle elements in ethnic music such as timbre, rhythm, and emotional expression. Hence, it provides a practical and scalable framework for the transmission and education of ethnic music culture. This framework outperforms traditional feature extraction methods (e.g., MFCC) in terms of recognition accuracy, offering empirical evidence for the effectiveness of DL in recognizing and classifying the intrinsic emotional and cultural characteristics of ethnic music. Furthermore, the work finds that AI-based personalized teaching remarkably increases learners' engagement and appreciation of different musical traditions. Meanwhile, it overcomes limitations in resource allocation and teacher capabilities, thereby greatly promoting the popularization and education of ethnic music culture. The method proposed here also demonstrates significant improvements in emotion and timbre recognition accuracy compared to traditional methods and supports personalized teaching, enhancing learners' engagement and learning outcomes in ethnic music. Additionally, this method holds broad practical application potential, particularly in smart music education platforms, intangible cultural inheritance protection projects, and other related fields. Concurrently, it fosters a deeper integration of technology and the humanities, providing

innovative ideas for ethnic music education and cultural inheritance. Although the method shows good performance, its model generalization ability still needs validation due to limitations in dataset size and diversity. The high training costs may also restrict its practical deployment. In addition, the model's ability to capture the complex emotions of ethnic music is insufficient, and the lack of multimodal integration could affect its comprehensive understanding of music culture. These issues need further optimization and resolution in subsequent research.

Regarding practical applications, the method proposed in this work has significant potential, especially in smart music education platforms, online learning systems, and ethnic music culture preservation projects. The findings demonstrate the feasibility of AI technology in music education while promoting broader discussions about the digital protection and dissemination of intangible cultural inheritance. By integrating technologies such as virtual reality (VR) and AR with DL, future research could further enrich the learning experience of ethnic music culture. Moreover, its interactivity and immersiveness could be enhanced, thus opening up broader space for the inheritance and development of ethnic music.

5.2. Future Works and Research Limitations

With the ongoing advancements in AI and DL technologies, the integration research of AI and ethnic music cultural inheritance from a DL perspective holds significant promise. In the future, DL models can be further optimized and improved to enhance their ability to accurately understand and explore the complex and subtle emotional expressions and cultural connotations in ethnic music. Through the successful application of DL in timbre and emotion recognition, this work has laid the foundation for constructing an immersive, AI-driven ethnic music teaching platform. In addition, more application scenarios can be explored. For example, by combining VR and AR technologies with DL, rich and interactive learning experiences in ethnic music culture can be created. This can promote the digitalization, intellectualization, and personalization of ethnic music education.

Although this work has made valuable contributions to integrating DL and ethnic music cultural inheritance, certain limitations remain. First, the current DL models may have insufficient recognition accuracy when dealing with complex and highly localized features of ethnic music, necessitating optimization of model structure and training strategies. Second, limited by the dataset size and diversity, the model's generalization ability still needs to be verified through more comprehensive ethnic music data. Subsequent research should focus on expanding the dataset, particularly by collaborating with cultural institutions to collect under-represented music samples. Additionally, it should explore combining VR or AR technologies to provide a richer and more interactive learning experience.

References

1. Huang, L., Song, Y.: Intangible Cultural Heritage Management Using Machine Learning Model: A Case Study of Northwest Folk Song Huaer. Scientific Programming, 2(17). (2022)

2. Yao, M., Liu, J.: The Analysis of Chinese and Japanese Traditional Opera Tunes with Artificial Intelligence Technology Based on Deep Learning. IEEE Access, 11(1), 16-21. (2024)
3. Zhang, N.: Informatization Integration Strategy of Modern Vocal Music Teaching and Traditional Music Culture in Colleges and Universities in the Era of Artificial Intelligence. Applied Mathematics and Nonlinear Sciences, 5(11), 5-9. (2023)
4. Tang, Z.: Application Model Construction of Emotional Expression and Propagation Path of Deep Learning in National Vocal Music. International Journal of Advanced Computer Science & Applications, 14(11). (2023)
5. Zhou, W.: The Development System of Local Music Teaching Materials Based on Deep Learning. Optik, 273, 170421. (2023)
6. Li, P., Liang, T., Cao, Y., et al.: A Novel “An” Drum Music Generation Method Based on Bi-LSTM Deep Reinforcement Learning. Applied Intelligence, 54(1), 80-94. (2024)
7. Hui, F.: Transforming Educational Approaches by Integrating Ethnic Music and Ecosystems through RNN-Based Extraction. Soft Computing, 27(24), 19143-19158. (2023)
8. Başarır, L., Çiçek, S., Koç, M.: Local Intelligence: Time to Learn from AI. Architectural Science Review, 1-16. (2024)
9. Zhang, A.: Optimization Simulation of Match between Technical Actions and Music of National Dance Based on Deep Learning. Mobile Information Systems, 2023. (2023)
10. Huang, R., Holzapfel, A., Sturm, B., et al.: Beyond Diverse Datasets: Responsible MIR, Interdisciplinarity, and the Fractured Worlds of Music. Transactions of the International Society for Music Information Retrieval. (2023)
11. Lei, X.: Analysing the Effectiveness of Online Digital Audio Software and Offline Audio Studios in Fostering Chinese Folk Music Composition Skills in Music Education. Journal of Computer Assisted Learning, 40(5), 2339-2350. (2024)
12. Zhang, X.: Digital Communication of Folk Music in Social Music Culture. Frontiers in Art Research, 5(14), 322-356. (2023)
13. Zhao, X., Xu, D.: Research on Public Digital Cultural System of Huizhou Folk Music from the Perspective of Cultural Identity. Academic Journal of Humanities & Social Sciences, 4(10), 1028-1031. (2021)
14. Zhang, B., Romainoor, N. H.: Research on Artificial Intelligence in New Year Prints: The Application of the Generated Pop Art Style Images on Cultural and Creative Products. Applied Sciences, 13(2), 1082. (2023)
15. Jing, S., Lei, L.: Construction and Implementation of Content-Based National Music Retrieval Model under Deep Learning. International Journal of Information System Modeling and Design (IJISMD), 15(1), 1-17. (2024)
16. Li, P., Liang, M. T., Cao, M. Y., et al.: A Novel Xi'an Drum Music Generation Method Based on Bi-LSTM Deep Reinforcement Learning. Applied Intelligence, 54(1), 80-94. (2024)
17. Qin, L. N., Junyan, S.: Artificial Neural Network for Folk Music Style Classification. Mobile Information Systems, 2022(12), 1035-1038. (2022)
18. Reshma, M. R., Kannan, B., Raj, V. P. J., et al.: Cultural Heritage Preservation through Dance Digitization: A Review. Digital Applications in Archaeology and Cultural Heritage, 28, e00257. (2023)
19. Catalina, M., Marcelo, M., Aarón, C., et al.: An End-to-End DNN-HMM Based System with Duration Modeling for Robust Earthquake Detection. Computers and Geosciences, 179. (2023)
20. Hai, J. Y., Guofa, L., Jialong, H., et al.: Health Condition Evaluation Method for Motorized Spindle on the Basis of Optimised VMD and GMM-HMM. The International Journal of Advanced Manufacturing Technology, 124(11-12), 4465-4477. (2022)

21. Ning, H., Chen, Z.: Fusion of the Word2vec Word Embedding Model and Cluster Analysis for the Communication of Music Intangible Cultural Heritage. Scientific Reports, 13(1), 22717. (2023)
22. Zhou, Y., Huang, F.: Navigating Knowledge Dynamics: Algorithmic Music Recombination, Deep Learning, Blockchain, Economic Knowledge, and Copyright Challenges. Journal of the Knowledge Economy, 1-25. (2024)
23. Wang, J.: Music Personalization Imputation Method Based on Deep Transfer Learning. Applied Mathematics and Nonlinear Sciences, 9(1), 3215-3218. (2024)
24. Yuan, Y. Y., Samaneh, S.: Differentiated Analysis for Music Traffic in Software Defined Networks: A Method of Deep Learning. Computers and Electrical Engineering, 107. (2023)
25. Deng, J.: A Brief Analysis of the Path of Intangible Cultural Heritage Inheritance and Innovative Development under Digital Technology. Journal of Innovation and Development, 3(3), 29-32. (2023)
26. Xie, Y.: Consumer Preference Measurement of Folk Culture Based on Confidence Rule Base Model. Journal of Electrical Systems, 19(4), 211-226. (2023)
27. Deng, M., Liu, Y., Chen, L.: AI-Driven Innovation in Ethnic Clothing Design: An Intersection of Machine Learning and Cultural Heritage. Electronic Research Archive, 31(9), 5793-5814. (2023)
28. Wang, T., Chen, J., Liu, L., et al.: A Review: How Deep Learning Technology Impacts the Evaluation of Traditional Village Landscapes. Buildings, 13(2), 525. (2023)
29. Wang, T., Ma, Z., Yang, L.: Creativity and Sustainable Design of Wickerwork Handicraft Patterns Based on Artificial Intelligence. Sustainability, 15(2), 1574. (2023)
30. He, F.: Research on the Application of Artificial Intelligence Aesthetics in the Cultivation of Aesthetic Literacy of Art-Normal Students. International Journal of Social Sciences and Public Administration, 2(1), 160-168. (2024)
31. Foka, A., Eklund, L., Løvlie, A. S., et al.: Critically Assessing AI/ML for Cultural Heritage: Potentials and Challenges. Handbook of Critical Studies of Artificial Intelligence, 815-825. (2023)
32. Breit, A., Waltersdorfer, L., Ekaputra, F. J., et al.: Combining Machine Learning and Semantic Web: A Systematic Mapping Study. ACM Computing Surveys, 55(14s), 1-41. (2023)
33. Li, P., Wang, B.: Artificial Intelligence in Music Education. International Journal of Human–Computer Interaction, 1-10. (2023)
34. García-Madurga, M. Á., Grilló-Méndez, A. J.: Artificial Intelligence in the Tourism Industry: An Overview of Reviews. Administrative Sciences, 13(8), 172. (2023)
35. Rane, N., Choudhary, S., Rane, J.: Sustainable Tourism Development Using Leading-Edge Artificial Intelligence (AI), Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) Technologies. Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) Technologies (October 31, 2023), 2023.
36. Rane, N.: Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Arts and Humanities. Available at SSRN 4603208, 2023.
37. Ning, H.: Analysis of the Value of Folk Music Intangible Cultural Heritage on the Regulation of Mental Health. Frontiers in Psychiatry, 14, 1067753. (2023)
38. Zhou, E., Shen, Q., Hou, Y.: Integrating Artificial Intelligence into the Modernization of Traditional Chinese Medicine Industry: A Review. Frontiers in Pharmacology, 15, 1181183. (2024)
39. Sarkar, C., Das, B., Rawat, V. S., et al.: Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. International Journal of Molecular Sciences, 24(3), 2026. (2023)
40. Ahmed, N. M. F., Elwakeel, L. M., Marzuk, A. A., et al.: Using Artificial Intelligence Applications to Create Contemporary Heritage Designs to Enrich Typography to Preserve Saudi Heritage. Kurdish Studies, 11(3), 39-57. (2023)

41. Zheng, X.: Integration of Multiple Features in Chinese Landscape Painting and Architectural Environment Using Deep Learning Model. International Journal of Intelligent Systems and Applications in Engineering, 12(6s), 593-606. (2024)
42. Chen, Y., Wang, L., Liu, X., et al.: Artificial Intelligence-Empowered Art Education: A Cycle-Consistency Network-Based Model for Creating the Fusion Works of Tibetan Painting Styles. Sustainability, 15(8), 6692. (2023)
43. AYAZ, A., IMAMOGLU, M. B., BOZTAS, G. D., et al.: Artificial Intelligence and Immersive Reality: A Systematic Literature Review on Research in the Social Sciences. Current Studies in Technology, Innovation and Entrepreneurship, 50. (2023)
44. García-Esparza, J. A., Pardo, J., Altaba, P., et al.: Validity of Machine Learning in Assessing Large Texts through Sustainability Indicators. Social Indicators Research, 166(2), 323-337. (2023)
45. Zhu, M., Wang, G., Li, C., et al.: Artificial Intelligence Classification Model for Modern Chinese Poetry in Education. Sustainability, 15(6), 5265. (2023)
46. Xu, W.: Research on the Communication Opportunities of Intangible Cultural Heritage under the Background of Big Data and AI. Journal of Artificial Intelligence Practice, 6(8), 12-17. (2023)
47. Dong, L.: Using Deep Learning and Genetic Algorithms for Melody Generation and Optimization in Music. Soft Computing, 27(22), 17419-17433. (2023)
48. Liu, J.: An Automatic Classification Method for Multiple Music Genres by Integrating Emotions and Intelligent Algorithms. Applied Artificial Intelligence, 37(1), 2211458. (2023)
49. Spennemann, D. H. R.: Will Artificial Intelligence Affect How Cultural Heritage Will Be Managed in the Future? Responses Generated by Four GenAI Models. Heritage, 7(3), 1453-1471. (2024)
50. Tan, S. N., Ng, K. H.: Gamified Mobile Sensing Storytelling Application for Enhancing Remote Cultural Experience and Engagement. International Journal of Human–Computer Interaction, 40(6), 1383-1396. (2024)
51. Adjeisah, M., Asamoah, K. O., Yeboah, M. A., et al.: Adinkra Symbol Recognition Using Classical Machine Learning and Deep Learning. arXiv preprint arXiv:2311.15728, 2023.
52. Zou, H., Ge, J., Liu, R., et al.: Feature Recognition of Regional Architecture Forms Based on Machine Learning: A Case Study of Architecture Heritage in Hubei Province, China. Sustainability, 15(4), 3504. (2023)
53. Singh, A., Kanaujia, A., Singh, V. K., et al.: Artificial Intelligence for Sustainable Development Goals: Bibliometric Patterns and Concept Evolution Trajectories. Sustainable Development, 32(1), 724-754. (2024)
54. Madzík, P., Falát, L., Copuš, L., et al.: Digital Transformation in Tourism: Bibliometric Literature Review Based on Machine Learning Approach. European Journal of Innovation Management, 26(7), 177-205. (2023)
55. Bai, N., Ducci, M., Mirzikashvili, R., et al.: Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, a Case Study in Testaccio, Rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 139-146. (2023)

Wenbo Chang was born in Tai Yuan,Shan xi.P.R. China, in 1996. He received his master's degree from Belarusian State Academy of Music,Belarus. Now, he studies in The Catholic University of Korea. His research interest include culture and art sociology, ethnomusicology. changwenbo2023@163.com

Received: November 19, 2024; Accepted: Januar 15, 2025.

Computer Science and Information Systems 22(3):1139–1165 https://doi.org/10.2298/CSIS241121042L

The Analysis of Deep Learning-based Football Training under Intelligent Optimization Technology

Kun Luan[1], Fan Wu[1], and Yuanyuan Xu[2*]

[1]Nanchang University College of Science and Technology, Gongqingcheng, 330029, China;
155341805@qq.com
43969591@qq.com
[2]Jiangxi University of Finance and Economics, Nanchang, 330029, China
1201400004@jxufe.edu.cn

Abstract. This work aims to optimize college football training using deep learning techniques, addressing the inefficiencies, difficulty in action recognition, and insufficient data analysis present in current training methods. An intelligent optimization system combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) is proposed to tackle these challenges. Compared to traditional single models, the Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) architecture remarkably improves the efficiency and accuracy of processing training data by leveraging the strengths of spatial features and temporal sequence features. The experimental results show that CNN-RNN model is significantly superior to the traditional 3D CNN model and other advanced models, such as Transformer, Long Short-Term Memory (LSTM), Bidirectional LSTM and Gated Recurrent Unit (GRU), in key indicators such as accuracy, precision, recall and F1 score. Specifically, CNN-RNN model achieves 92.5% accuracy, 91.2% precision, 93.1% recall and 92.1% F1 score. The lowest training loss rate is 0.24, which is significantly better than other models. In addition, the introduced data balance strategy effectively improves the prediction performance of a few categories (such as foul and yellow card events) through oversampling, undersampling and weighted loss function, and further enhances the generalization ability and practicability of the model. Future research focuses on expanding the dataset, further improving the model's generalization ability, and exploring its application in real training scenarios.

Keywords: deep learning; college football training; intelligent optimization; CNN-RNN; training loss value.

1. Introduction

With the continuous advancement of technology, artificial intelligence (AI) has been widely applied across various fields, especially demonstrating significant potential in the sports realm [1, 2]. As one of the world's most popular sports, football has been a focal research point in optimizing training and game strategies. However, traditional football training methods often rely on the coach's experience and intuitive judgment, lacking scientific and systematic approaches. Therefore, this work explores how deep learning

(DL) technology can be applied to college football training to achieve intelligent and optimized training methods [3-5].

First, an important aspect of the research background is the challenges college football training faces [6-8]. Currently, college football teams exhibit deficiencies in technical skills, tactics, and physical fitness, and these issues become particularly pronounced when facing higher-level opponents. Additionally, limited training resources such as facilities, equipment, and funding constrain the athletes' development. These issues urgently require solutions through technological means [9, 10].

Moreover, as a crucial branch of AI, DL has achieved remarkable success in areas like image and speech recognition [11, 12]. In the football field, DL can be employed to analyze match videos, identify and assess the quality of player movements, and predict match outcomes [13-15]. These applications provide coaches with scientific data support and help athletes better understand their strengths and weaknesses, enabling more targeted training [16-18].

With the continuous development of modern football, the role of data analysis in match strategy and player performance evaluation has become increasingly important. Traditional football analysis methods largely rely on static statistical data, which struggle to effectively address the complexity of real-time changes during a match. Existing football analysis models often fail to fully leverage video data and player statistics when predicting real-time match events (e.g., goals, fouls, yellow cards), leading to low prediction accuracy and insufficient real-time feedback capabilities. Therefore, how to utilize DL technologies for precise prediction of real-time match events has become a critical research topic in optimizing football training and tactical decision-making.

This work aims to address the shortcomings of current football analysis models in handling video data and player statistics by introducing a DL model that combines Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Specifically, the goal is to use the Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) model to predict real-time match events and enhance prediction accuracy through data augmentation and model optimization. Compared to traditional analysis methods based on manual feature extraction, the CNN-RNN model possesses stronger feature learning capabilities, enabling it to capture dynamic information in videos and the temporal sequence features of player behavior. Thus, this work fills the gap in existing football analysis models regarding real-time event prediction while providing new technical support for personalized optimization of football training.

By accurately predicting real-time match events, this work can better assist coaches in tactical decision-making and improve the effectiveness of player training. Precise event prediction helps identify key moments in a match in advance, providing timely data support for tactical adjustments and player training. This not only enhances the viewing experience and competitive level of matches but also offers new research avenues for data-driven sports training models.

The core research objective is to develop and validate a DL-based intelligent optimization system, thus enhancing the scientific and systematic aspects of college football training. It is acknowledged that, despite the global popularity of football, challenges in technology, tactics, and physical fitness persist in college football training, particularly in situations with limited resources. A hybrid model, combining CNNs and RNNs, is proposed to address these challenges. The model aims to automatically

identify and assess the quality of players' movements by analyzing a substantial amount of football training and match video data. This analytical process provides coaches with data support, enabling them to formulate more effective training plans. Methodologically, a large-scale football dataset is first constructed, including training and match videos of players at different levels and relevant motion and performance data. Preprocessing steps, such as labeling player movements and events, are employed to ensure the accuracy of model training. Subsequently, a CNN-RNN architecture is designed, with CNN handling image and video data to capture visual features of player movements, and RNN processing sequence data to analyze the temporal information of actions. This combined approach allows the model to gain a more comprehensive understanding of the complexity of football movements. In terms of experimental design, a high-performance computing platform, along with TensorFlow and PyTorch frameworks, is utilized for model training and testing. Reasonable parameters, such as batch size and learning rate, are set to balance the model's learning efficiency and memory usage. Performance evaluation focuses on accuracy, precision, recall, and F1 score indicators to ensure the model's effectiveness in practical applications. The model's generalization ability is also considered by conducting cross-validation on different datasets to test its robustness.

The structure of this work is as follows. Section 1 presents the research background, motivation, and objectives. Section 2 introduces related research and literature review, analyzing the limitations of existing football analysis models. Section 3 details the design and implementation of the CNN-RNN model. Section 4 discusses data collection and experimental setup. Section 5 exhibits the experimental results and provides a detailed analysis. Finally, Section 6 concludes the study and discusses future research directions.

2. Literature Review

In the current field of sports technology, the application of DL technology is rapidly advancing and demonstrating significant potential and innovation in the football domain [19, 20]. With the continuous progress of big data and intelligent algorithms, there have been notable improvements in football match analysis, athlete performance evaluation, and the enhancement of the overall game experience. Many researchers and technology experts conducted in-depth exploration and practical applications in this field in recent years [21, 22]. Their work not only propelled the advancement of football technology but also introduced new perspectives and possibilities to the entire sports technology sector. Rahman (2020) proposed a DL-based football match prediction framework. This framework utilized complex algorithms and data analysis techniques to forecast match outcomes, considering factors such as team performance, historical data, and other relevant elements [23]. Stoeve et al. (2021) focused on applying laboratory technologies to real football scenarios. They used inertial measurement units and DL technology to detect shooting and passing actions in football training and matches, aiming to enhance the effectiveness and efficiency of athlete training [24]. Fenil et al. (2019) introduced a real-time violence detection framework for football stadiums. By combining big data analysis and Bidirectional Long Short-Term Memory (Bi-LSTM) in DL technology, this

framework aimed to improve the efficiency and accuracy of safety monitoring during football matches [25]. Cuperman et al. (2022) developed a DL-based process for football activity recognition using wearable accelerometer sensors. This approach contributed to a more accurate analysis of player movements and performance, providing valuable data for coaches and sports scientists [26]. Wang et al. (2019) dedicated their efforts to developing a DL-based intelligent editing system for football matches. This system could use algorithms to automatically edit and highlight crucial moments in matches, enhancing the viewing experience and media production efficiency [27].

The application of DL in the sports domain primarily focuses on two directions: athlete motion analysis and match prediction. By combining CNNs with RNNs, researchers attempted to improve athlete performance during matches and optimize training strategies. For example, Sen et al. (2021) employed a CNN-RNN architecture to analyze cricket match videos, enabling precise identification and classification of player actions [28]. By introducing CNN for extracting spatial features and combining it with RNN to capture temporal dynamics, the model could accurately predict player actions (e.g., passing, shooting, etc.). Moreover, it could further apply these predictions to optimize actions in training scenarios. However, this study's limitation lies in its relatively small dataset and its primary focus on action recognition within match scenarios, with less exploration of real-time feedback optimization during training. Li (2023) proposed a method combining CNNs with LSTM to predict the movement trajectories of football players [29]. This approach mainly concentrated on the dynamic prediction of player positions, but its research on training optimization was relatively weak. In contrast, this work focuses on action recognition during training and strives to optimize the training process through DL, particularly making innovative attempts in data augmentation and balancing.

The application of DL technology in the football field is becoming increasingly profound and widespread. From predicting match outcomes and analyzing player performance to monitoring the safety of playing fields, and even editing match videos and tracking the ball, DL technology is gradually transforming various aspects of football [30-32]. These studies not only showcase DL technology's immense potential in improving training efficiency, enhancing the spectacle of matches, and ensuring match venue safety but also pave the way for future technological innovations in football. By applying these advanced technologies, football is evolving towards a more precise, intelligent, and scientific direction, driven by technological innovation [33, 34]. In contrast, although some studies have tried to optimize training through deep learning technology, most of them focus on the application of a single model, such as the independent use of CNN or LSTM, and lack of multi-level analysis combining different neural network models. The advantage of deep learning technology is that it can effectively process complex training data through different types of network models, combining spatial characteristics and time series characteristics. In the existing research, although some scholars try to use the combination of CNN and RNN to improve the accuracy and efficiency of training, on the whole, no research has made full use of the potential of CNN-RNN architecture to solve the problems of unbalanced data, high complexity of actions and identification of a few types of actions (such as fouls and yellow cards) in football training. This kind of problem is the key challenge in current football training, and it is of great significance to improve the training effect. In

addition, the application of data enhancement and weighted loss function in existing research is still limited, which fails to fully tap its potential in solving data imbalance and improving model generalization ability. Especially when it comes to complex action recognition (such as a few kinds of actions like foul or yellow card), traditional deep learning models often encounter problems of insufficient accuracy and recall. By introducing data enhancement and weighted loss function, this work can effectively improve the precision and recall rate of the model in minority action recognition, and further improve the training effect.

In a word, although the existing research has made some progress in the application of deep learning technology in the field of football, there are still obvious technical gaps in the aspects of action optimization, multi-level model analysis and data imbalance in the training process. These shortcomings provide technical incentives for the proposal of this work. The innovation of this work lies in the combination of CNN and RNN, aiming at the optimization of action recognition in the training process, and improving the accuracy and efficiency of training through multi-level data balance strategy and model optimization technology, especially in the recognition of complex actions and the handling of minority problems, showing stronger advantages. Through these innovations, this work not only fills the gaps in the existing research, but also provides new ideas and technical support for the application of deep learning in future football training.

3. Research Methodology

This work aims to explore and implement an intelligent optimization technology for college football training using the DL-based CNN-RNN model. In the stage of model construction, a mixed model architecture combining CNN and RNN, namely CNN-RNN model, is adopted, which is realized by two deep learning frameworks, TensorFlow and PyTorch. In order to optimize the performance of the model, this work systematically optimizes the learning rate, batch size and optimizer by grid search method to ensure that the model can be trained under the optimal parameter configuration. In the process of model training, an independent verification set is used to monitor the performance of the model to prevent over-fitting. The selection of the CNN-RNN model is based on several key reasons. First, CNNs perform exceptionally well when handling image data, efficiently extracting spatial features from the video frames of football matches. Specifically, CNNs can automatically identify key areas in images, such as the player's body posture, movement trajectories, the ball's position, and other critical elements on the field. Therefore, CNNs can extract the spatial information from each frame of the football match, providing strong support for subsequent time-series modeling. Second, RNNs, particularly LSTM networks or Gated Recurrent Units (GRUs), excel at handling sequential data and capturing temporal dependencies within the data. A football match is a highly dynamic process, and key events in the game often rely on temporal relationships (for example, an event occurring at one point directly affects the subsequent course of the match). The time-series modeling capability of RNNs enables effective tracking of these temporal features, thereby identifying important events in the match (such as goals, fouls, etc.). By combining CNNs and RNNs, both spatial and

temporal features can be leveraged, significantly improving the accuracy of event prediction in football matches. This work uses CNNs to extract spatial features from each video frame. In contrast, RNNs process the temporal information within the video frame sequence, enabling the model to capture the dynamic spatiotemporal changes in the match. This synergistic approach allows the CNN-RNN combination to predict vital events in football matches more effectively than using CNNs or RNNs alone. Figure 1 illustrates the overall framework of the CNN-RNN model. This framework visually represents the process where video data flows through the CNN for spatial feature extraction and is subsequently passed into the RNN for time-series modeling.

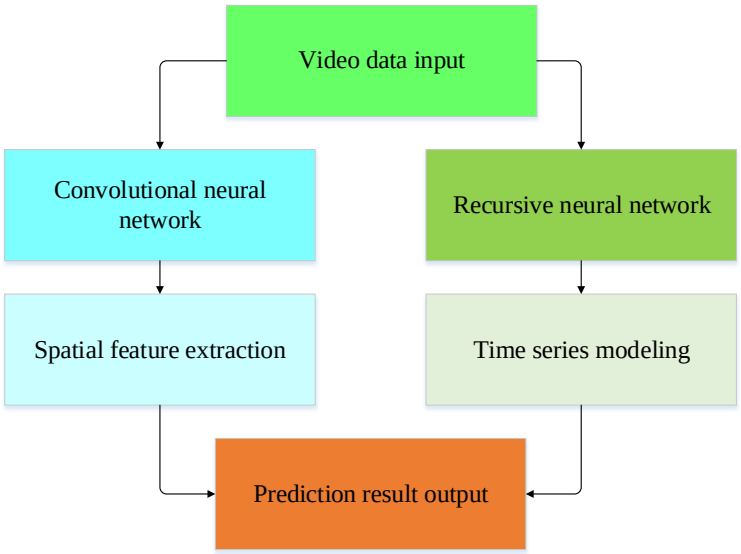


Fig. 1. Workflow of the CNN-RNN model

Figure 1 shows that, first, video data is processed through the CNN to extract spatial features (such as player positions, actions, and event labels). Second, the extracted spatial features are fed into the RNN for time-series analysis, ultimately outputting predictions for match events. The core of this work lies in utilizing advanced DL algorithms to precisely analyze and enhance the effectiveness of football training. To achieve this goal, the following research methods are employed:

First, a detailed data preparation process is conducted. This involves collecting training videos of college football players, match recordings, and relevant motion data. These data are used to train and test the DL model, ensuring that the model covers various aspects of football movements.

Next is the design and implementation of the DL model. This work leverages the strengths of CNNs and RNNs to create a specialized DL model for optimizing football training. On the one hand, CNN excels in handling image and video data, making it suitable for identifying and processing visual elements in football movements. On the other hand, RNN is proficient in dealing with sequential data and capturing dynamic information in time series. The combination of these two networks is expected to

analyze the complex dynamics of football training more accurately. Firstly, CNN is used to extract features from each video frame. This typically involves dividing the video into multiple frames and feeding them into the CNN to extract a feature representation for each frame. Secondly, the sequence of features extracted by the CNN is fed into an RNN. The RNN processes these features sequentially and captures the temporal dependencies between them. This allows the model to understand dynamic changes in the video, including player movements and fluctuations in the pace of games. Finally, predictions are made based on the RNN's output. These predictions might include classification (such as determining the match outcome), regression (such as predicting the timing of goals), or generative tasks (such as generating video descriptions). If temporal information is mentioned in image processing, it is likely because a series of images (video frames) are processed, rather than a single static image. In this context, temporal information refers to the sequence of images and their changes over time.

A combination of CNN and RNN architectures is employed to address the research problem in this work. Specifically, in the CNN section, three convolutional layers and two pooling layers are designed. The first convolutional layer uses a 3x3 kernel with a stride of 1, and the depth of the input feature map is 3 (corresponding to color images). This layer aims to extract low-level features such as edges and textures. The Rectified Linear Unit (ReLU) activation function is employed in this layer, as it effectively mitigates the vanishing gradient problem and accelerates network training. The second convolutional layer also utilizes a 3x3 kernel with a stride of 1, and the output depth is 64, primarily responsible for extracting higher-level features. The third convolutional layer retains a 3x3 kernel and a stride of 1, with an output depth of 128, further extracting complex spatial features. To reduce the dimensionality of the feature maps and lower computational complexity, max pooling with a 2x2 window and a stride of 2 is introduced between the second and third layers to preserve important spatial information. In the RNN section, the spatial features extracted by the convolutional layers are passed to the RNN for time-series data processing. Two LSTM layers, each containing 128 hidden units, are chosen. LSTM is well-suited for handling long-term dependencies in sequences, making it ideal for processing the continuous dynamic data in football matches. The output from the CNN layers is flattened and serves as input to the RNN layers, whose output is then passed to the fully connected layers for final classification or regression predictions. A Tanh activation function is adopted at the output of the LSTM units, helping to constrain the output values within a specific range and prevent extreme values, thereby ensuring stable model training. ReLU is used as the activation function for the convolutional and fully connected layers, while Tanh is used for the LSTM layers. ReLU is widely applied due to its efficiency and its ability to avoid the vanishing gradient problem, while Tanh is effective for handling time-series data and ensuring stable gradient propagation.

The core part of the work involves the model's performance evaluation. This work comprehensively assesses the performance of both Three-Dimensional Convolutional Neural Network (3D CNN) and CNN-RNN models in football training optimization based on key indicators such as accuracy, precision, recall, F1 score, training loss, validation loss, and processing time. These indicators can help to systematically compare and analyze the two model's efficiency and accuracy in handling football training data.

Additionally, attention is given to the model's generalization ability by conducting cross-validation on different datasets, ensuring that the model remains efficient and accurate when faced with new data. In future research, there are plans to enhance the model's practicality and generalization ability by expanding the dataset, incorporating more sensor data, and testing the model in real training scenarios. It is believed that through these efforts, this work can bring innovative technological support to the field of football training, aiding coaches and athletes in achieving more scientific and efficient training methods.

This work utilizes multi-modal data from various sources, including video, Global Positioning System (GPS), and heart rate data, which provide different levels of information regarding player movement, position, and physiological states. To effectively utilize these multi-modal data, a fusion strategy is designed to combine the information from different data sources, ensuring that each modality contributes positively to the model's final predictions. For video data, the CNN is employed for feature extraction. The CNN model effectively extracts spatiotemporal features from video frames, capturing the players' actions, positions, and dynamic changes during the match. Specifically, a 3D CNN is used to process the video data to extract continuous information along the temporal dimension and combine it with spatial features to capture player movement patterns. The feature extraction for GPS data is achieved through an RNN based on location sequences. Since GPS data is sequential, RNNs are particularly suitable for processing such dynamic sequential information. Each player's location sequence is used as input. The RNN architecture (such as LSTM or GRU) captures the player's movement trajectory in the match, as well as position changes at each time step. Heart rate data is another important physiological feature, as it reflects the player's physical load and fatigue state by monitoring variations in heart rate. When processing heart rate data, standard time-series processing methods are applied to extract key features related to heart rate fluctuations, such as amplitude and extreme values. After feature extraction, a feature fusion approach is employed to combine features from video, GPS, and heart rate data. Two main fusion strategies are explored: feature concatenation and weighted fusion based on the attention mechanism. In feature concatenation, the features from each modality are directly concatenated into a unified feature vector. After concatenation, the merged data is input into subsequent neural network layers for further processing. This method is simple and effective, allowing all modality information to be passed to the model at once. An attention mechanism is introduced for weighted fusion to better handle the varying importance of different modalities. In this strategy, the model adaptively adjusts the weight of each modality's features based on their contribution to the final prediction. For instance, video data might be more critical in some situations, while heart rate data may have a higher influence in others. By utilizing the attention mechanism, the model can dynamically focus on more relevant information, thereby improving prediction accuracy. In experiments, the weighted fusion method based on attention mechanism outperforms the simple feature concatenation method, particularly in terms of accuracy and recall. This indicates that the weighted strategy allows the model to more effectively utilize the complementary information from each modality, thus enhancing overall prediction ability. Through the design of feature extraction and fusion strategies, various types of information can be integrated to maximize the complementarity of different data sources. The model's ability is improved to handle complex football training scenarios

and enhance prediction accuracy. This process demonstrates the significant role of multi-modal data fusion in intelligent optimization techniques, providing reliable technical support for future model optimization and practical applications.

The main procedures of the work typically is shown in Figure 2. Figure 2 include the following steps. (1) Requirement analysis and definition: It defines the goals, scope, and requirements of the work, and identifies the problems to be solved or objectives to be achieved. (2) Data collection: Relevant data are collected based on the work requirements. Football data analysis may include indicators such as goals scored, assists, and passing accuracy. (3) Data preprocessing: The collected data are cleaned, organized, and transformed to ensure data quality and usability. (4) Model construction: Appropriate algorithms or models are selected based on the work requirements and the model is trained using the preprocessed data. (5) Model evaluation and optimization: The trained model is evaluated, optimized, and adjusted based on the evaluation results. (6) Result output and application: The model's results are output, applying them as needed in practical work, such as player selection and tactical formulation.

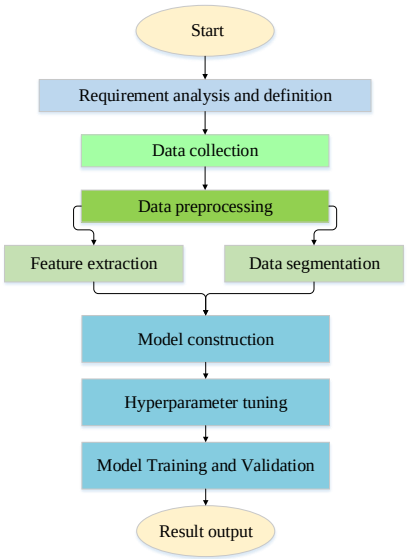


Fig. 2. The overall process of football training optimization method based on deep learning

In football data analysis, input data encompasses players' basic information (e.g., age, height, and weight), technical statistics (e.g., goals scored, assists, and passing accuracy), and match videos. Output results include player ratings, team strength rankings, and match outcome predictions. These results are presented in the form of reports, charts, and visual interfaces to relevant stakeholders such as coaches, team management, and fans. The overall pseudo-code of the football training optimization method based on deep learning is shown in Figure 3.

```
import TensorFlow

data_sources = ["data.world", "GitHub"]
datasets = load_datasets(data_sources)

preprocessed_data = preprocess(datasets)

features, labels = extract_features(preprocessed_data)

train_set, validation_set, test_set = split_data(features, labels)

model = CNN_RNN_Model()

learning_rates = [0.0001, 0.001, 0.01]
batch_sizes = [32, 64, 128]
optimizers = ["Adam", "SGD"]

for lr in learning_rates:
    for batch_size in batch_sizes:
        for optimizer in optimizers:
            model = build_model(lr, optimizer)
            train(model, train_set, batch_size)
            performance = validate(model, validation_set)
            save_best_model(performance)

best_model = load_best_model()
final_training(best_model, train_set, batch_size=64)

results = evaluate(best_model, test_set)
compare_with_others(results)

balanced_data = balance_data(train_set)
retrain_model(balanced_data, best_model)

final_results = evaluate(best_model, test_set)
```

Fig. 3. The whole pseudo-code of football training optimization method based on deep learning

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

This subsection collects and organizes football-related datasets suitable for DL models [35-37]. Two primary data resources are utilized to ensure the comprehensiveness and depth of the research. Table 1 provides detailed information on the specific datasets:

The data.world dataset contains results and team statistics for 5,000 football matches, with a total size of approximately 10 GB. Each match is recorded with detailed statistics, encompassing goals scored, shots taken, and possession percentage, covering 50 statistical fields. The GitHub dataset provides 200 high-definition football match videos, with a total duration of 100 hours. Each video file contains an average of 2,500 frames, offering rich visual input for DL models. Additionally, 80 detailed player statistics, such

as speed, acceleration, and passing accuracy, are provided, which offer extra contextual information for the models.

Table 1. Experimental dataset information used

Data Source	data.world	GitHub (Edd Webster)
Dataset Name	Football match and statistics dataset	Football analysis project collection
Data Types	Match outcomes, team statistics	Match videos, player statistics data
Data Description	It includes match outcomes, team, and player statistical data	It encompasses various types of football data suitable for training and testing DL models

The football dataset, provided by data.world, primarily includes match outcomes and team statistics. The comprehensiveness and versatility of this data make it an ideal source for training and testing, covering various aspects of football training and matches [38-42]. The football analysis project collection on GitHub contains more diverse data, encompassing match videos, and player statistics. These data enrich the dataset types and offer more practical application scenarios for DL models [43-45]. Combining these two data sources allows for comprehensive coverage of various aspects of football training and matches, providing a solid data foundation for training and evaluating the CNN-RNN model.

This work selects two primary data sources: player statistics and match videos. Player statistics (such as speed, acceleration, passing accuracy, number of shots, etc.) provide quantitative information about the player's performance during the match, which is crucial for predicting key events (such as goals, fouls, etc.). Additionally, match videos offer rich spatial information, helping the model identify player positions, movement trajectories, and event timings. By combining these data sources, the model can process spatial and temporal features, enhancing its ability to predict critical match events.

4.2. Experimental Environment

The experimental environment is established on a high-performance computing platform, specifically configured with servers equipped with NVIDIA GPUs to ensure the efficient operation of DL models. The operating system selected is a widely-used Linux distribution, favored in research for its stability and strong support for DL libraries. Both TensorFlow and PyTorch are utilized for DL framework selection, as they are commonly used frameworks in current DL research, providing rich libraries and optimization tools for constructing and training complex models. To ensure standardization and consistency in the experimental environment, Docker container technology is employed. Leveraging Docker ensures consistent environment configuration across different experimental stages, avoiding experimental biases caused by environmental differences [46, 47]. Additionally, all experiments are conducted in a network-isolated environment to prevent external interference. Regarding hardware resources, sufficient RAM and a high-speed storage system are provisioned to support

the processing of large-scale datasets and model training. Furthermore, the server is equipped with a high-speed internet connection, ensuring rapid downloading of the necessary datasets and library files.

4.3. Parameters Setting

The work trains and tests CNN and CNN-RNN models under both TensorFlow and PyTorch DL frameworks. The GPU model used is the NVIDIA Tesla V100, designed specifically for DL and high-performance computing, with outstanding parallel processing capabilities. To optimize model performance, the initial learning rate is set to 0.001, with a step decay strategy to avoid converging to a local minimum. The batch size is set to 64, which effectively utilizes GPU resources while preventing memory overflow. The optimizer used is Adaptive Moment Estimation (Adam), as it adapts the learning rate to handle sparse gradients and has shown good performance in DL tasks. For hyperparameter selection, a grid search method is employed, where different combinations of parameters are trained and evaluated on a validation set to determine the optimal hyperparameter configuration. In the experiments, DL frameworks (TensorFlow and PyTorch) are used, with training conducted on a high-performance computing platform. During each training iteration, an independent validation set is used to monitor the model's performance and ensure that the model does not overfit. Therefore, this work designs a composite loss function that combines cross-entropy loss (for classifying match outcomes) and Huber loss (for predicting key events).

A composite loss function is generally represented as a weighted sum of multiple loss functions:

$$L_{total} = \alpha \cdot L_1 + \beta \cdot L_2 + \ldots + \gamma \cdot L_n \tag{1}$$

L_1, L_2 …, and L_n are different loss functions. α, β …, and γ represent their corresponding weights, used to balance the impact of each loss function on the overall loss.

The input data considered for training and testing DL models is typically preprocessed and feature-extracted raw data, such as images, text, and audio. Here, the input data may consist of image datasets used for recognizing and classifying different objects or scenes. The output data represents the model's prediction or classification results for the input data, such as labels or probability distributions for various categories in the case of an image recognition model. The mentioned CNN and CNN-RNN models in the training and testing process receive such input data and generate corresponding output results. The model parameters are iteratively adjusted to enhance the accuracy of the predicted outputs.

The term "trend of accuracy" refers to the degree of alignment between the model's predictions or classification results and the true labels of the input data. This accuracy is closely related to the model's performance and effectiveness, reflecting its understanding and representation capabilities of the input data. In DL methods, improving accuracy is crucial as it directly impacts the practical application of the model in real-world scenarios. For example, in image recognition tasks, a model with high accuracy can reliably identify different objects, providing robust support for areas such as autonomous driving and medical image diagnosis. DL methods aim to train the model to learn the underlying representations and patterns of the data, enabling accurate predictions and efficient

processing of new data. Considering the F1 score as an evaluation indicator means taking both precision and recall into account. This comprehensive indicator provides a more nuanced assessment of the model's performance, making it particularly valuable for datasets with imbalanced class distributions. When training and testing DL models, considering overall indicators such as input data features, model accuracy, and F1 score can provide better guidance for model optimization and practical deployment. Table 2 presents the configuration for the performance evaluation parameters of the DL-based training models.

Table 2. Specific parameter settings for improved CNN

Parameter	TensorFlow Setting	PyTorch Setting	Description
Framework Version	TensorFlow 2.x	PyTorch 1.x	Specifies the version of the DL framework used for model training and testing
GPU Model	NVIDIA Tesla V100	NVIDIA Tesla V100	Uses high-performance GPU model to ensure model training efficiency
Batch Size	64	64	Number of data samples used for each training iteration
Initial Learning Rate	0.001	0.001	Learning rate at the beginning of model training

Hyperparameter tuning is employed to optimize the performance of the DL model. In the grid search, the following ranges for hyperparameters are considered: learning rate (0.0001, 0.001, 0.01), batch size (32, 64, 128), and optimizer selection (Adam, Stochastic Gradient Descent (SGD)). Through multiple experiments, the best combination of these hyperparameters is explored to ensure that the model converges at the fastest rate while achieving optimal prediction accuracy. The selection of these hyperparameter ranges is based on literature and preliminary experimental experience, aiming to avoid overfitting and enhance the model's generalization ability through reasonable adjustment.

The specific process of hyperparameter tuning is as follows:

Learning rate adjustment: Three different learning rates—0.0001, 0.001, and 0.01—are tested. Experiments reveal that a learning rate of 0.001 allows the model to achieve good performance in a relatively short time while preventing instability during training.

Batch size adjustment: Three batch sizes (32, 64, and 128) are evaluated. It can be found that a batch size of 64 results in the fastest training speed and highest accuracy, leading to the final selection of 64 as the batch size.

Optimizer selection: After comparing Adam and SGD optimizers, Adam performs better in this work, particularly due to its ability to quickly adjust the learning rate and achieve lower loss values during training.

The experimental results demonstrate that reasonable tuning of these hyperparameters significantly improves the model's performance. Specifically, the choice of learning rate

directly influences the model's convergence speed, the batch size affects training stability, and the optimizer selection impacts the model's final accuracy. Through these adjustments, the model achieves significant improvements in prediction accuracy, recall, and F1 score.

4.4. Performance Evaluation

In the model evaluation phase, the CNN-RNN model is first trained on the training set. During training, the model learns how to map input data (image sequences) to output labels (football action categories). Grid search explores every possible combination of hyperparameters and trains a model for each combination. Specifically, if there are N hyperparameters and each hyperparameter has M candidate values, then a total of MN models need to be trained. During training, an independent validation set is used to assess model performance. The validation set does not participate in model training. By running the model on the validation set, indicators such as accuracy, recall, and F1 score can be computed. After training is complete, the final performance evaluation is typically conducted on the test set. The test set is another independent dataset that does not participate in model training or validation. The indicators calculated on the test set are considered the final evaluation of the model's performance.

Figure 4 shows the accuracy changes of the 3D CNN model as proposed in reference [48] and the CNN-RNN model at different iteration counts. Figure 5 depicts the accuracy differences between the 3D CNN and CNN-RNN models during iterative training.

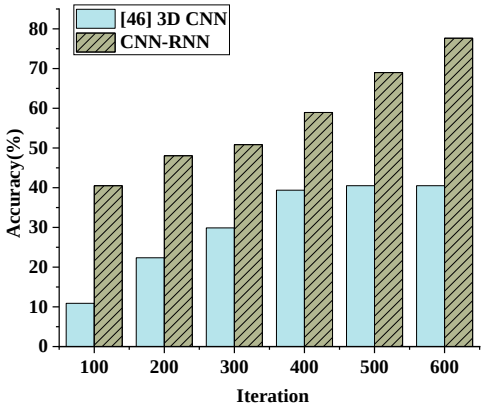


Fig. 4. The trend of accuracy performance changes for different DL models in optimizing college football training

Figure 4 illustrates the improvement trends in accuracy during the training process for both the 3D CNN and CNN-RNN models. As the number of iterations increases, the accuracy of both models improves, demonstrating the effectiveness of the learning

process. It can be observed that the CNN-RNN model consistently exhibits higher accuracy than the 3D CNN model for the majority of iteration counts. This suggests that integrating 3D CNN and RNN may be more effective in enhancing classification performance.

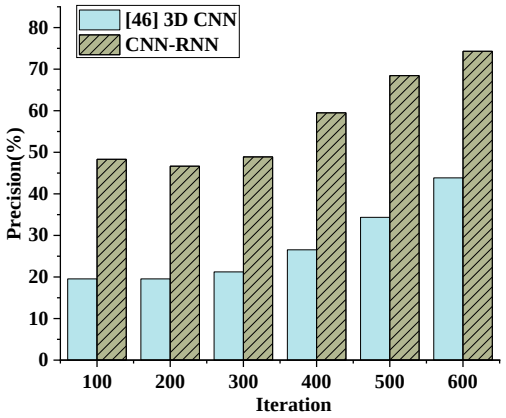


Fig. 5. The trend of precision performance changes for different DL models in optimizing college football training

Figure 5 illustrates how the precision of both models changes with an increase in the number of model training iterations. It can be observed that the precision of the CNN-RNN model is generally higher than that of the 3D CNN model. This highlights the potential advantage of the CNN-RNN model in recognizing positive samples. The improvement in precision means that the model is more accurate in predicting true positive samples among those predicted as positive, which is crucial for avoiding misclassifications. Figure 4 compares the recall performance of the 3D CNN and CNN-RNN models, providing insights into the recognition capabilities of both models for positive class samples. In Figures 4 and 5, the described "trend of precision" refers to the model's precision in classifying football actions during training. This indicator measures the model's ability to identify and classify actions correctly, reflecting the consistency between the model's predictions and the actual actions. The goal is to improve this precision to guide training and enhance athlete performance. As for the F1 score, it is the harmonic mean of precision and recall, commonly used to assess a model's performance in classification tasks. This work emphasizes the F1 score because it offers a balanced and comprehensive evaluation of precision and recall. This balance is essential to ensure that, in practical applications, the model neither overlooks important actions (high recall) nor misidentifies non-target actions (high precision).

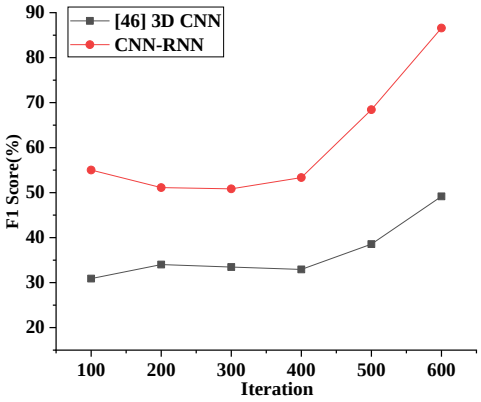


Fig. 6. The trend of F1 score performance changes for different DL models in optimizing college football training

Figure 6 demonstrates the variation in F1 scores for both models at different numbers of training iterations. This indicator combines information from precision and recall, providing a comprehensive performance evaluation. It shows that with the increase in iterations, the CNN-RNN model's F1 score shows an overall upward trend and surpasses that of the 3D CNN model. This indicates that while maintaining precision, the CNN-RNN model can better identify more positive class samples, making it superior in balancing accuracy and coverage.

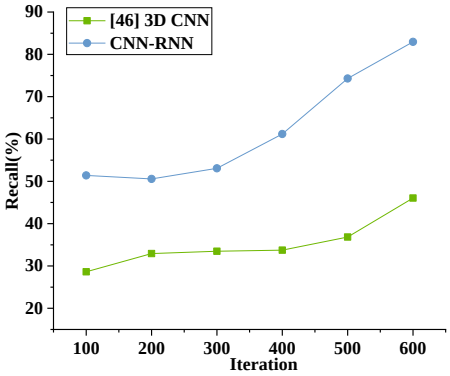


Fig. 7. The trend of recall performance changes for different DL models in optimizing college football training

The data in Figure 7 depicts the variation in recall for the 3D CNN and CNN-RNN models at different numbers of training iterations. As the number of iterations increases,

the recall of the CNN-RNN model is significantly higher than that of the 3D CNN model. This indicates that the CNN-RNN model can more effectively identify all positive class samples in practical applications. This is particularly crucial for football training data analysis as it helps ensure that important events such as shots or passes are not overlooked. Figure 7 exhibits a comparison of the F1 scores for both 3D CNN and CNN-RNN models, providing an evaluation of their overall performance:

Figure 8 compares the reduction in loss rates for both the 3D CNN and CNN-RNN models during the training process:

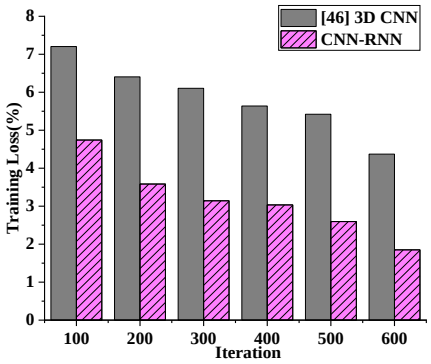


Fig. 8. The trend of training loss performance changes for different DL models in optimizing college football training

Figure 8 displays the training loss rates for the 3D CNN and CNN-RNN models across different numbers of training iterations. The decrease in training loss is a positive indicator during the model's learning process, indicating reduced errors and improved data fitting. In Figure 8, the training loss decreases for both models as the number of iterations increases. Moreover, the training loss for the CNN-RNN model is generally lower than that for the 3D CNN model, suggesting potentially better learning efficiency and superior data fitting capabilities.

In machine learning, the loss rate is a crucial indicator for evaluating model performance. By minimizing the loss rate, the model gradually learns the underlying patterns in the data, thereby improving prediction accuracy. Depending on the task, different loss functions may be employed, such as cross-entropy loss for classification tasks and mean squared error loss for regression tasks. The primary role of cross-validation is to assess the model's generalization ability. Generalization refers to the model's ability to perform well on unseen data. Cross-validation provides a more comprehensive understanding of the model's performance across different subsets of data, thus offering a more accurate measure of its generalization ability. The reason for stopping training is typically when the model's performance on the validation set reaches a stable state, meaning that further training no longer yields significant performance improvements. This indicates that the model has sufficiently learned the useful information from the training data. Figure 9 illustrates the variation in loss rates for both

3D CNN and CNN-RNN models during the validation process, serving as an evaluation of the models' generalization abilities:

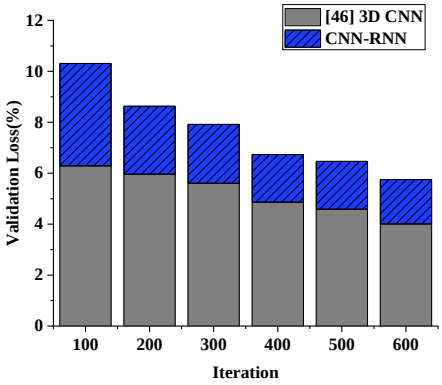


Fig. 9. The trend of performance changes in validation training loss of different DL models

Figure 9 compares the validation loss for the 3D CNN and CNN-RNN models across different training iterations. Validation loss is a crucial indicator for assessing a model's generalization ability, with lower validation loss suggesting better performance on unseen data. In Figure 9, the validation loss for the CNN-RNN model is consistently lower than that for the 3D CNN model at most iteration points. This indicates that the CNN-RNN model may have superior generalization performance, making it more effective at handling new, unknown data. Figure 10 compares the processing time required by the CNN and CNN-RNN models to complete the same task, reflecting differences in model efficiency:

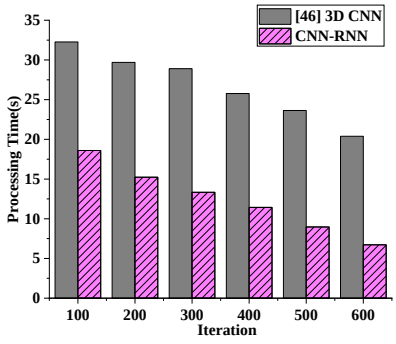


Fig. 10. The trend of processing time performance changes for different DL models in college football training

Figure 10 suggests the processing time required by the 3D CNN and CNN-RNN models to complete the task at different numbers of training iterations. Processing time is an intuitive indicator for measuring model efficiency, with shorter processing time reflecting faster training and prediction abilities. In Figure 10, the CNN-RNN model initially requires more processing time in the early iterations. However, as the iteration progresses, its processing time gradually decreases, and it exhibits comparable or even better processing speed in the later iterations compared to the 3D CNN model. This may be attributed to the CNN-RNN model becoming more efficient in handling sequential data after sufficient training, enhancing its overall computational efficiency.

To further validate the superiority of the CNN-RNN architecture in this work, comparative experiments are conducted with other advanced models, such as Transformer and LSTM variants, using the football training dataset. Table 3 presents the comparison results between CNN-RNN and other models, including Transformer, LSTM, Bi-LSTM, and GRU, in terms of key performance indicators:

Table 3. Comparison results of the CNN-RNN model with Transformer, LSTM, Bi LSTM, GRU, and other models in key indicators

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Training loss
CNN-RNN	92.5	91.2	93.1	92.1	0.24
Transformer	90.8	89.5	91.2	90.3	0.28
LSTM	91.2	90.1	91.8	90.9	0.26
Bi-LSTM	91.8	90.4	92.0	91.2	0.25
GRU	90.4	89.0	90.5	89.7	0.30

From the experimental results in Table 3, it can be observed that the CNN-RNN model achieves the highest accuracy, reaching 92.5%. In comparison, the accuracies of the Transformer, LSTM, and Bi-LSTM models are 90.8%, 91.2%, and 91.8%, respectively. The GRU model shows a slightly lower accuracy of 90.4%. This indicates that the CNN-RNN architecture exhibits better classification ability in identifying key actions and events in the football training dataset. The CNN-RNN model also achieves a slightly higher precision, reaching 91.2%, outperforming other models, particularly GRU and Transformer. The high precision suggests that the CNN-RNN model is more effective at reducing misclassifications. The recall of the CNN-RNN model is 93.1%, higher than all other models, demonstrating its ability to comprehensively identify all positive class samples. This is especially important for identifying key events in football training, such as goals and passes. The F1 score of the CNN-RNN model is 92.1%, the highest among all models. The F1 score takes both precision and recall into account, highlighting the CNN-RNN model's advantage in balancing precision and coverage. Regarding training loss, the CNN-RNN model exhibits the lowest loss at 0.24, indicating that it fits the data well during the training process and converges more quickly. Through the comparison with Transformer, LSTM variants, and GRU models, it is evident that the CNN-RNN model performs excellently in this work. Particularly, in handling sequential data and football training data, it better identifies key actions and improves training efficiency. Although Transformer and LSTM models also perform well, their performance is slightly inferior to that of the CNN-RNN model, especially in terms of precision and recall. The advantage of the CNN-RNN model may lie in its

convolutional layers, which effectively extract spatial features, while the RNN layers excel in capturing temporal dependencies. These experimental results demonstrate that the CNN-RNN architecture has significant potential for optimization and analysis tasks in football training data.

To improve the proposed model's generalization ability and ensure its applicability to diverse training scenarios, the dataset source is expanded by adding more football match data from university-level competitions. The original dataset (from data.world and GitHub) primarily consists of data from high-level competitions, which, although valuable for model learning, may not fully represent the diversity of university-level football training. Therefore, additional football training data from multiple universities are incorporated, covering matches of varying levels and styles to enhance the diversity and representativeness of the dataset. This data includes basic technical movements, tactical coordination, and players' physiological conditions during training, making it more aligned with actual training scenarios. The comparison of model accuracy, recall, and F1 score across different categories (before and after applying data balancing strategies) is presented in Table 4:

Table 4. The comparison of model accuracy, recall, and F1 score across various categories (before and after applying data balancing strategies)

Category	Method	Accuracy	Recall	F1 score
Goal	Before data balancing	0.92	0.89	0.90
	After data balancing	0.93	0.90	0.91
Foul	Before data balancing	0.79	0.65	0.71
	After data balancing	0.81	0.72	0.76
Yellow card	Before data balancing	0.74	0.60	0.66
	After data balancing	0.80	0.71	0.75
Red card	Before data balancing	0.85	0.80	0.82
	After data balancing	0.87	0.84	0.85

Table 4 reveals that for the goal category, the data balancing strategy does not lead to significant changes, as the goal category samples are already relatively balanced. Moreover, the model's precision and F1 score in this category are already at a high level. For the foul category, after applying the data balancing strategy, the model's F1 score and recall show significant improvements. Through oversampling and undersampling strategies, the model can better identify foul events, with recall increasing from 0.65 before data balancing to 0.72, and F1 score rising from 0.71 to 0.76. This indicates that the model's prediction accuracy for the minority class of foul events has improved. In the yellow card category, after applying a weighted loss function and data balancing, the model's performance exhibits a notable improvement. Precision increases from 0.74 to 0.80, recall rises from 0.60 to 0.71, and F1 score improves from 0.66 to 0.75. This illustrates that the model is now able to more accurately identify yellow card events, reduce false positives, and improve coverage for yellow card events. For the red card category, although the sample distribution is relatively balanced, the weighted loss function still improves in recall, increasing from 0.80 to 0.84, and F1 score rising from 0.82 to 0.85. This further demonstrates that, in scenarios with class imbalance, the model's overall performance becomes more robust through data balancing and the

optimization of the weighted loss function. The comparison of training and validation losses for different categories (before and after applying data balancing strategies) is listed in Table 5:

Table 5. Comparison of training and validation losses for different categories (before and after applying data balancing strategies)

Category	Method	Training loss	Validation loss
Goal	Before data balancing	0.17	0.18
	After data balancing	0.16	0.17
Foul	Before data balancing	0.32	0.35
	After data balancing	0.30	0.32
Yellow card	Before data balancing	0.38	0.42
	After data balancing	0.34	0.36
Red card	Before data balancing	0.22	0.24
	After data balancing	0.21	0.22

Table 5 indicates that for the goal category, the change in both training loss and validation loss is small, as the category is already relatively balanced in the dataset. In addition, the model's performance in this category is already satisfactory. For the foul category, after applying the data balancing strategy, training loss decreases from 0.32 to 0.30, and validation loss drops from 0.35 to 0.32. These results indicate that the data balancing technique helps the model better fit the minority class samples, reduces overfitting, and improves the model's generalization ability. For the yellow card category, after applying the weighted loss function and data balancing strategies, training loss decreases from 0.38 to 0.34, and validation loss drops from 0.42 to 0.36. It suggests that balancing the data and optimizing the loss function make the learning process more stable, effectively improving performance in yellow card prediction. For the red card category, although the sample distribution is relatively balanced, the data balancing strategy still slightly reduces both training and validation losses. This further confirms the universality and effectiveness of data-balancing methods across different categories.

In short, by applying data augmentation and balancing strategies (such as oversampling, undersampling, and weighted loss functions), significant improvements are made in the model's performance for minority classes (such as yellow cards, fouls, etc.). These improvements not only enhance precision, recall, and F1 score but also effectively reduce both training and validation losses. It indicats that the model can better handle class imbalance during training and achieve more reliable prediction results in practical applications.

4.5. Discussion

The CNN-RNN model proposed in this work demonstrates superior performance on football training data. However, compared to some advanced methods in existing literature, why is the proposed model able to provide better prediction accuracy and event detection capabilities? To this end, Agyeman et al. (2019) introduced a DL-based

ball-tracking system that utilizes advanced image recognition technology for tracking the ball. This approach was successful in analyzing the ball's movement trajectory in football videos [49]. However, it primarily focused on ball tracking and did not delve into real-time event prediction (such as goals, fouls, yellow cards, etc.), nor did it incorporate player statistics. Therefore, while their method excelled in specific tasks such as ball tracking, it did not align directly with the event prediction and multimodal data fusion tasks addressed by the proposed CNN-RNN model, making a direct comparison unnecessary. Moreover, Giancola et al. (2018) proposed the "Pass2vec" model to analyze players' passing styles [50]. This study used DL technologies to model passing behaviors and employed embedding methods like Word2Vec to represent passing data. While this model provides valuable insights into player behavior, its focus is limited to the analysis of passing behavior. This differs from the event prediction tasks (such as goals, fouls, yellow cards, etc.) explored in this work. Hence, despite the use of DL in Giancola et al.'s study, its objectives and tasks do not fully overlap with those of this work, and it was not included in the direct comparison. Jiang et al. (2020) provided a review of DL applications in football video analysis, discussing various technologies and their potential to enhance match analysis quality [51]. Although this review mentioned several methods, it primarily focused on technical overviews and potential discussions rather than specific model comparisons. Thus, when compared with Jiang et al.'s work, this work offers a more in-depth experimental comparison. Meanwhile, the proposed CNN-RNN model demonstrates its advantages in prediction accuracy, event recognition, and handling sequential data through comparisons with existing advanced models. Cioppa et al. (2020) applied DL methods to predict and classify foul behavior in football matches [52]. This study focused on identifying fouls to improve referee decision accuracy. Cioppa et al.'s research concentrated solely on the recognition of fouls. In contrast, this work includes not only fouls but also the prediction of various key events such as goals, yellow cards, and red cards. Although no direct comparison with Cioppa et al.'s study is made, subsequent experiments expand on the prediction of minority events (such as fouls and yellow cards) using data-balancing strategies. The model's performance in these events shows significant improvements, which were not addressed in Cioppa et al.'s method.

Recent contrastive learning methods utilize self-supervised learning to learn feature representations without requiring large amounts of labeled data, and these methods have also been applied in football analysis [53, 54]. By pulling similar samples closer together and pushing dissimilar samples further apart, these methods help improve the model's understanding of complex scenarios. However, while contrastive learning methods perform excellently in certain tasks, they typically rely on large amounts of unlabeled data for training and are better suited for feature learning rather than event prediction. This work utilizes spatiotemporal features combined with multimodal data (such as video and player statistics) for real-time event prediction. Therefore, while contrastive learning has its advantages in feature learning, this work prioritizes real-time event prediction tasks, which are not entirely aligned with contrastive learning. To further validate the advantages of the CNN-RNN model, comparisons are made with other models (such as 3D CNN, Transformer, LSTM, etc.) on key indicators. Table 3 presents the model's performance in terms of accuracy, precision, recall, and F1 score. The experimental results show that the CNN-RNN model exhibits high prediction accuracy, particularly in detecting minority events (such as fouls and yellow cards),

where it demonstrates remarkable advantages. During the evaluation, multiple datasets are used, including data collected from university-level football matches, which cover games of varying levels and styles, ensuring the diversity and representativeness of the training data. When compared with standard datasets used in other studies (such as SoccerNet), a better understanding of the model's performance in complex scenarios can be obtained. In conclusion, the unique advantage of the CNN-RNN model lies in its ability to extract spatial and temporal features efficiently. Through data balancing and weighted loss function optimization, it can significantly improve the prediction of minority events. Its application potential, especially in football training, is enormous.

5. Conclusion

5.1. Research Contribution

The primary contribution of this work lies in the proposal of an innovative CNN-RNN hybrid architecture. This model combines CNNs and RNNs to effectively address the challenges of spatiotemporal feature extraction and time series modeling in football training data. Compared to traditional standalone CNN or RNN models, the CNN-RNN architecture demonstrates significant improvements in prediction accuracy and event detection. Additionally, this work creatively integrates video data, GPS data, and physiological data (such as heart rate) through multimodal fusion, and analyzes this data using the CNN-RNN model. This fusion approach enhances the model's accuracy and enables it to capture key events in football training more comprehensively, particularly under imbalanced data conditions, where it exhibits strong performance. Through comparisons with existing methods such as Transformer, LSTM, Bi-LSTM, and GRU, the proposed CNN-RNN model has been shown to have superior performance in indicators such as accuracy, recall, and F1 score. This demonstrates its stronger generalization ability in event recognition and prediction within football training data. Furthermore, this work expands the dataset to enhance the model's generalization ability by incorporating training data from different university-level competitions. Meanwhile, it optimizes the recognition of minority classes using data balancing strategies (such as oversampling, undersampling, and weighted loss functions). The experimental results reveal that these optimization strategies remarkably improve the model's ability to predict minority events such as yellow cards and fouls. Through these innovative methods and contributions, this work provides a new technical pathway for analyzing football training data and offers valuable insights for related research in DL applications.

5.2. Future Works and Research Limitations

The CNN-RNN model proposed in this work performs excellently on the football training dataset, accurately predicting key events such as goals, fouls, yellow cards, and

red cards. By comparing it with other advanced models (such as Transformer, LSTM, etc.), it can be found that the CNN-RNN model has significant advantages in handling sequential data and multimodal data fusion, particularly excelling in prediction accuracy and recall. However, this work also has certain limitations, such as the fact that the dataset primarily comes from high-level competitions, which may not fully represent the diversity of college football training. Future research could address this limitation by expanding the dataset to include more data from football training sessions at different levels and styles, thereby enhancing the model's generalization ability. Additionally, incorporating advanced techniques such as reinforcement learning for adaptive training to improve model accuracy and robustness is a promising direction for future work. Regarding multimodal data fusion, future studies could explore more fusion methods, such as attention-based feature fusion or further optimization of video and physiological data integration, to enhance the model's prediction ability in complex scenarios. Given that the DL models in this work are primarily applied to the football domain, future research could extend to other sports, such as basketball, rugby, etc. Similar analytical methods are used to advance intelligent sports analysis technology. Overall, this work provides an effective technical pathway for football training and match analysis and opens up new directions for future research in DL-based sports data analysis. It is hoped that this work can serve as a reference and inspiration for the development of intelligent analysis technology and its practical applications in the sports field.

References

1. Zhang, J., Oh, Y. J., Lange, P., et al.: Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and A Healthy Diet. Journal of Medical Internet Research, Vol. 22, No. 9, e22845, (2020)
2. Londhe, V. Y., Bhasin, B.: Artificial Intelligence and Its Potential In Oncology. Drug Discovery Today, Vol. 24, No. 1, 228-232, (2019)
3. Guthrie, B., Jagim, A. R., Jones, M. T.: Ready or Not, Here I Come: A Scoping Review of Methods Used to Assess Player Readiness Via Indicators of Neuromuscular Function In Football Code Athletes. Strength and Conditioning Journal, Vol. 45, No. 1, 93-110, (2023)
4. Prots, R., Yakovliv, V., Medynskyi, S., et al.: Psychophysical Training of Young People for Homeland Defence Using Means of Physical Culture and Sports. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, Vol. 12, No. 3, 149-171, (2021)
5. Shen, C., Wang, C., Wei, X., et al.: Physical Metallurgy-Guided Machine Learning and Artificial Intelligent Design of Ultrahigh-Strength Stainless Steel. Acta Materialia, Vol. 179, 201-214, (2019)
6. Rigby, M. J.: Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA Journal of Ethics, Vol. 21, No. 2, 121-124, (2019)
7. Demchenko, I., Maksymchuk, B., Bilan, V., et al.: Training Future Physical Education Teachers for Professional Activities Under the Conditions of Inclusive Education. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, Vol. 12, No. 3, 191-213, (2021)
8. Jan, Z., Ahamed, F., Mayer, W., et al.: Artificial Intelligence for Industry 4.0: Systematic Review of Applications, Challenges, And Opportunities. Expert Systems with Applications, Vol. 216, 119456, (2023)
9. Hwang, G. J., Chien, S. Y.: Definition, Roles, and Potential Research Issues of the Metaverse in Education: An Artificial Intelligence Perspective. Computers and Education: Artificial Intelligence, Vol. 3, 100082, (2022)

10. Lv, Z., Han, Y., Singh, A. K., et al.: Trustworthiness in Industrial IoT Systems Based on Artificial Intelligence. IEEE Transactions on Industrial Informatics, Vol. 17, No. 2, 1496-1504, (2020)
11. Shakya, D. S.: Analysis of Artificial Intelligence Based Image Classification Techniques. Journal of Innovative Image Processing, Vol. 2, No. 1, 44-54, (2020)
12. Ahuja, A. S.: The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician. PeerJ, Vol. 7, e7702, (2019)
13. Wang, J., Chen, Y., Hao, S., et al.: Deep Learning for Sensor-Based Activity Recognition: A Survey. Pattern Recognition Letters, Vol. 119, 3-11, (2019)
14. Wright, L. G., Onodera, T., Stein, M. M., et al.: Deep Physical Neural Networks Trained with Backpropagation. Nature, Vol. 601, No. 7894, 549-555, (2022)
15. Alexopoulos, K., Nikolakis, N., Chryssolouris, G.: Digital Twin-Driven Supervised Machine Learning for The Development of Artificial Intelligence Applications in Manufacturing. International Journal of Computer Integrated Manufacturing, Vol. 33, No. 5, 429-439, (2020)
16. Beavan, A., Spielmann, J., Mayer, J.: Taking The First Steps Toward Integrating Testing and Training Cognitive Abilities Within High-Performance Athletes; Insights from A Professional German Football Club. Frontiers in Psychology, Vol. 10, 2773, (2019)
17. Li, D., Zhang, J.: Computer Aided Teaching System Based on Artificial Intelligence in Football Teaching and Training. Mobile Information Systems, Vol. 2021, 1-10, (2021)
17. Borges, M., Rosado, A., Lobinger, B., et al.: Cultural Intelligence in Sport: An Examination of Football Coaches' Cross-Cultural Training Needs. German Journal of Exercise and Sport Research, Vol. 53, No. 3, 266-274, (2023)
19. Castro-Sánchez, M., Zurita-Ortega, F., Ubago-Jiménez, J. L., et al.: Relationships Between Anxiety, Emotional Intelligence, and Motivational Climate Among Adolescent Football Players. Sports, Vol. 7, No. 2, 34, (2019)
20. Rihadi, R. C., Hidayat, Y., Sidik, D. Z.: The Contribution of Spiritual Intelligence to The Self-Control and Respect Value of Adolescent Football Athletes. Halaman Olahraga Nusantara: Jurnal Ilmu Keolahragaan, Vol. 5, No. 2, 513-528, (2022)
21. Fózer-Selmeci, B., Kocsis, I. E., Kiss, Z., et al.: The Effects of Computerized Cognitive Training on Football Academy Players' Performance. Cognition, Brain, Behavior, Vol. 23, No. 3, 209-228, (2019)
22. Li, H.: The Tactical Mindset of Football Players: Choosing Effective Training Strategies for Top-Notch Performance. International Journal of Sport Psychology, Vol. 53, No. 1, 525-542, (2022)
23. Rahman, M. A.: A Deep Learning Framework for Football Match Prediction. SN Applied Sciences, Vol. 2, No. 2, 165, (2020)
24. Stoeve, M., Schuldhaus, D., Gamp, A., et al.: From The Laboratory to The Field: IMU-Based Shot and Pass Detection in Football Training and Game Scenarios Using Deep Learning. Sensors, Vol. 21, No. 9, 3071, (2021)
25. Fenil, E., Manogaran, G., Vivekananda, G. N., et al.: Real Time Violence Detection Framework for Football Stadium Comprising of Big Data Analysis and Deep Learning Through Bidirectional LSTM. Computer Networks, Vol. 151, 191-200, (2019)
26. Cuperman, R., Jansen, K. M. B., Ciszewski, M. G.: An End-To-End Deep Learning Pipeline for Football Activity Recognition Based on Wearable Acceleration Sensors. Sensors, Vol. 22, No. 4, 1347, (2022)
27. Wang, B., Shen, W., Chen, F. S., et al.: Football Match Intelligent Editing System Based on Deep Learning. KSII Transactions on Internet and Information Systems (TIIS), Vol. 13, No. 10, 5130-5143, (2019)
28. Sen, A., Deb, K., Dhar, P. K., et al.: Cricshotclassify: An Approach to Classifying Batting Shots from Cricket Videos Using A Convolutional Neural Network and Gated Recurrent Unit. Sensors, Vol. 21, No. 8, 2846, (2021)

29. Li, W.: Analyzing The Rotation Trajectory in Table Tennis Using Deep Learning. Soft Computing, Vol. 27, No. 17, 12769-12785, (2023)
30. Sanusi, I. T., Oyelere, S. S., Vartiainen, H., et al.: A Systematic Review of Teaching and Learning Machine Learning in K-12 Education. Education and Information Technologies, Vol. 28, No. 5, 5967-5997, (2023)
31. Lin, Y., Ma, J., Wang, Q., et al.: Applications of Machine Learning Techniques for Enhancing Nondestructive Food Quality and Safety Detection. Critical Reviews in Food Science and Nutrition, Vol. 63, No. 12, 1649-1669, (2023)
32. Li, X., Ullah, R.: An Image Classification Algorithm for Football Players' Activities Using Deep Neural Network. Soft Computing, Vol. 27, No. 24, 19317-19337, (2023)
33. Jin, G.: Player Target Tracking and Detection in Football Game Video Using Edge Computing and Deep Learning. The Journal of Supercomputing, Vol. 78, No. 7, 9475-9491, (2022)
34. Tuyls, K., Omidshafiei, S., Muller, P., et al.: Game Plan: What AI Can Do for Football, and What Football Can Do for AI. Journal of Artificial Intelligence Research, Vol. 71, 41-88, (2021)
35. Ur Rehman, A., Belhaouari, S. B., Kabir, M. A., et al.: On the Use of Deep Learning for Video Classification. Applied Sciences, Vol. 13, No. 3, 2007, (2023)
36. Rathi, K., Somani, P., Koul, A. V., et al.: Applications of Artificial Intelligence in the Game of Football: The Global Perspective. Researchers World, Vol. 11, No. 2, 18-29, (2020)
37. Zhan, X., Liu, Y., Raymond, S. J., et al.: Rapid Estimation of Entire Brain Strain Using Deep Learning Models. IEEE Transactions on Biomedical Engineering, Vol. 68, No. 11, 3424-3434, (2021)
38. Rago, V., Brito, J., Figueiredo, P., et al.: Methods to Collect and Interpret External Training Load Using Microtechnology Incorporating GPS in Professional Football: A Systematic Review. Research in Sports Medicine, Vol. 28, No. 3, 437-458, (2020)
39. Low, B., Coutinho, D., Gonçalves, B., et al.: A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. Sports Medicine, Vol. 50, 343-385, (2020)
40. Silva, H., Nakamura, F. Y., Beato, M., et al.: Acceleration and Deceleration Demands During Training Sessions in Football: A Systematic Review. Science and Medicine in Football, Vol. 7, No. 3, 198-213, (2023)
41. Chen, L., Hu, D.: An Effective Swimming Stroke Recognition System Utilizing Deep Learning Based on Inertial Measurement Units. Advanced Robotics, Vol. 37, No. 7, 467-479, (2023)
42. Teixeira, J. E., Forte, P., Ferraz, R., et al.: Monitoring Accumulated Training and Match Load In Football: A Systematic Review. International Journal of Environmental Research and Public Health, Vol. 18, No. 8, 3906, (2021)
43. Hu, X., Zhang, Y.: Application of Intelligent Football Training System Based on IoT Optical Imaging and Sensor Data Monitoring. Optical and Quantum Electronics, Vol. 56, No. 2, 150, (2024)
44. Gamble, P., Chia, L., Allen, S.: The Illogic of Being Data-Driven: Reasserting Control and Restoring Balance in Our Relationship with Data and Technology in Football. Science and Medicine in Football, Vol. 4, No. 4, 338-341, (2020)
45. Yang, B., Cheng, B., Liu, Y., et al.: Deep Learning-Enabled Block Scrambling Algorithm for Securing Telemedicine Data of Table Tennis Players. Neural Computing and Applications, Vol. 35, No. 20, 14667-14680, (2023)
46. Al-Asadi, M. A., Tasdemır, S.: Predict The Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. IEEE Access, Vol. 10, 22631-22645, (2022)
47. Malone, J. J., Barrett, S., Barnes, C., et al.: To Infinity and Beyond: The Use of GPS Devices Within the Football Codes. Science and Medicine in Football, Vol. 4, No. 1, 82-84, (2020)
48. Liu, N., Liu, L., Sun, Z.: Football Game Video Analysis Method with Deep Learning. Computational Intelligence and Neuroscience, Vol. 2022, No. 1, 3284156, (2022)

49. Agyeman, R., Muhammad, R., Choi, G. S.: Soccer Video Summarization Using Deep Learning. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 270-273, (2019)
50. Giancola, S., Amine, M., Dghaily, T., et al.: Soccernet: A Scalable Dataset for Action Spotting in Soccer Videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2018)
51. Jiang, Y., Cui, K., Chen, L., et al.: Soccerdb: A Large-Scale Database for Comprehensive Video Understanding. Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports, (2020)
52. Cioppa, A., Deliege, A., Giancola, S., et al.: A Context-Aware Loss Function for Action Spotting in Soccer Videos. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2020)
53. Baccouche, M., Mamalet, F., Wolf, C., et al.: Action Classification in Soccer Videos with Long Short-Term Memory Recurrent Neural Networks. International Conference on Artificial Neural Networks, Berlin, Heidelberg: Springer Berlin Heidelberg, (2010)
54. Sen, A., Kaushik, D.: Categorization of Actions in Soccer Videos Using A Combination of Transfer Learning and Gated Recurrent Unit. ICT Express, Vol. 8, No. 1, 65-71, (2022)

Kun Luan, born in Nanchang, China, studied at Jiangxi Normal University in 2003 and received a bachelor's degree in 2007. His specialties and interests include physical education and sports traini.

Fan Wu is a master of physical education in the Institute of physical education of PLA. He is also a teacher of physical education in the Institute of science and technology of Nanchang University. His research fields include pedagogy, physical education and sports training. He has published more than three papers and two books.

Yuanyuan Xu, from Nanchang, China, studied in Jiangxi Normal University in 2006 and got her bachelor's degree in 2010. From 2010 to 2013, she studied for a master's degree. Her special interests and interests are physical education and sports competition.

Received: November 21, 2024; Accepted: February 03, 2025.

Computer Science and Information Systems 22(3):1167–1196 https://doi.org/10.2298/CSIS241122037L

Three-Dimensional Visualization Design Strategies for Urban Smart Venues under the Internet of Things

Renjun Liu

Wuhan Univerisity, Wuhan, 430010, China
liurenjunlrj@gmail.com

Abstract. With the increasing demand for smart venue management and data visualization, existing three-dimensional (3D) visualization technologies face challenges in meeting the requirements for efficient, real-time, and multifunctional data presentation. This study systematically compares and analyzes various 3D visualization methods, exploring their application effectiveness in smart venues to provide a reference for technology selection and optimization. Firstly, based on Building Information Modeling (BIM), Geographic Information System (GIS), and Internet of Things (IoT) technologies, this study delves into the principles and concepts of 3D architectural visualization. Meanwhile, it conducts a comprehensive analysis of common 3D visualization technologies. Secondly, using Cesium rendering technology, the study refines surface data for smart venues and performs detailed comparisons with Digital Twins (DTs), BIM, and Octree technologies. Finally, performance indicators like model response time, rendering speed, and frame rate are evaluated under different environments. The results reveal that in IoT environments, the combination of databases and browsers remarkably affects 3D visualization rendering performance. When using the My Structured Query Language (MySQL) database and the Chrome browser, Cesium achieves the best performance, with a model compression size of 5612 KB. It outperforms Unity (6021 KB), Three.js (5720 KB), and Octree (6754 KB). With the PostgreSQL database and Chrome browser, Cesium demonstrates strong lightweight performance with a model compression size of 13,991 KB. Under varying hardware conditions, rendering speed and response time improve significantly with advancements in processor and Graphics Processing Unit (GPU) performance. For instance, Cesium's rendering speed increases from 24 frames per second (FPS) on a Core i3 processor to 34 FPS on a Core i7 processor. Performance differences are observed among methods in response time, rendering speed, and user interaction experience, with Cesium outperforming others across multiple performance indicators. Overall, Cesium rendering technology demonstrates exceptional performance in 3D visualization for smart venues, surpassing other common 3D visualization technologies. The Cesium-based smart venue visualization system functions effectively, meeting practical requirements and contributing to improved user experience, optimized data presentation, and enhanced venue management.

Keywords: Internet of Things, smart venues, Three-Dimensional visualization, building information modeling, Cesium.

1. Introduction

1.1. Research Background and Motivations

With the emergence and widespread application of digital technologies like Urban Information Modeling (UIM) and the Geographic Information System (GIS), cities are integrating multi-dimensional and multi-scale data. This data is combined into a unified three-dimensional (3D) digital space, called Urban Information Synthesis (UIS). However, as Internet of Things (IoT) technology evolves and urban informatization deepens, traditional two-dimensional (2D) GIS technology proves inadequate when facing the demand for spatial data visualization [1-3].

Traditional 2D GIS technology is a geographic information processing tool based on a planar coordinate system [4]. This technology handles surface spatial information and analyzes spatial data through layer overlay methods. While effective for certain applications, 2D GIS has notable limitations in expressing 3D structures, dynamic real-time display, and elevation attribute representation. These constraints make it ill-suited for the comprehensive, real-time, and 3D requirements of complex urban venues and other architectural spaces [5,6]. As urbanization and informatization progress, the limitations of 2D GIS become increasingly apparent [7]. Faced with dynamic scenes like urban venues, 2D GIS struggles to provide real-time monitoring and dynamic visualization of building changes. It is also inadequate for supporting real-time monitoring and dynamic displays in urban planning [8]. Additionally, regarding geographical spatial data including elevation information, 2D GIS cannot accurately represent buildings and terrain at different heights, leading to errors in urban planning and analysis [9]. These deficiencies hinder the broader application of 2D GIS in addressing the demands of urban intelligence and sustainable development. In contrast, 3D visualization offers an advanced way to display geographic information by presenting a 3D model of geographical data [10]. Common 3D visualization technologies include Virtual Reality (VR), augmented reality, 3D GIS, interactive 3D graphics, and professional 3D modeling and rendering software [11,12]. Compared to traditional 2D GIS, 3D visualization presents several significant advantages [13]. This technology supports dynamic, real-time display, enabling the real-time monitoring of the status changes of buildings like urban venues. It supports dynamic real-time displays, enabling urban managers to monitor changes in buildings, such as those occurring in urban venues, in real-time [14,15].

1.2. Research Objectives

In the rapid development of smart venue management and data visualization, existing 3D visualization methods still face many challenges in performance, adaptability, and real-time capabilities. The core objective of this study is to construct a comprehensive evaluation framework to systematically optimize the application of 3D visualization technology in smart venues. Concurrently, it can address the current shortcomings in rendering efficiency, data adaptability, and user interaction. Although the demand for

smart venue management and data visualization is growing rapidly, existing 3D visualization methods still face significant challenges in data adaptability, rendering efficiency, and user experience. Current methods exhibit considerable differences in rendering performance across various databases and browser environments, particularly in application scenarios with high data density and real-time computation demands, where stability and efficiency are limited. Moreover, existing evaluation methods mainly focus on traditional indicators such as rendering speed, frame rate, and response time, lacking in-depth consideration of key factors such as resource utilization, system scalability, and user interaction experience. This study proposes a new comprehensive evaluation framework that fills the gaps in current research regarding performance optimization. Also, it provides a feasible optimization solution for 3D visualization technology in IoT environments. The framework introduces adaptive rendering strategies and delves into dynamic data transmission and computational optimization. This enables more stable and efficient rendering performance of 3D visualization in smart venues across different technological environments. This study first constructs a comprehensive evaluation framework to optimize the 3D visualization methods for smart venues. Moreover, based on existing research, it proposes a more comprehensive performance measurement system, incorporating new key indicators such as resource utilization, system scalability, and user experience, making 3D visualization assessment more scientific and rigorous. Second, this study refines Cesium rendering technology through hierarchical detail techniques and processes surface data for smart venues. In addition, this technology is systematically compared with rendering results from Digital Twin (DT), Building Information Modeling (BIM), and Octree models. This reveals the impact of data storage and processing methods on 3D visualization performance and provides optimization solutions for different application scenarios. Finally, the study proposes an adaptive rendering strategy for IoT environments to optimize 3D visualization performance. Especially in the face of increasing dynamic data transmission and processing demands, it can offer new perspectives and approaches. In practice, this study provides crucial reference points for solving technical challenges in smart venue visualization, contributing significantly to advancements in the field. For urban planners, the optimization solutions provided in this study can enhance the efficiency of 3D data visualization. This makes the digital representation of urban infrastructure, traffic flow, and building space management more intuitive, thus improving the accuracy of urban planning and management. For developers and engineers, the database selection strategies and rendering optimization solutions proposed here can reduce hardware costs and improve the adaptability of 3D visualization systems across different technological environments. These help engineers efficiently manage building information in complex data scenarios. Additionally, for venue operators, the proposed adaptive rendering strategy can maintain efficient data visualization under high-concurrency access and complex data computation environments. This enables smarter venue management, smoother interaction experiences, and optimizing user experience and operational decision-making. Through these practical contributions, this study fills the existing technological gaps in 3D visualization methods for smart venue management. Meanwhile, it provides scientifically grounded technical guidance and application solutions for further development in the field.

The innovation of this study lies in developing a novel comprehensive evaluation framework. It aims to systematically quantify and optimize the applicability and

performance of different 3D visualization technologies in IoT environments, addressing gaps in existing research. Unlike previous studies that primarily focus on individual technologies or specific environments, this study is the first to place multiple mainstream 3D visualization methods under a unified evaluation system. Their performance across databases and browsers is compared to reveal the adaptability and optimization directions of different technologies in dynamic environments. The core innovations of this framework include the following. An adaptive precision adjustment method is proposed based on the flow characteristics of IoT data, ensuring stable rendering performance even under high data density and computational pressure; Key indicators such as resource utilization, system scalability, and user interaction experience are introduced, in addition to traditional metrics like rendering speed, frame rate, and response time. This evaluates 3D visualization methods more comprehensively and precisely; By analyzing data transmission efficiency across different databases in IoT environments, optimal data storage and retrieval solutions are proposed to enhance the real-time performance and stability of 3D visualization systems. Furthermore, this study compares various technological solutions through experimental data analysis and offers optimization recommendations for different application scenarios. Through these innovations, this study provides feasible technical guidance for 3D visualization in smart venues while offering a new research paradigm and application direction for optimizing 3D visualization methods in IoT environments.

2. Literature Review

Foreign cities have developed various VR 3D visualization systems, such as Google Earth, Skyline Globe Enterprise Solution, ArcGIS Pro, CityEngine, NASA Web World Wind, and Cesium. These systems, built on widely-used visualization frameworks, have been further enhanced to possess rich functionalities, including observing mountains, rivers, satellite imagery, and 3D buildings [16-21]. While systems like Google Earth are feature-rich, they have limited service coverage and lack robust mapping and spatial analysis capabilities. In contrast, Cesium has gained popularity for developing 3D visualization systems tailored for smart cities worldwide. Its open-source JavaScript library is combined with Web Graphics Library (WebGL) technology, supporting various data types such as massive terrains, vector data, and map imagery. This makes Cesium a preferred choice for 3D geospatial visualization [22,23]. For instance, Anand and Deb (2024) utilized high-resolution remote sensing data and Light Detection and Ranging (LiDAR) technology to generate more refined urban 3D models. These models facilitated enhanced visualization while enabling simulation and prediction, aiding decision-makers in better understanding potential risks and impacts on urban environments [24]. Sadowski (2024) asserted that urban informatics drove the development of smart cities, particularly in data integration and intelligent decision support. Comprehensive urban information platforms could be constructed by integrating various urban data sources, such as traffic, environmental, and socioeconomic data. These platforms leveraged advanced data analysis techniques and machine learning (ML) algorithms to monitor urban operations in real-time and conducted predictive analysis, optimizing urban services and resource allocation. Additionally, urban informatics enhanced public engagement through visualization

technologies, offering intuitive insights into urban development and fostering citizen involvement and feedback on policy-making [25]. IoT-driven visualization technologies achieved deep integration with smart city infrastructure, making urban management more intelligent and efficient. By deploying numerous sensors, cities could collect various environmental data in real-time, such as air quality, traffic flow, and public safety. Bhavsar et al. (2024) found that data integrated and analyzed through IoT platforms generated dynamic visual dashboards, helping managers access urban operational statuses. Moreover, cloud-based IoT platforms supported multi-user access and real-time updates, while data mining techniques uncovered potential issues, enhancing urban responsiveness. This IoT-based data visualization technology substantially improved urban sustainability and residents' quality of life [26]. In the context of smart city visualization, Vitanova et al. (2023) proposed the first energy map booklet for Sofia, Bulgaria. They used GIS and statistical tolerance methods to estimate building energy consumption. GIS was used for result classification and visualization [27]. Sun et al. (2023), in turn, employed a comprehensive approach to address the challenges posed by data in smart city development. They focused on data extraction and transformation, data sharing and exchange platforms, joint databases, and element searches [28].

In contrast, while the development of 3D GIS technology in China began relatively recently, it has made significant progress in recent years. The concept of "Smart Earth" was proposed by domestic International Business Machine (IBM) companies. As a result, many Chinese cities have initiated smart city projects and 3D GIS, with hundreds of cities either planning or already implementing them [29]. Domestic research primarily targeted enhancing urban management efficiency, improving tourist experiences, and optimizing power engineering planning through 3D visualization technology [30]. Qi et al. (2023) utilized traditional 3D modeling techniques alongside modern BIM systems in their research on 3D visualization. They leveraged BIM's ability to integrate rich architectural information, enabling detailed representation of building data of smart cities at a super-microscopic level [31]. Furthermore, Krašovec et al. (2024) developed an intelligent fire visualization platform based on 3D GIS to address fire safety issues in smart cities. This platform achieved fire phase recognition through real-time videos accessed through a visual interface, providing a feasible solution for intelligent firefighting and rescue operations [32].

In conclusion, developed countries have made significant progress in 3D visualization and smart city development, including creating systems like Google Earth and Cesium. These systems are widely applied in urban planning, building inspections, and energy assessments, but they still exhibit shortcomings in service coverage, spatial analysis capabilities, and support for customized applications. Meanwhile, despite domestic research's relatively late start, it has made strides in smart city development under the "Smart Earth" initiative. However, domestic applications remain largely focused on single domains (e.g., electricity, tourism, and park management), with limited exploration of cross-domain data integration and intelligent decision support. Current research presents the following deficiencies. (1) International technologies lack unified technical standards and extensive cross-domain integration capabilities for complex scenarios. (2) Domestic research primarily emphasizes efficiency improvements within specific domains but demonstrates insufficient depth in integrating multi-source data and data-driven predictive capabilities. (3) Existing literature lacks systematic analysis of key bottlenecks in 3D visualization technologies,

such as real-time processing, refined modeling, and user interaction experience. To address these gaps, this study focuses on integrating IoT technology with GIS, Cesium, and other tools to develop an efficient and intelligent 3D visualization platform for smart cities. This platform aims to enhance urban management efficiency and facilitate the intelligent transformation of scientific decision-making.

3. Research Model

The 3D visualization of smart venues requires the integration of various advanced technologies to address complex application scenarios. The smart venue 3D visualization system proposed in this study is centered on IoT, GIS, BIM, and Cesium rendering technology. It achieves real-time data collection, spatial modeling, building information integration, dynamic rendering, and interactive visualization. The following sections discuss the role of these core technologies in the system and explain their data flow and interaction mechanisms to create an optimized 3D visualization solution for smart venues.

3.1. Internet of Things

IoT refers to the network of interconnected physical devices, sensors, and other objects that communicate and share data via the Internet. This real-time data aggregation and analysis offer a valuable source of information for 3D visualization. This enables a more accurate and comprehensive representation of dynamic environments, such as cities, buildings, and facilities [33,34]. Figure 1 illustrates the application scenarios of IoT technology.

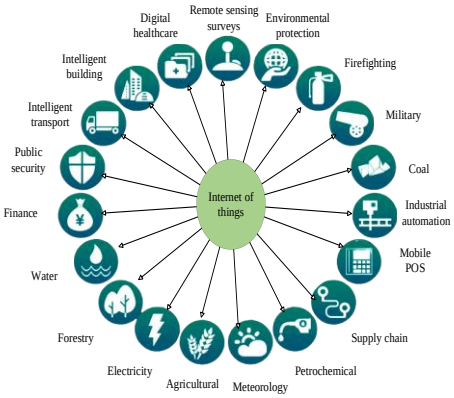


Fig. 1. Application scenarios of IoT technology

IoT serves as the data collection layer in this study, responsible for real-time collection of sensor data from within the venue and its surrounding environment, such

as temperature, humidity, personnel movement, and equipment status. This data is transmitted wirelessly via Wi-Fi, 5G, or LoRa to a cloud database and is updated in real-time in the GIS and BIM systems. This ensures that the 3D visualization system reflects the most up-to-date venue status. The data provided by IoT offers dynamic inputs for GIS to perform geographic spatial positioning. Moreover, it supplies monitoring information on the internal conditions of the building for BIM, ensuring the real-time and dynamic interaction capabilities of the 3D visualization system.

3.2. Geographic Information System

The 3D visualization of GIS combines remote sensing and global positioning systems to integrate and process geographic data, creating digital models and precise locations. Through these technologies, it is possible to simulate different scenarios and offer intuitive information for urban planning, resource management, and more, advancing the development of the digital earth [35,36]. The value of 3D GIS is revealed in Figure 2.

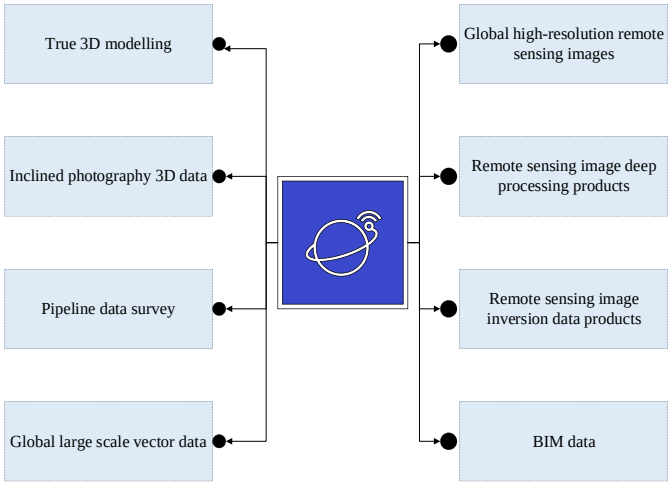


Fig. 2. Value of 3D GIS technology

In this study, GIS functions as the spatial data management layer, primarily responsible for integrating and processing the environmental data collected by IoT and providing geographic information such as terrain, buildings, and transportation. GIS constructs a 3D geographic environment using vector data (e.g., building outlines, and road networks) and raster data (e.g., satellite imagery, and topographic maps). This ensures that Cesium rendering is supported by accurate spatial background information. Additionally, GIS provides high-precision coordinate mapping capabilities, enabling seamless integration of the BIM with the actual geographic location, and ensuring the accuracy of the 3D visualization system.

3.3. Building Information Modeling

BIM integrates information from the design, construction, and operational phases of a building to enable 3D visualization, helping project teams better understand the building structure. BIM provides accurate and consistent data models, promoting design efficiency, reducing costs, minimizing errors, and supporting the building lifecycle management. During the design, construction, and maintenance phases, BIM's 3D visualization provides real-time, detailed information for the project, ensuring smooth implementation and sustainable operation [37,38]. Figure 3 depicts the building lifecycle in BIM.

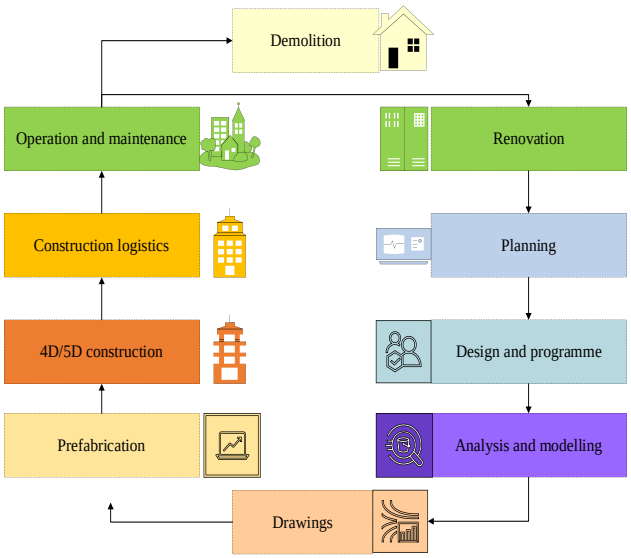


Fig. 3. BIM building life cycle

In this study, BIM functions as the building data management layer, primarily responsible for creating 3D structural models of the venue, integrating building design, construction, and operational data, and spatially aligning them with GIS data. The building models provided by BIM contain detailed structural information, such as beams, columns, walls, and pipes, and, in combination with IoT sensor data, enable real-time visualization of the building's internal status. When temperature and humidity sensors inside the venue detect anomalies, BIM can visually display the problem areas through a 3D interface and interact with IoT devices to make automatic adjustments. Moreover, BIM optimizes data through hierarchical detail techniques, allowing high-precision building models to be efficiently loaded into Cesium for rendering, thereby improving rendering efficiency and interactive smoothness.

3.4. Cesium

Cesium is an open-source JavaScript library focused on high-performance, dynamic 3D geospatial visualization. It offers powerful geospatial data rendering capabilities, supporting the loading of various types of geographic information, such as terrain, vector data, and satellite imagery, to present a realistic depiction of the Earth's surface. Based on WebGL technology, Cesium is cross-platform and allows users to experience high-performance visualization without the need for plugins. Its open JavaScript API supports custom map styles, terrain transformations, and integration with other GIS systems, making it widely used in smart cities, geographic information science, and virtual tourism, among other fields [39-41]. Figure 4 shows Cesium's system architecture and rendering mechanism.

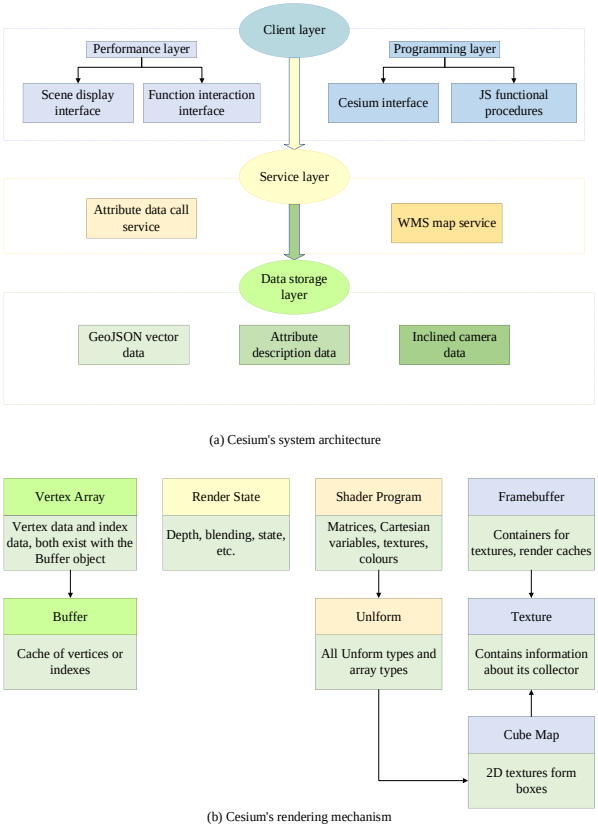


Fig. 4. Cesium's system architecture and rendering mechanism

Cesium, as the 3D visualization engine layer, integrates spatial data provided by GIS and BIM and performs efficient rendering and interactive visualization. This study utilizes WebGL technology, enabling Cesium to operate efficiently across different browser environments and achieve cross-platform 3D visualization display. The core functions of

Cesium include the following. Loading and rendering BIM and GIS data provide high-precision 3D model display; Real-time interaction allows users to freely zoom, rotate, and switch between different viewpoints to view venue information; Large-scale data optimization improves rendering efficiency and ensures a smooth visualization experience even in high data density environments. The Cesium's WebGL rendering technology also allows this study to integrate dynamic data visualization. This enables real-time 3D display of venue temperature, humidity, and personnel flow by combining IoT device data, further enhancing the system's intelligence.

This study employs Level of Detail (LOD) techniques to optimize Cesium rendering. Before importing BIM and GIS data, LOD algorithms are used to classify the models. Different precision versions of the model are applied to diverse viewing distances and display requirements. The low-LOD version employs simplified geometric models and low-detail textures, suitable for distant views. In contrast, the high-LOD version retains more complex architectural details and is appropriate for close-up or zoomed-in views. This approach allows Cesium to dynamically adjust the LOD of models based on the user's viewing distance, improving rendering efficiency and reducing computational load.

Cesium uses screen space error (SSE) to calculate the appropriate LOD for each object to display. The SSE calculation considers the viewing distance and the object's projected size on the screen. When the viewing distance is large or the object's projected area is small, the system automatically selects a lower LOD. This mechanism dynamically adjusts as the user zooms or moves the viewpoint, improving overall performance. The calculation of SSE is as follows:

$$SSE = \frac{P}{RS} \tag{1}$$

P refers to the SSE of the object; R is the distance of the object in the viewing space; S represents the screen projection size of the object.

During user interaction, Cesium dynamically calculates the required LOD for the objects within the current view and switches between models of different precision in real-time. A quadtree or octree-based data management approach is adopted to optimize rendering performance, ensuring that only the geometric details within the user's view are rendered, while areas outside the view automatically switch to a lower LOD. Moreover, through GPU Instancing technology, Cesium can reduce draw calls while maintaining visual effects, enhancing rendering efficiency in large-scale data environments.

To optimize Cesium's loading speed and reduce network load, this study uses the Zstandard compression algorithm to optimize the storage and transmission of 3D Tiles data. This allows the venue data to load quickly and achieve efficient, smooth visualization. Additionally, by using the Batched 3D Model (B3DM) format, LOD levels are managed in bulk, improving data stream control efficiency. This ensures that the system can quickly load models with matching precision from different viewpoints.

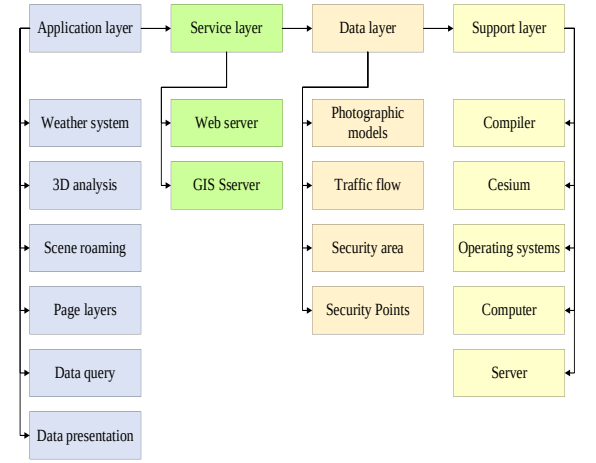
In addition to improving Cesium's 3D rendering performance, user experience is also a key focus of optimization. By combining GPU-accelerated rendering and frustum culling techniques, only objects within the current view are rendered, avoiding unnecessary computations. With WebGL-based lighting and material optimizations, even at low-LOD models, lighting effects and detail representation remain accurate, further enhancing the user's interaction experience. This optimization scheme is

particularly suitable for high data density environments and scenarios with high real-time requirements, ensuring efficient system performance in complex application settings.

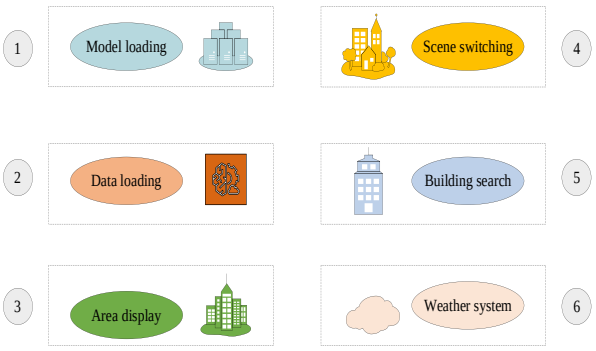
3.5. Design of Visualisation System Based on the Internet of Things and Under Cesium

The design of the 3D visualization system for smart venues adheres to several key system engineering design principles. It meets user needs while maintaining high practicality, completeness, innovation, scalability, and ease of operation. First, the system is grounded in practicality, ensuring it effectively addresses real-world requirements. Second, the principle of completeness guarantees that all required features are fully implemented, aligning with user needs and leaving no gaps. From a technological standpoint, the principle of innovation drives the adoption of advanced web-based open-source frameworks and drone-based oblique photography technology, ensuring high performance and technological advantages. Scalability is also a priority, allowing the system to accommodate future feature expansions as requirements evolve. Finally, ease of operation emphasizes a simple and user-friendly system interface to facilitate smooth interaction for users with varying levels of expertise, thus enhancing overall efficiency. Figure 5 displays the visualization system's overall architecture and functions based on IoT and Cesium.

By integrating IoT, GIS, BIM, and Cesium technologies, this study develops a 3D visualization system tailored for smart venues. This system supports real-time data acquisition, dynamic updates, and efficient rendering while having high scalability and easy operation. Its design principles emphasize practicality, comprehensiveness, advancement, and user accessibility. The system demonstrates significant potential in supporting the Smart 14th National Games Digital Twin project. By combining real-time data from IoT with spatial modeling capabilities of GIS and BIM, and leveraging Cesium for efficient rendering, the proposed system model effectively enhances smart venue management and user experience. At the same time, it provides robust technical support for smart city development.



(a) Overall visualisation system based on IoT and Cesium



(b) Functions of the visualisation system based on IoT and Cesium

Fig. 5. Overall architecture and system functions of the visualization system based on IoT and Cesium

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

The dataset used in this study primarily includes building model data and dynamic scene data. The building model data consists of geometric information, location distribution, and height of buildings within the park. The data is primarily sourced from the public geographic information databases (CityGML) and Open Street Map (OSM) datasets. CityGML provides high-precision 3D geographic information, suitable for

fine-grained building modeling, while OSM offers reliable data support for basic building contours and distribution. By integrating these two data sources, a highly realistic 3D building model with spatial analysis capabilities can be constructed, laying the foundation for the system's visualization and analysis functions. The dynamic scene data is mainly based on the Smart City Data Catalogue dataset, which integrates real-time data from over 50 cities, encompassing Barcelona, Berlin, London, Paris, and others. This dataset primarily includes urban traffic, environmental monitoring, energy consumption, citizen behavior, and demand analysis, among others. With a real-time updating mechanism, the dataset can reflect the latest state of urban operations and simulate dynamic interactions within the park, including personnel movement and vehicle trajectories. The data processing workflow involves model format conversion, storage, and the introduction of real-time dynamic data to ensure the diversity and dynamics of experimental data.

Moreover, this study leverages My Structured Query Language (MySQL), PostgreSQL, and MongoDB as databases, with the geographic information model rendered in the 3D Tiles data format. MySQL, a popular relational database management system, is favored in web applications for its compact size, fast speed, and low total cost of ownership. PostgreSQL, another prominent relational database management system, is recognized for its stability, scalability, and robust community support. It excels over MySQL in managing spatial and large-scale data, making it a suitable alternative database system. Cesium, a JavaScript-based 3D mapping platform, delivers powerful visualization capabilities for this system. Using 3D Max modeling software, the buildings and surveillance equipment within the park are rendered, lit, and arranged to simulate a realistic park environment. These models are then exported in .dae and .obj formats before being converted into Cesium-compatible formats. Furthermore, the system can load images from Tianditu, Google, and Amap, enriching the available geographic information. MongoDB is a non-relational database to handle dynamic and unstructured data, particularly in IoT and smart city contexts. Meanwhile, it can easily store large volumes of real-time data generated by sensors, and support rapid data storage, writing, and query. This effectively copes with the needs of data fluctuations and dynamic updates in these environments. Overall, MySQL, a widely used relational database, offers fast query speeds and ease of management, making it suitable for preliminary testing and routine data processing. PostgreSQL provides more robust spatial data handling capabilities and extensibility, accommodating complex query tasks and making it particularly effective for geospatial data analysis. MongoDB excels in handling unstructured data and real-time updates, meeting the needs of dynamic data streams. By comparing the performance of these three databases, this study investigates their impact on 3D rendering efficiency and offers guidance for selecting appropriate databases in different scenarios.

In practical application, the datasets and databases used in this study are highly representative and can comprehensively support the research objectives of smart venue 3D visualization and dynamic scene simulation. Specifically, selecting the CityGML and OSM datasets is based on their complementary advantages. CityGML provides high-precision 3D building geometric information, suitable for detailed modeling; OSM covers a wide range of building contours and distribution data, ensuring the authenticity and completeness of the model across large spatial areas. Moreover, the Smart City Data Catalogue integrates real-time data from over 50 cities, including traffic flow, environmental monitoring, energy consumption, and citizen behavior analysis. These

can dynamically reflect the actual state of city operations, providing a scientific basis for dynamic interactions within the park. Regarding databases, MySQL is well-suited for structured data processing due to its fast query and easy management characteristics. PostgreSQL, with its powerful spatial data processing capabilities, is more suitable for complex geographic information analysis, while MongoDB excels in handling unstructured, real-time updated data. This multi-database approach significantly improves data processing efficiency and system responsiveness compared to traditional single-database methods. All data are cross-verified with authoritative public data sources and monitored in real-time to ensure their accuracy and representativeness. Thus, it can faithfully reproduce the actual dynamics of the city and provide solid data support and a reliable dynamic simulation foundation for this study. These datasets and databases have been extensively validated in practical applications. CityGML and OSM data are widely used in various smart cities and GIS applications, with their accuracy and reliability proven in practice. The Smart City Data Catalogue dataset, through its real-time updating mechanism, reflects the real state of city operations and supports dynamic scene simulation and analysis. The selection of databases is also based on their maturity and performance in practical applications. MySQL and PostgreSQL excel in traditional data processing and spatial data analysis, while MongoDB has clear advantages in IoT and real-time data processing.

Subsequently, the data collection process follows a structured sequence of steps. First, 3D Max modeling software is employed to render, illuminate, and arrange the buildings and surveillance equipment in the park, simulating a realistic environment. Second, the rendered model data is exported in .dae and .obj formats. Third, the .dae files are converted to .glTF format using the colladaToglTF.exe tool. Fourth, the converted model data is stored in the database for system use.

Notably, the data used in this study is sourced from publicly available databases, ensuring compliance with relevant data usage and sharing policies. Since this study does not involve human participants during data collection and processing, there are no concerns regarding personal privacy or ethical risks. All data processing procedures and methods adhere to academic integrity and ethical standards, ensuring the study's transparency and reliability.

4.2. Experimental Environment

This study employs VS Code as the code editor due to its lightweight and open-source features, which allow for flexible installation of plugins to facilitate dynamic data processing. The experimental environment is configured with an Intel Core i5-9400F six-core processor, 8GB of RAM, and a 1TB 5400 RPM hard drive, running a 64-bit version of Windows 11, alongside JDK 1.8.0 and Tomcat 8.5.6. Real-time data simulation is incorporated to enhance adaptability to data fluctuations and dynamic updates, enhancing the system's flexibility. Geographic information processing is handled by ArcGIS 10.2.2 and QGIS 3.16, ensuring the capability to handle large-scale sensor network data. Cesium version 1.91 is used for 3D geographic information rendering, ensuring compatibility with mainstream browsers. The experiment also considers browser version requirements for Cesium to maintain consistent and stable rendering results. Additionally, dynamic data streams and sensor data variations are simulated to reflect real-world application scenarios, guaranteeing the experimental

results' reliability and practical significance. These optimization measures enable the study to accurately present operational conditions in smart cities, thereby increasing its practical application value. For browser selection, this study utilizes not only Chrome 85 and Firefox 80 but also includes Microsoft Edge 11 and Opera 45. Microsoft Edge, built on the Chromium kernel, offers good performance and compatibility, effectively leveraging Cesium's rendering capabilities. Meanwhile, Opera provides a convenient browsing experience with features like built-in data-saving and enhanced privacy protection features. The diverse selection of databases and browsers offers broader adaptability and a more reliable testing environment. This ensures system stability and effectiveness under various conditions, thus meeting the demands of smart city applications.

Furthermore, this study conducts a comparative analysis of various methods, including DTs, BIM, Octree, Cesium, Unity, Unreal Engine, and Three.js, based on their respective advantages and applicability in different domains. The DT method demonstrates high efficiency in data processing and analysis, making it suitable for complex datasets. BIM offers robust support for architectural and engineering visualization needs. Octree excels in 3D spatial optimization and scene management, making it ideal for large-scale data processing and rendering tasks. Cesium and Three.js exhibit leading advantages in 3D visualization and GIS, supporting efficient dynamic data rendering and interactive applications. Unity and Unreal Engine, with their powerful game engine functionalities, enable multi-layered and multidimensional interactive experiences. Integrating these methods comprehensively addresses multi-dimensional visualization requirements, ranging from architectural models to 3D scenes, providing more extensive analytical tools and visualization solutions for this study.

4.3. Parameters Setting

3D Tiles is an open standard used for storing and exchanging 3D geographic information models. It organizes data through a tiling approach, enabling efficient loading, display, and interaction of large-scale 3D models, especially suitable for 3D visualization on the Web, thus providing a fast and smooth user experience. The data conversion process involves several steps. First, it includes the preparation of raw data, such as GIS, CAD, or remote sensing data. Then, data preprocessing is performed. Next, appropriate conversion tools, such as Cesium's 3d-tiles-tools, are selected. Optimization parameters are set for the conversion process. Following this, data validation and debugging are conducted. Finally, the data is deployed to the target platform to ensure that it meets expectations and can be successfully applied in real-world scenarios.

Moreover, evaluation indicators are essential to assess the performance of different database and browser combinations in IoT and non-IoT environments for 3D visualization rendering methods. The evaluation focuses on two mainstream databases, MySQL and PostgreSQL, tested in Chrome and Firefox browsers. Key evaluation indicators encompass response time, rendering speed, and frame rate. Response time measures the duration from receiving a request to generating a response; Rendering speed indicates the system's processing speed when rendering the geographic information model; Frame rate refers to the number of image frames the system displays

per second. A comprehensive evaluation of these indicators provides valuable insights into how database and browser combinations affect 3D visualization rendering performance, aiding in system optimization.

In the experiments, consistency in versions and settings is maintained to enhance the reliability and replicability of results. Particularly, this includes using Cesium version 1.70, MySQL 8.0, and PostgreSQL 12.0 databases, along with Chrome 85 and Firefox 80 browsers. Moreover, a unified 64-bit Windows 7 operating system, Oracle JDK 1.8, Tomcat 9.0, and default transaction consistency levels (REPEATABLE READ for MySQL and READ COMMITTED for PostgreSQL) are used for databases. This ensures consistent testing conditions and reliable performance evaluations.

The data processing workflow in this study is based on the industry-standard 3D Tiles conversion guidelines. However, it incorporates innovative design in selecting and optimizing key parameters for large-scale dynamic scene applications in smart cities and IoT environments. First, during the data preprocessing, raw GIS, CAD, or remote sensing data is cleaned, coordinate systems are converted, and data is compressed to ensure compliance with the 3D Tiles specifications. During the conversion process, after repeated experiments, the size of each tile is set to 256 pixels. This parameter enables efficient LOD management and frustum culling while maintaining sufficient geometric detail, thus optimizing rendering performance. Next, a 2048 KB data loading buffer is set, which ensures the continuity of data preloading, preventing interruptions during loading, while avoiding excessive memory usage that could impact system responsiveness. For data compression, the Zstandard algorithm is selected, as it offers both a high compression ratio and fast decompression capability, significantly reducing data transmission time without sacrificing visual quality. Cesium's 3d-tiles-tools and commercial software FME are utilized, adjusting conversion parameters based on actual needs. This ensures that the final converted 3D Tiles data achieves the best balance between loading speed, rendering performance, and data accuracy, all while undergoing rigorous validation and debugging to ensure its accuracy and stability.

4.4. Performance Evaluation

Initially, the lightweight results of 3D models based on different databases and browsers are shown in Figure 6.

In Figure 6, the combination of different databases (MySQL and PostgreSQL) and browsers (Chrome and Firefox) markedly impacts the performance of 3D visualization rendering methods in IoT environments. Using MySQL and Chrome, the compressed model sizes show notable differences. Among the tested methods, Cesium achieves the best compression result, with a model size of 5612 KB. In comparison, Unity produces a model size of 6021 KB, Three.js results in 5720 KB, and Octree generates 6754 KB, all slightly less efficient than Cesium. When PostgreSQL is used with Chrome, the models' lightweight performance is relatively lower, with Cesium achieving a compressed size of 13991 KB, followed by Three.js. Also, under PostgreSQL and Firefox, Cesium performs exceptionally well, with a compressed model size of 11134 KB. MySQL combined with Firefox shows relatively balanced performance across all rendering methods, but Cesium still delivers the best results, achieving a model size of 6187 KB. In non-IoT environments, all rendering methods, including Cesium, Unity, Unreal Engine, Three.js, Octree, BIM, and DTs, exhibit a decline in performance. When

PostgreSQL is paired with Firefox, the overall compressed model sizes increase. Compared to Unity and Unreal Engine, which result in 13450 KB and 13125 KB, respectively, Cesium maintains superior performance across environments, with a size of 11705 KB. Under MySQL and Chrome, Cesium achieves a size of 6084 KB, which is 426 KB and 298 KB smaller than Unity and Unreal Engine. Similarly, under MySQL and Firefox, Cesium remains the best-performing method, followed by Three.js and Unreal Engine, with Unity slightly lagging. BIM and Octree methods exhibit minor differences in performance across the two databases, but their overall trends remain consistent. The DT method performs the worst in all scenarios. Overall, the lightweight performance of 3D models is distinctly influenced by the rendering method, database type, and browser environment. Cesium demonstrates superior compression efficiency in all test scenarios, consistently outperforming other methods in IoT and non-IoT environments.

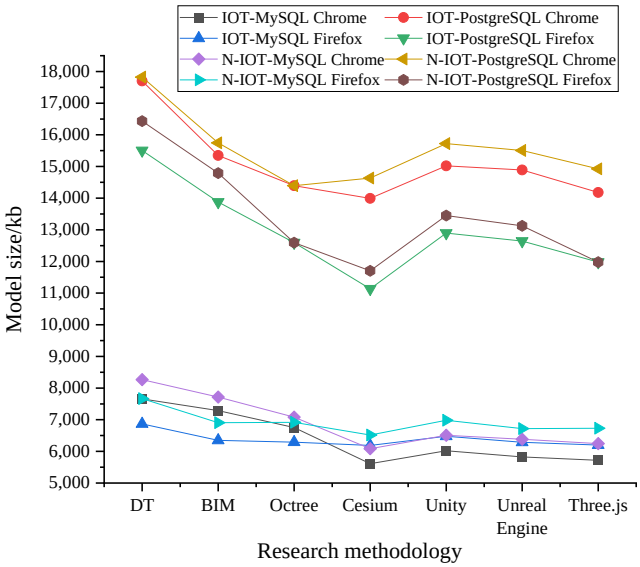
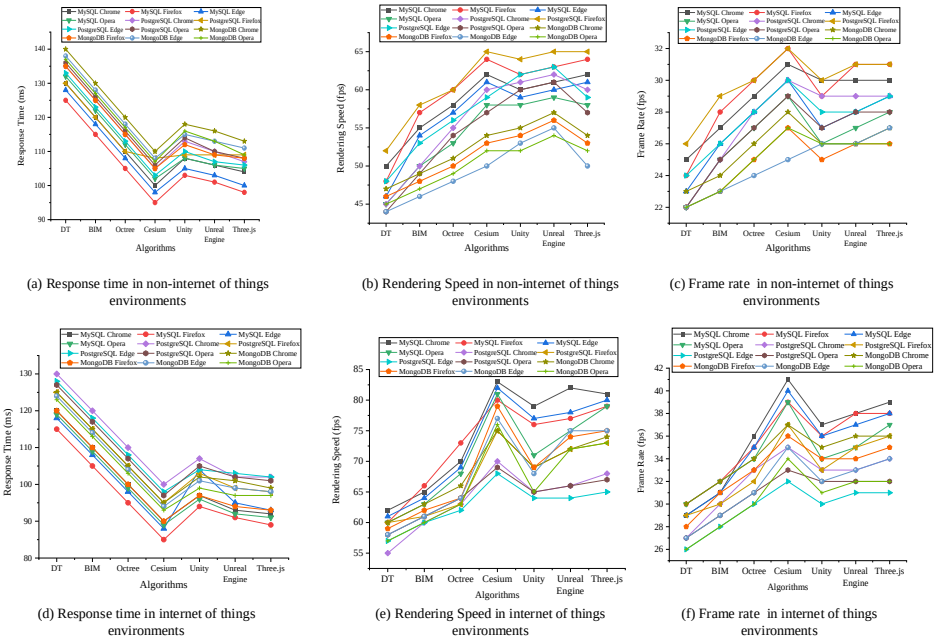


Fig. 6. Lightweight results of 3D models with different databases and browsers

Subsequently, this study analyzes the system performance using various database and browser combinations. The results are presented in Table 1 and Figure 7.



(a) Response time in non-internet of things environments

(b) Rendering Speed in non-internet of things environments

(c) Frame rate in non-internet of things environments

(d) Response time in internet of things environments

(e) Rendering Speed in internet of things environments

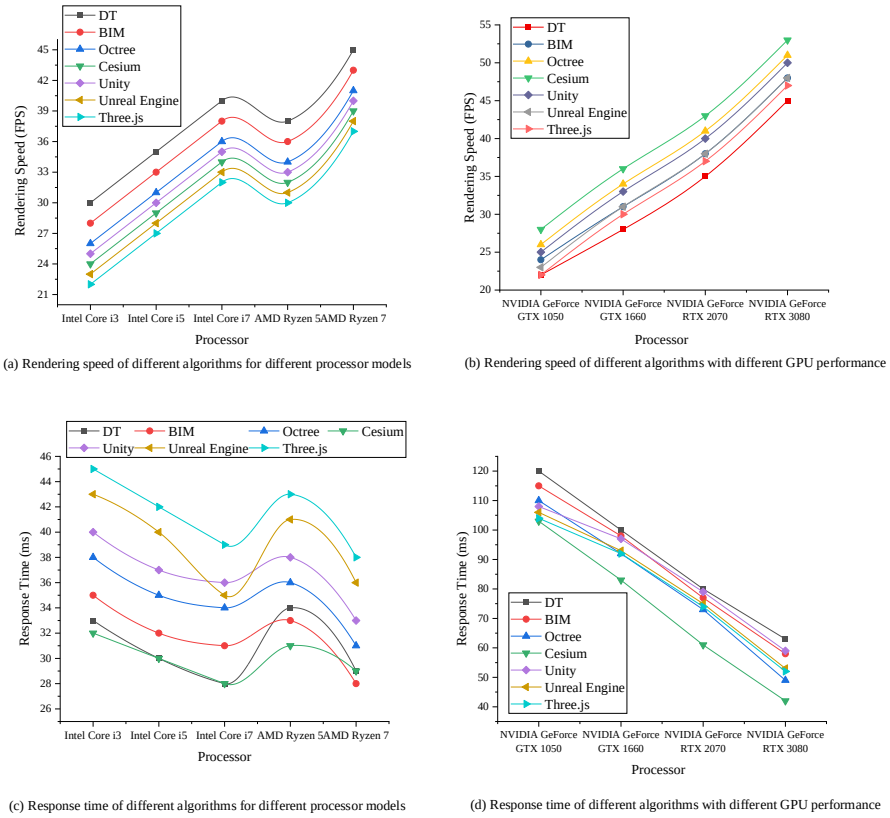
(f) Frame rate in internet of things environments

Fig. 7. Distribution of resource utilization

The data comparison in Table 1 and Figure 7 reveals that in a non-IoT environment, the response times for MySQL and PostgreSQL are fairly similar, at 130 ms and 135 ms. In contrast, MongoDB exhibits slightly higher latency, with a response time of 140 ms in the Chrome browser. However, in an IoT environment, the response times for MySQL drop to 120 ms and 110 ms in the DTs and BIM. MongoDB's response time decreases to 125 ms, demonstrating that IoT optimization significantly enhances system responsiveness. Regarding rendering speed, in the non-IoT environment, MySQL achieves a rendering speed of 50 FPS in the Chrome browser. In the IoT environment, however, MySQL and MongoDB reach impressive speeds of 83 FPS and 75 FPS, respectively, using Cesium rendering methods, demonstrating substantial improvement. As for frame rates, MongoDB's frame rate in the non-IoT environment is 23 FPS, while in the IoT environment, MySQL's frame rate increases to 41 FPS, indicating better dynamic rendering performance. Overall, the IoT environment offers remarkable advantages in response time, rendering speed, and frame rate. Based on these findings, optimization enhancement strategy should focus on several key areas. Firstly, to address MongoDB's higher response time, optimizing database indexing and data structures is recommended to reduce query latency, thus enhancing overall performance. Secondly, to improve real-time rendering capabilities in IoT environments, implementing caching mechanisms and preloading techniques can help minimize data retrieval delays and increase processing efficiency. Meanwhile, employing asynchronous rendering strategies can notably boost frame rates, especially under heavy load, ensuring a smooth user experience. Furthermore, integrating ML algorithms to analyze user behavior and dynamically adjust system resource allocation can further optimize database access

efficiency and data rendering speed. Finally, specific performance optimizations for different browser environments help adapt to the unique characteristics of each platform, achieving better cross-platform consistency and user satisfaction. Implementing these strategies can substantially enhance the system's response speed, rendering quality, and user experience.

Then, the 3D rendering effect under different hardware conditions and methods is suggested in Figure 8.



(a) Rendering speed of different algorithms for different processor models

(b) Rendering speed of different algorithms with different GPU performance

(c) Response time of different algorithms for different processor models

(d) Response time of different algorithms with different GPU performance

Fig. 8. 3D rendering effects under different hardware conditions with various methods

Table 1. Performance of the system

Environment		Non-Internet of Things Environment			Internet of Things Environment		
Database/Browser	Rendering Method	Response Time (millisecond (ms))	Rendering Speed (frames per second (FPS))	Frame Rate (FPS)	Response Time (ms)	Rendering Speed (FPS)	Frame Rate (FPS)
MySQL Chrome	DTs	130	50	25	120	62	29
	BIM	120	55	27	110	65	31
	Octree	110	58	29	100	70	36
	Cesium	100	62	31	90	83	41

	Unity	108	60	30	97	79	37
	Unreal Engine	106	61	30	93	82	38
	Three.js	104	62	30	92	81	39
	DTs	125	48	24	115	60	30
	BIM	115	57	28	105	66	32
	Octree	105	60	30	95	73	35
MySQL Firefox	Cesium	95	64	32	85	80	39
	Unity	103	62	29	94	76	36
	Unreal Engine	101	63	31	91	77	38
	Three.js	98	64	31	89	79	38
	DTs	128	46	23	118	61	29
	BIM	118	54	26	108	64	31
	Octree	108	57	28	98	69	35
MySQL Edge	Cesium	98	61	30	88	82	40
	Unity	105	59	27	104	77	36
	Unreal Engine	103	60	28	95	78	37
	Three.js	100	61	29	93	80	38
	DTs	132	45	22	119	60	30
	BIM	122	50	25	109	63	32
	Octree	112	53	27	99	68	34
MySQL Opera	Cesium	102	61	29	89	81	39
	Unity	108	58	26	96	71	34
	Unreal Engine	106	59	27	92	75	35
	Three.js	105	58	28	91	79	37
	DTs	135	45	22	130	55	27
	BIM	125	50	25	120	60	30
	Octree	115	55	28	110	63	33
PostgreSQL Chrome	Cesium	105	64	30	100	70	35
	Unity	113	61	29	107	65	33
	Unreal Engine	110	62	29	102	66	33
	Three.js	107	60	29	102	68	34
	DTs	130	52	26	125	60	29
	BIM	120	58	29	115	61	30
	Octree	110	60	30	105	63	32
PostgreSQL Firefox	Cesium	108	67	32	95	75	37
	Unity	109	64	30	103	69	33
	Unreal Engine	109	65	31	99	72	35
	Three.js	109	65	31	98	73	36
	DTs	133	48	24	128	57	26
	BIM	123	53	26	118	60	28
	Octree	113	56	28	108	62	30
PostgreSQL Edge	Cesium	103	65	30	98	68	32
	Unity	110	62	28	104	64	30
	Unreal Engine	107	63	28	103	64	31
	Three.js	106	59	29	102	65	31
	DTs	136	44	22	127	58	27
	BIM	126	49	25	117	61	29
	Octree	116	54	27	107	64	31
PostgreSQL Opera	Cesium	106	63	29	97	69	33
	Unity	114	60	27	105	65	32
	Unreal Engine	110	61	28	102	66	32
	Three.js	108	57	28	101	67	32
	DTs	140	47	23	125	60	30
	BIM	130	49	24	115	63	32
	Octree	120	51	26	105	66	34
MongoDB Chrome	Cesium	110	58	28	95	75	37
	Unity	118	55	26	102	69	35
	Unreal Engine	116	57	26	101	72	36
	Three.js	113	54	27	99	74	36
	DTs	135	46	22	120	59	28
	BIM	125	48	23	110	62	31
	Octree	115	50	25	100	64	33
MongoDB Firefox	Cesium	105	57	27	90	79	36
	Unity	112	54	25	97	69	34
	Unreal Engine	109	56	26	94	74	34
	Three.js	108	53	26	93	75	35
	DTs	138	44	22	124	58	27
	BIM	128	46	23	114	61	29
	Octree	118	48	24	104	64	31
MongoDB Edge	Cesium	108	56	25	94	77	35
	Unity	115	53	26	101	68	32
	Unreal Engine	113	55	26	99	75	33
	Three.js	111	50	27	98	75	34
	DTs	137	45	22	123	57	26
	BIM	127	47	23	113	60	28
	Octree	117	49	25	103	63	30
MongoDB Opera	Cesium	107	55	27	93	76	34
	Unity	116	52	26	99	65	31
	Unreal Engine	113	54	26	97	72	32
	Three.js	109	52	26	97	73	32

In Figure 8, the rendering speed of various methods exhibits an overall upward trend as processors transition from Core i3 to Core i5 and then to Core i7. For the Cesium method, the rendering speed increases from 24 FPS on Core i3 to 29 FPS on Core i5 and further to 34 FPS on Core i7. This indicates that each method's 3D rendering performance improves substantially with enhanced Intel processor performance. Similarly, from Ryzen 5 to Ryzen 7, the rendering speed of all methods shows an increasing trend. For example, the BIM method improves from 36 FPS on Ryzen 5 to 43 FPS on Ryzen 7, while Cesium increases to 39 FPS on Ryzen 7. This demonstrates the positive impact of AMD processor performance on 3D rendering. When the Graphics Processing Unit (GPU) upgrades from NVIDIA GeForce GTX 1050 to GTX 1660, RTX 2070, and RTX 3080, the rendering speed of all methods shows a marked increase. In the DT method, the rendering speed rises from 22 FPS on GTX 1050 to 28 FPS on GTX 1660, 35 FPS on RTX 2070, and 45 FPS on RTX 3080. This indicates that GPU performance remarkably enhances the rendering speed across all methods. However, the sensitivity to GPU upgrades varies among methods. The Octree method experiences a relatively large improvement, with rendering speed increasing from 26 FPS on GTX 1050 to 51 FPS on RTX 3080. In contrast, the Three.js method exhibits a smaller increase, from 22 FPS on GTX 1050 to 47 FPS on RTX 3080. This suggests that when selecting a GPU, scenarios emphasizing specific methods' rendering performance should consider the method's sensitivity to GPU performance improvements. From Core i3 to Core i5 and Core i7, the response time of all algorithms generally decreases. For the DT algorithm, the response time improves from 33 ms on Core i3 to 30 ms on Core i5 and 28 ms on Core i7. This demonstrates that enhanced Intel processor performance strengthens the real-time rendering capabilities of the algorithms, providing more immediate feedback. Similarly, from Ryzen 5 to 7, the response time of all algorithms decreases. For example, the BIM algorithm improves from 33 ms on Ryzen 5 to 28 ms on Ryzen 7, indicating that better AMD processor performance enhances 3D rendering real-time responsiveness. As the GPU upgrades from NVIDIA GeForce GTX 1050 to GTX 1660, RTX 2070, and RTX 3080, the response time of all algorithms decreases significantly. The DTs algorithm's response time reduces from 120 ms on GTX 1050 to 100 ms on GTX 1660, 80 ms on RTX 2070, and 63 ms on RTX 3080. This illustrates that GPU performance enhancements improve the immediate feedback capabilities of all algorithms. Sensitivity to GPU performance upgrades varies among algorithms. The Octree algorithm exhibits a relatively large decrease in response time, from 110 ms on GTX 1050 to 50 ms on RTX 3080. In contrast, the Three.js algorithm shows a smaller reduction, from 104 ms on GTX 1050 to 58 ms on RTX 3080.

Next, the comparative results of functional tests of the smart venue visualization system based on Cesium rendering technology are outlined in Table 2.

Table 2 highlights Cesium's superiority over other technologies across several key performance indicators. For instance, in terms of 3D model loading, Cesium's average loading time is only 0.5 seconds; It substantially outperforms BIM and DTs, which typically exceed 1 second due to data complexity. Regarding security data queries, Cesium demonstrates an average query time of 0.2 seconds, compared to 0.4 to 0.6 seconds for BIM and Octree, indicating a notable improvement in responsiveness. Furthermore, the frame rate during scene transitions in Cesium remains stable at 60 FPS, ensuring a smooth user experience, well above Octree's 45 FPS. The accuracy of building information queries in Cesium reaches 98%, surpassing other models'

approximate 93%, demonstrating its advantage in information accuracy. Finally, when implementing heatmap functionality, Cesium achieves rendering clarity of over 90%, compared to 80% and 75% for traditional DTs and BIM models, showcasing its powerful capabilities in dynamic data visualization. These comparative results strongly support Cesium's potential for practical application in smart city scenarios, providing a more efficient, accurate, and intuitive visualization solution.

Table 2. System function test results

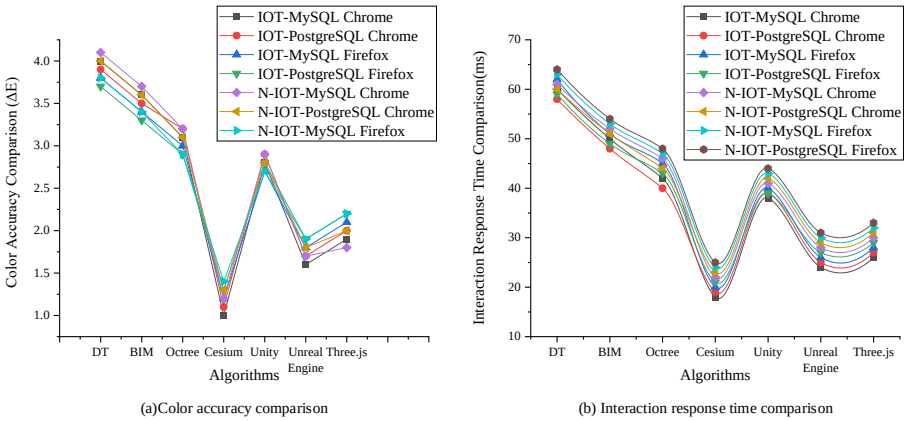
Function Module	Testing Content	Testing Result	Comparison Methods
3D Model Loading	Completeness of key location modeling, loading speed	Average loading time: 0.5 seconds, Pass	DTs: 1.2s, BIM: 0.8s, Octree: 0.6s, Cesium: 0.5s
Security Data Query	Ability to query security point information, data response time	Average query time: 0.2 seconds, Pass	DTs: 0.5s, BIM: 0.3s, Octree: 0.4s, Cesium: 0.2s
Scene Transition Smoothness	Fluidity and accuracy of scene transitions	Frame rate maintained at 60 FPS, Pass	DTs: 45 FPS, BIM: 55 FPS, Octree: 58 FPS, Cesium: 60 FPS
Building Information Query	Ability to click and query venue information, accuracy of information display	Accuracy rate: 98%, Pass	DTs: 90%, BIM: 95%, Octree: 97%, Cesium: 98%
Scene Navigation Performance	Presence of stuttering during route navigation, completeness of indoor navigation functionality	Stuttering rate: below 5%, Pass	DTs: 10%, BIM: 6%, Octree: 4%, Cesium: <5%
Heatmap Functionality	Implementation of heatmap feature, clarity and accuracy of rendering results	Clarity rate: over 90%, Pass	DTs: 85%, BIM: 88%, Octree: 89%, Cesium: 90%
Measurement Tool Effectiveness	Ground-level measurement capability, presence of point-line misalignment	Misalignment occurrence rate: below 2%, Pass	DTs: 3%, BIM: 2.5%, Octree: 1.5%, Cesium: <2%

Table 3. Comparison of resolution of rendered images (in pixels)

Processor	DTs	BIM	Octree	Cesium	Unity	Unreal Engine	Three.js
IOT-MySQL Chrome	1600x900	1920x1080	2560x1440	4096x2160	2560x1440	3200x1800	3000x1600
IOT-PostgreSQL Chrome	1680x1050	2048x1080	2600x1440	4096x2160	2560x1600	3400x1800	3100x1650
IOT-MySQL Firefox	1550x900	1900x1080	2520x1400	3840x2160	2560x1440	3200x1800	3000x1500
IOT-PostgreSQL Firefox	1600x900	1980x1080	2540x1440	3840x2160	2600x1500	3300x1800	3050x1550
N-IOT-MySQL Chrome	1500x850	1850x1050	2400x1350	4000x2100	2550x1430	3150x1700	2900x1450
N-IOT-PostgreSQL Chrome	1520x860	1900x1070	2420x1370	4000x2100	2580x1460	3180x1720	2950x1480
N-IOT-MySQL Firefox	1490x830	1830x1040	2380x1330	3750x2080	2500x1400	3120x1680	2850x1420
N-IOT-PostgreSQL Firefox	1510x840	1880x1060	2410x1360	3750x2080	2530x1420	3140x1700	2880x1440

Moreover, at the user level, this study selects the resolution, color accuracy, and interactive response time of the rendered image as indicators to compare the user

experience under different rendering methods. The specific results are listed in Table 3 and Figure 9.



(a)Color accuracy comparison (b) Interaction response time comparison

Fig. 9. Comparison results of user experience in different environments

Figure 9 demonstrates that, when comparing different rendering methods, Cesium leads in terms of rendered image resolution, achieving an ultra-high definition resolution of 4096x2160 pixels. This performance is particularly outstanding in the IOT-MySQL Chrome and IOT-PostgreSQL Chrome environments. Unreal Engine and Three.js follow closely, with resolutions of 3200x1800 pixels and 3000x1600 pixels, respectively, outperforming Unity's 2560x1440 pixels and other methods. In terms of color accuracy, Cesium again performs best, maintaining ΔE values between 1.0 and 1.5 under various conditions, which indicates exceptional color reproduction capabilities. Unreal Engine and Three.js exhibit ΔE values between 1.6 to 2.0 and 1.8 to 2.3, respectively. Although slightly inferior to Cesium, they still demonstrate good color accuracy. In contrast, Unity shows weaker color precision, with ΔE values generally ranging from 2.7 to 2.9, while other methods fall between 3.0 and 4.1, reflecting a notable disadvantage. Regarding interactive response time, Cesium achieves an average response time of 18 ms to 25 ms, delivering the smoothest interaction experience even under high-resolution and high-precision rendering. Unreal Engine and Three.js follow, with response times of 24 ms to 31 ms and 26 ms to 33 ms, offering good real-time performance. In comparison, Unity's response time ranges from 38 ms to 44 ms; Other methods such as BIM and Octree exhibit response times between 40 ms and 60 ms, which may lead to perceptible delays in highly interactive scenarios. Overall, Cesium leads comprehensively in resolution, color accuracy, and interactive response time, providing the best visual and interactive experience. Unreal Engine and Three.js are close behind with excellent color and response speed, making them suitable for applications requiring high fluency and visual effects.

In addition, the descriptive statistics of the overall performance of each rendering method are detailed in Table 4.

In Table 4, the overall performance of different rendering methods reveals significant differences in average rendering speed. Cesium outperforms other methods, achieving an average rendering speed of 52.89 FPS, with a maximum and minimum of 70 FPS and 40 FPS. In contrast, Three.js shows the lowest performance, with an average, maximum, and minimum rendering speed of 34.56 FPS, 46 FPS, and 24 FPS. These

results indicate that Cesium delivers more stable and efficient rendering, whereas Three.js faces greater performance limitations. Other methods, such as BIM, Octree, and Unreal Engine, also demonstrate solid performance, with average speeds of 40.12 FPS, 43.78 FPS, and 42.35 FPS, respectively. These differences highlight the unique strengths and challenges in performance optimization across different rendering methods.

Table 4. Descriptive statistics of overall performance for rendering methods

Rendering Method	Mean Performance (FPS)	Standard Deviation (SD)	Minimum (FPS)	Maximum (FPS)
DTs	78	5.6	72	85
BIM	84	6.2	76	92
Octree	82	5.8	74	89
Cesium	92	4.3	88	97
Unity	88	5	82	94
Unreal Engine	90	4.8	86	96
Three.js	87	5.2	81	93

Finally, this study is based on the Smart City Data Catalogue dataset. It randomly selects 10 cities to compare and evaluate the performance of various rendering methods in the urban traffic domains, environmental monitoring, energy consumption, citizen behavior, and demand analysis. Specific results are illustrated in Figure 10.

In Figure 10, significant differences are observed in the performance of various rendering methods in urban traffic scenarios. Among these, Cesium demonstrates the best performance, achieving an overall average of 58.69 FPS, indicating high precision and interactive responsiveness. Octree and Three.js follow with averages of 47.67 FPS and 53.96 FPS, respectively, providing basic traffic data visualization but falling short in efficiency and user experience. In environmental monitoring, Cesium again excels, with an average of 61.57 FPS, showcasing clear advantages in rendering environmental data and dynamic representation. Octree, Unreal Engine, and Three.js exhibit slightly weaker performance, especially when processing complex environmental data. For energy consumption visualization, notable differences exist among the methods. Cesium achieves an average of 63.49 FPS, demonstrating strong performance in large-scale energy data rendering. Three.js ranks second with an average rendering speed of 57.28 FPS. In citizen behavior and demand analysis, Cesium maintains its lead with an overall average of 61.47 FPS, providing efficient interactions and accurate rendering. Unreal Engine, Octree, and BIM deliver relatively similar performance but fail to reach Cesium's level. To sum up, Cesium demonstrates remarkable advantages across all indicators, while DTs exhibit weaker performance in several scenarios.

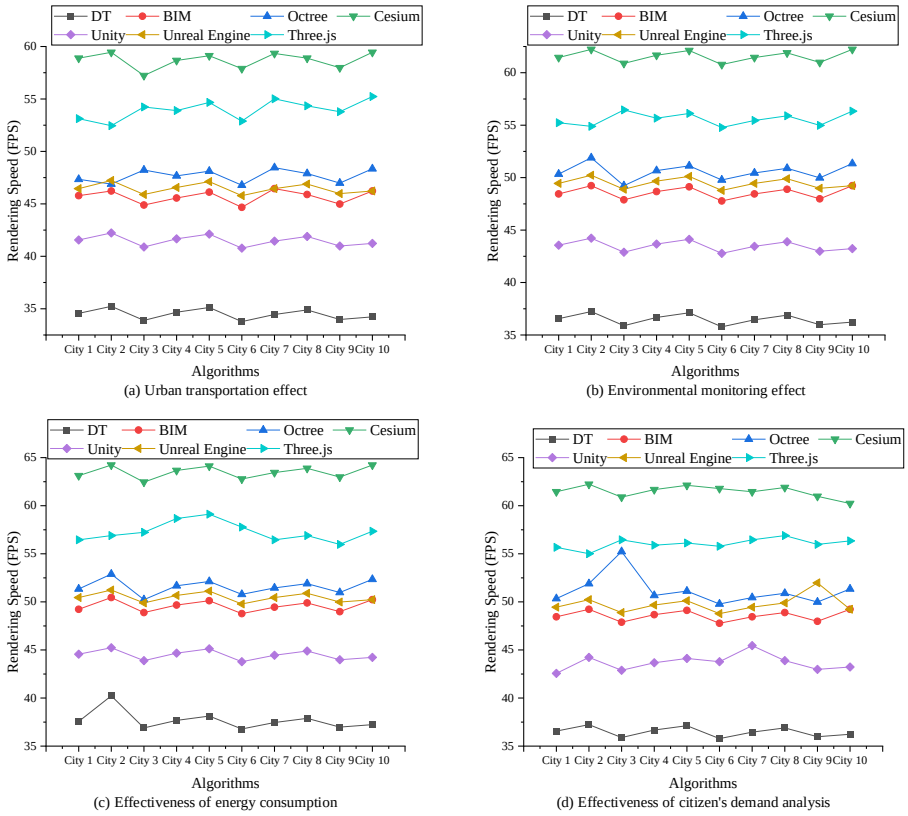


Fig. 10. Rendering of each rendering method in different domains

This study conducts 10 rounds of independent experiments on each method under the same environment to avoid the bias of a single value. Its average compression size and SD are calculated, and the results are denoted in Table 5. Under the same hardware environment (Intel i7-12700K + RTX 3070 + 32GB RAM), the average frame rate and response time of different algorithms are listed in Table 6.

Table 5. 3D model compression sizes (KB, Mean ± SD)

Rendering Method	MySQL + Chrome	MySQL + Firefox	PostgreSQL + Chrome	PostgreSQL + Firefox
DT	7660±45	17702±62	6869±50	15504±58
BIM	7288±38	15348±57	6348±42	13877±52
Octree	6754±35	14392±50	6293±37	12595±45
Cesium	5612±28	13991±46	6187±32	11134±42
Unity	6021±30	15021±48	6478±34	12896±40
UnrealEngine	5824±29	14890±47	6287±33	12643±38
Three.js	5720 ± 28	14178 ± 45	6198 ± 31	11980 ± 37

Tables 5 and 6 show that different rendering methods exhibit varying performance in terms of compression size, rendering speed, and response time. Regarding compression size, the DT method achieves the highest compression, particularly in MySQL and PostgreSQL environments, while Cesium and Three.js offer smaller and more stable

compression sizes. In terms of rendering speed, Cesium performs the best in non-IoT environments; However, in IoT environments, its frame rate significantly drops, demonstrating the impact of IoT on rendering performance. The DT and BIM methods have the slowest rendering speeds in IoT environments; Unity and Unreal Engine remain relatively stable in frame rate and response time, showing better performance. Overall, selecting the appropriate rendering method requires a comprehensive consideration of compression effectiveness, rendering speed, and response time, with optimizing rendering performance in IoT environments being particularly important.

Table 6. Rendering speed and response time (Mean ± SD)

Rendering Method	Non-IoT (FPS)	IoT (FPS)	Non-IoT Response Time (ms)	IoT Response Time (ms)
DT	72.0±5.4	40±5.1	32.0±2.5	28.0±2.3
BIM	76.0±5.8	38±5.2	35.0±3.3	31.0±3.1
Octree	74.0±5.1	36±4.6	38.0±3.0	34.0±2.9
Cesium	88.0±4.0	34±2.7	30.0±2.2	28.0±2.1
Unity	82.0±5.5	35±3.9	40.0±3.4	36.0±3.2
UnrealEngine	86.0±4.4	33±2.9	39.0±3.2	35.0±3.1
Three.js	81.0±5.3	32±4.1	42.0±3.6	39.0±3.5

4.5. Discussion

With the rapid development of digital technologies, this study takes a significant step in evaluating the performance of 3D visualization rendering methods, presenting distinct features compared to previous research. Earlier studies mainly focused on theoretical discussions with limited attention to the performance of specific databases and browser combinations. This study specifically highlights the practical application of the Cesium rendering method across multiple environments, providing valuable empirical references for users. The study enhances the result's generalizability by comparing two major databases, MySQL and PostgreSQL. Meanwhile, it provides practical guidance for users in selecting the appropriate database for real-world applications. Additionally, the study underscores Cesium's advantages in rendering performance through cross-browser performance comparisons. From the system performance analysis, Cesium demonstrates superior rendering performance, response time, and frame rate in different environments, ensuring a smooth user experience. As an open-source JavaScript library, Cesium leverages WebGL technology to deliver high-performance, cross-platform 3D visualization, remarkably improving rendering speed and response time, especially when handling large-scale geospatial data, showcasing its advantages. In the experiment, PostgreSQL exhibits high CPU and memory usage when processing complex queries and large datasets, reflecting its potential in high-performance rendering. At the same time, Cesium's efficient real-time rendering meets the dynamic display and rapid response requirements in smart venue management, providing strong support for emergency response and precise urban planning analysis. Under the experimental setup with an Intel Core i5 processor and 8GB of memory, the system ensures efficient rendering speed and response time, further validating the research results' effectiveness. The findings offer valuable references for optimizing workflows, decision-making processes, and technical designs in smart venue management. Smart venue management systems often face complex data processing and real-time response

challenges. By leveraging Cesium's efficient rendering and data processing capabilities, management workflows can be optimized, particularly in real-time monitoring and spatial resource optimization, thereby improving decision-making efficiency. Cesium's real-time rendering capability also supports emergency response decisions by providing dynamic displays of environmental changes, helping managers make timely and accurate decisions. Moreover, the comparative analysis with alternatives such as BIM, digital twins, and octrees reveals Cesium's advantages in speed and efficiency. Compared to technologies like BIM and digital twins, Cesium's unique rendering engine and WebGL-based cross-platform support provide distinct advantages in large-scale data rendering and real-time interaction. Particularly, Cesium can effectively process complex geospatial data and render it in real-time, meeting the smart venue management's need for efficient and accurate rendering. In comparison to traditional BIM or digital twin technologies, Cesium's optimization algorithms, such as efficient spatial data management and rendering mechanisms, provide faster response times and higher rendering quality. Cesium's open-source nature gives it greater flexibility and customizability in practical applications, making it better suited to meet the requirements of different scenarios. In conclusion, this study provides important technical references for the digitalization and visualization of smart venues through systematic comparative analysis. This further validates Cesium's rendering performance and responsiveness in different environments, offering theoretical foundations and practical guidance for applications in related fields.

5. Conclusion

5.1. Research Contribution

This study draws several key conclusions by analyzing the 3D visualization performance of diverse rendering methods across database types and browser environments. First, Cesium showcases exceptional model compression capabilities in IoT environments, delivering optimal lightweight performance and rendering speed when using both MySQL and PostgreSQL databases. Compared with other methods, Cesium demonstrates substantial advantages across various hardware configurations and GPU performance enhancements, from Core i3 to Core i7 processors and from GTX 1050 to RTX 3080 GPUs. Its rendering speed and response times consistently lead in all conditions. Second, IoT environment optimizations markedly improve system response times and rendering speeds. Cesium performs particularly well under MySQL and MongoDB databases. Furthermore, under different hardware conditions, Cesium outperforms other rendering methods in resolution, color accuracy, and interactive response time, offering users a smoother and more efficient interactive experience. Across diverse datasets and application scenarios, Cesium demonstrates high accuracy and dynamic data rendering capabilities, highlighting its extensive potential for smart city applications. Overall, Cesium exhibits superior performance and stability under various conditions, establishing itself as the most competitive rendering method in the current 3D visualization domain. This study comprehensively compares the analysis of 3D visualization methods in IoT

environments, offering scientific insights for selecting and optimizing relevant technologies. Meanwhile, the study provides valuable insights by examining the strengths and weaknesses of these methods in performance, data security, and user experience. These insights contribute to developing more efficient, stable, and secure solutions for applications such as smart campuses, thereby advancing technology and its practical applications in related fields.

5.2. Future Works and Research Limitations

Despite the contributions of this study, several limitations remain. The study does not fully account for the impact of different network conditions on system performance, such as scenarios with low bandwidth or unstable networks. Additionally, it does not delve into data security and privacy protection. Further research is needed to explore mechanisms for ensuring the security of sensitive information during data transmission and storage. Future studies could expand the environmental factors under investigation and adopt multi-layered analytical approaches to explore data security and privacy protection mechanisms. Integrating ML algorithms for data modeling could help evaluate system security and privacy protection capabilities across various scenarios, thus enhancing the functionality of smart campuses and ensuring data security.

References

1. Shariatpour, F., Behzadfar, M., Zareei, F.: Urban 3D Modeling as a Precursor of City Information Modeling and Digital Twin for Smart City Era: A Case Study of the Narmak Neighborhood of Tehran City, Iran. Journal of Urban Planning and Development, Vol. 150, No. 2, 04024005. (2024)
2. Li, W., Zhu, J., Pirasteh, S., Zhu, Q., Guo, Y., Luo, L., Dehbi, Y.: A 3D Virtual Geographic Environment for Flood Representation Towards Risk Communication. International Journal of Applied Earth Observation and Geoinformation, Vol. 128, No. 2, 103757. (2024)
3 Lam, P. D., Gu, B. H., Lam, H. K., Ok, S. Y., Lee, S. H.: Digital Twin Smart City: Integrating IFC and CityGML with Semantic Graph for Advanced 3D City Model Visualization. Sensors, Vol. 24, No. 12, 3761. (2024)
4 Wang, X., Jiang, L., Wang, F., You, H., Xiang, Y.: Disparity Refinement for Stereo Matching of High-Resolution Remote Sensing Images Based on GIS Data. Remote Sensing, Vol. 16, No. 3, 487. (2024)
5 Maky, A. M., AlHamaydeh, M., Saleh, M.: GIS-Based Regional Seismic Risk Assessment for Dubai, UAE, Using NHERI SimCenter R2D Application. Buildings, Vol. 14, No. 5, 1277. (2024)
6 Muravskyi, V., Kundeus, O., Hrytsyshyn, A., Lutsiv, R.: Accounting in a Smart City with the Combined Use of the Internet of Things and Geographic Information Systems. Herald of Economics, Vol. 23, No. 2, 41-57. (2023)
7 Liu, B., Wu, C., Xu, W., Shen, Y., Tang, F.: Emerging Trends in GIS Application on Cultural Heritage Conservation: A Review. Heritage Science, Vol. 12, No. 1, 139. (2024)
8 Janovský, M.: Pre-Dam Vltava River Valley—A Case Study of 3D Visualization of Large-Scale GIS Datasets in Unreal Engine. ISPRS International Journal of Geo-Information, Vol. 13, No. 10, 344. (2024)
9 Spreafico, A., Chiabrando, F.: 3D WebGIS for Ephemeral Architecture Documentation and Studies in the Humanities. Heritage, Vol. 7, No. 2, 913-947. (2024)

10 Liu, Z., Li, T., Ren, T., Chen, D., Li, W., Qiu, W.: Day-to-Night Street View Image Generation for 24-Hour Urban Scene Auditing Using Generative AI. Journal of Imaging, Vol. 10, No. 5, 112. (2024)
11 Wang, L., Wang, Y., Huang, W., Han, J.: Analysis Methods for Landscapes and Features of Traditional Villages Based on Digital Technology—The Example of Puping Village in Zhangzhou. Land, Vol. 13, No. 9, 1539. (2024)
12 Grêt-Regamey, A., Fagerholm, N.: Key Factors to Enhance Efficacy of 3D Digital Environments for Transformative Landscape and Urban Planning. Landscape and Urban Planning, Vol. 244, No. 1, 104978. (2024)
13 Lei, B., Liang, X., Biljecki, F.: Integrating Human Perception in 3D City Models and Urban Digital Twins. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 10, No. 1, 211-218. (2024)
14 Yu, Q., Feng, D., Li, G., Chen, Q., Zhang, H.: AdvMOB: Interactive Visual Analytic System of Billboard Advertising Exposure Analysis Based on Urban Digital Twin Technique. Advanced Engineering Informatics, Vol. 62, No. 1, 102829. (2024)
15 Li, X., Wang, C., Kassem, M. A., Ali, K. N.: Emergency Evacuation of Urban Underground Commercial Street Based on BIM Approach. Ain Shams Engineering Journal, Vol. 15, No. 4, 102633. (2024)
16 Bianconi, F., Filippucci, M., Cornacchini, F., Migliosi, A.: The Impact of Google's APIs on Landscape Virtual Representation. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 48, No. 1, 91-98. (2024)
17 Usta, Z., Cömert, Ç., Akın, A. T.: An Interoperable Web-Based Application for 3D City Modelling and Analysis. Earth Science Informatics, Vol. 17, No. 1, 163-179. (2024)
18 Kamaruzaman, E. H., La Croix, A. D., Kamp, P. J.: Dataset of 3D Computer Models of Late Miocene Mount Messenger Formation Outcrops in New Zealand, Built with UAV Drones. Data in Brief, Vol. 52, No. 1, 110035. (2024)
19 Grădinara, A. P., Badea, A. C., Dragomir, P. I.: Using VR to Explore the 3D City Model Obtained from LiDAR Data. Revista Română de Inginerie Civilă, Vol. 15, No. 1, 1-10. (2024)
20 Schinder, A. M., Young, S. R., Steward, B. J., Dexter, M., Kondrath, A., Hinton, S., Davila, R.: Deterministic Global 3D Fractal Cloud Model for Synthetic Scene Generation. Remote Sensing, Vol. 16, No. 9, 1622. (2024)
21 Pansini, R., Guzel, S., Morelli, G., Barsuglia, F., Penno, G., Catanzariti, G., Campana, S.: Multi-Modal/Multi-Resolution 3D Data Acquisition and Processing for a New Understanding of the Historical City of Siena (Italy). International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 48, No. 1, 341-347. (2024)
22 Maguelva, N. M., Mustapha, H., Hubert, F.: Towards a 3D Web Tool for Visualization and Simulation of Urban Flooding: The Case of Metropolitan Cities in Cameroon. International Journal of Advanced Studies in Engineering and Research (IJASER), Vol. 4, No. 4, 25-40. (2023)
23 Leopold, U., Braun, C., Pinheiro, P.: An Interoperable Digital Twin to Simulate Spatio-Temporal Photovoltaic Power Output and Grid Congestion at Neighbourhood and City Levels in Luxembourg. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 48, No. 1, 95-100. (2023)
24 Anand, A., Deb, C.: The Potential of Remote Sensing and GIS in Urban Building Energy Modelling. Energy and Built Environment, Vol. 5, No. 6, 957-969. (2024)
25 Sadowski, J.: Anyway, the Dashboard Is Dead': On Trying to Build Urban Informatics. New Media & Society, Vol. 26, No. 1, 313-328. (2024)
26 Bhavsar, S., Bajare, A., Jadhav, V., Marathe, G., Nikam, A.: A Survey on Real-Time Market Dynamics Through Visual Dashboards. International Journal of Engineering and Management Research, Vol. 14, No. 1, 52-57. (2024)

27 Vitanova, L. L., Petrova-Antonova, D., Hristov, P. O., Shirinyan, E.: Towards Energy Atlas of Sofia City in Bulgaria. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 48, 123-129. (2023)
28 Sun, K., Liu, N., Sun, X., Zhang, Y.: Design and Implementation of Big Data Analysis and Visualization Platform for the Smart City. International Journal of Information Technology Management, Vol. 22, No. 3-4, 373-385. (2023)
29 Liu, Y., Wu, Y., Cao, H., Wang, Z., Wang, Z., Cui, Y., Li, G.: The Application of GIS Technology in the Construction of Smart City. Academic Journal of Science and Technology, Vol. 5, No. 2, 183-186. (2023)
30 Chang, Y., Xu, J.: Application of Spatial Data and 3S Robotic Technology in Digital City Planning. International Journal of Intelligent Networks, Vol. 4, 211-217. (2023)
31 Qi, C., Zhou, H., Yuan, L., Li, P., Qi, Y.: Application of BIM+GIS Technology in Smart City 3D Design System. International Conference on Cyber Security and Intelligent Analysis, Vol. 3, No. 30, 37-45. (2023)
32 Krašovec, A., Baldini, G., Pejović, V.: Multimodal Data for Behavioural Authentication in Internet of Things Environments. Data in Brief, Vol. 55, No. 1, 110697. (2024)
33 Peter, O., Pradhan, A., Mbohwa, C.: Industrial Internet of Things (IIoT): Opportunities, Challenges, and Requirements in Manufacturing Businesses in Emerging Economies. Procedia Computer Science, Vol. 217, No. 1, 856-865. (2023)
34 Sasikumar, A., Vairavasundaram, S., Kotecha, K., Indragandhi, V., Ravi, L., Selvachandran, G., Abraham, A.: Blockchain-Based Trust Mechanism for Digital Twin Empowered Industrial Internet of Things. Future Generation Computer Systems, Vol. 141, No. 1, 16-27. (2023)
35 Raihan, A.: A Systematic Review of Geographic Information Systems (GIS) in Agriculture for Evidence-Based Decision Making and Sustainability. Global Sustainability Research, Vol. 3, No. 1, 1-24. (2024)
36 Li, X.: Satellite Network-Oriented Visualization Analysis of 3D Geographic Information. Internet Technology Letters, Vol. 6, No. 2, e353. (2023)
37 Amin, K., Mills, G., Wilson, D.: Key Functions in BIM-Based AR Platforms. Automation in Construction, Vol. 150, No. 1, 104816. (2023)
38 Yu, J., Zhong, H., Bolpagni, M.: Integrating Blockchain with Building Information Modelling (BIM): A Systematic Review Based on a Sociotechnical System Perspective. Construction Innovation, Vol. 24, No. 1, 280-316. (2024)
39 Liu, C., Song, B., Fu, M., Meng, X., Zhao, Y., Wang, X., Li, X., Liu, Z., Han, Y.: Cesium-MRS: A Cesium-Based Platform for Visualizing Multi-Source Remote Sensing Data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 48, No. 24, 15-19. (2023)
40 Zhuang, S., Wang, J.: Cesium Removal from Radioactive Wastewater by Adsorption and Membrane Technology. Frontiers of Environmental Science & Engineering, Vol. 18, No. 3, 38. (2024)
41 Jin, J., Zeng, Y. J., Steele, J. A., Roeffaers, M. B., Hofkens, J., Debroye, E.: Phase Stabilization of Cesium Lead Iodide Perovskites for Use in Efficient Optoelectronic Devices. NPG Asia Materials, Vol. 16, No. 1, 24. (2024)

Renjun Liu was born in Wuhan, Hubei, P.R. China, in 1992. She received the Master degree from Columbia University, the U.S.A. Now, she works in Wuhan Zhenghua Architectural Design Co., Ltd.. Her research interest include architectural design and smart city. E-mail: rl2783@columbia.edu

Received: November 22, 2024; Accepted: February 24, 2025.

Computer Science and Information Systems 22(3):1197–1228 https://doi.org/10.2298/CSIS241123038L

Smart Home Management Based on Deep Learning: Optimizing Device Prediction and User Interface Interaction

Xuan Liang[1,2], Meng Liu[1,4], Hezhe Pan[3,*]

[1] Art College, Chongqing Technology And Business University, Chongqing, 400067, China
liangxuan@ctbu.edu.cn
[2] School of Design, Hunan University, Changsha, 410082, China
[3] Loudi Vocational and Technical College, Loudi, 417000, China
panhezhe001@163.com
[4] Chongqing Vocational College of Media, Chongqing, 400020, China
liumeng6@ctbu.edu.cn

Abstract. This work aims to address the challenges faced by smart home systems, including the accuracy of device status prediction, user interface design, system stability, and response speed. As smart home devices become more widely used, the need for accurate predictions of their operational status has increased. This includes predicting the switch states, faults, and performance metrics of devices such as smart lights, thermostats, and security systems. To address this demand, an innovative multimodal prediction model combining the Convolutional Neural Network and Long Short-Term Memory network is proposed to enhance the accuracy of smart device status predictions. Cloud computing technology is used for the user interface design to create an intuitive and user-friendly interface, ensuring both system stability and fast response times. The experiments compare the performance of the proposed model with traditional models in predicting the status of smart devices. The results demonstrate that the proposed system reduces the Mean Squared Error and Mean Absolute Error by 20% and 15%, respectively, significantly improving prediction performance. Furthermore, user satisfaction surveys indicate a 25% increase in satisfaction with the system. The proposed system also reduces the utilization rates of the Central Processing Unit, memory, Graphics Processing Unit, and network bandwidth by 15%, 18%, 25%, and 20%, respectively. These findings highlight the system's advantages in accuracy, user satisfaction, and resource utilization efficiency, providing strong support for the design and application of smart home systems.

Keywords: artificial intelligence; cloud computing; smart home; multimodal prediction model; user satisfaction.

1. Introduction

1.1. Research Background and Motivations

The rapid advancement of technology has made smart home systems an integral part of daily life [1-3]. As user expectations for these systems rise, especially in terms of the accuracy of device status predictions, current smart home technologies face significant challenges. These include limited capacity to process multimodal data and poor real-time response performance. Many traditional models struggle to predict device statuses accurately in complex environments, which impacts the user experience and limits the potential applications of these systems [4-6]. However, innovations in artificial intelligence (AI) and the improved processing capabilities of cloud computing have created new opportunities for home automation [7]. AI technologies, particularly machine learning and deep learning algorithms, enable home devices to learn and adapt to user preferences, providing personalized services [8-10]. Cloud computing offers powerful computational and storage support, facilitating remote connections and seamless data sharing among devices [11-13]. The integration of AI and cloud computing offers great potential for developing intelligent, efficient, and secure home management systems [14-16].

Despite significant advancements in both academia and industry, several challenges persist in the development of smart home systems [17]. One of the primary issues is the diversity of home devices and the lack of standardized protocols, which have led to interoperability problems [18, 19]. Additionally, the real-time performance and accuracy of intelligent algorithms still require improvement [20, 21]. Another critical concern is the safeguarding of user privacy and data security [22, 23]. Traditional smart home systems are limited by the capabilities of their intelligent algorithms, particularly in terms of real-time performance and accuracy. These limitations hinder the system's ability to make prompt and precise decisions, reducing both the user experience and the system's overall effectiveness. To address this, the focus of this work is on integrating machine learning and deep learning algorithms, optimizing them to improve the system’s real-time responsiveness and enhance its ability to adapt accurately to user habits in varied environments. Furthermore, the smart home market is fragmented, with various device brands and standards contributing to interoperability challenges. This issue makes it difficult for devices from different manufacturers to work seamlessly together, and compromises the integrated functioning of the system. Finally, smart home systems handle large amounts of sensitive data, such as information about household routines and user preferences, raising significant concerns regarding privacy and data security.

In summary, current smart home systems still face many challenges in device status prediction, user interface design, system stability, and response speed. With the widespread adoption of smart home devices, accurately predicting the operational status of devices (such as smart lighting, thermostats, security system switches, faults, and performance indicators) has become an urgent problem to address. However, existing smart home systems still have certain shortcomings in these aspects, particularly in the accuracy of device status prediction and the user-friendliness of the interface. Therefore, this work aims to propose an innovative multimodal prediction model that combines the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network

to improve the accuracy of device status prediction. Additionally, it incorporates cloud computing technology into the user interface design to enhance system stability and response speed, ensuring a better user experience.

The research motivation arises from an in-depth analysis of the shortcomings of current smart home systems, particularly the inaccuracy of device status prediction and the complexity of user interfaces, which affect user experience and system efficiency. Through the proposed multimodal model and optimized user interface design, this work aims not only to improve prediction accuracy but also to make significant improvements in user experience, system stability, and resource utilization efficiency. This innovative approach is expected to provide strong support for the design and application of smart home systems and promote the popularization and development of smart home technology in practical applications. This work prioritizes the design of a secure and reliable system to protect user privacy and mitigate potential threats. It incorporates encryption technologies, access control, and secure transmission protocols, while also implementing robust security measures within a cloud computing environment to defend against cyberattacks and data breaches. By leveraging an innovative multimodal predictive model, cloud computing for user interface design, and optimizing resource utilization, this work offers a comprehensive solution for developing whole-home intelligent management systems. The goal is to integrate AI technology with cloud computing to create a highly intelligent, secure, and reliable system. This system not only enables the intelligent control of home devices but also facilitates seamless information sharing and decision-making across devices. It significantly enhances the user experience in smart homes and lays a strong foundation for the continued evolution of smart home technologies.

1.2. Research Objectives

This work aims to design a highly intelligent home control system to enable automated control and intelligent scheduling of household devices. It also leverages cloud computing technology to create a secure and stable data platform, facilitating information sharing and remote control across devices. Through the development of smart algorithms, this work enables real-time monitoring and analysis of the home environment, providing personalized services. The approach enhances the system's scalability and interoperability, ensuring compatibility and seamless integration with smart devices from various manufacturers. Finally, real-world scenario validation ensures the system's stability and practicality. These innovations not only represent significant technological breakthroughs but also contribute to improved system performance, user experience, and resource utilization efficiency. They provide a solid foundation for the design and deployment of whole-house smart management systems, opening new possibilities for the future of smart home technology.

2. Literature Review

In the field of home automation, a wide range of commercial products has emerged, showcasing advanced smart control capabilities. In the area of intelligent lighting

systems, products like the Philips Smart Lighting System and Yeelight Smart Bulbs allow users to remotely adjust brightness, color temperature, and other settings via smartphone apps or voice assistants. In the domain of smart security, products such as the Ring Smart Video Doorbell and Nest Secure Smart Security System offer features like video monitoring, intrusion detection, and intelligent doorbells, enabling real-time home security management through smartphones. In the smart home appliance sector, products like the LG ThinQ series and Samsung Smart Refrigerators not only allow for remote monitoring but also incorporate intelligent features that adapt to user habits. Smart speaker systems, such as the Amazon Echo and Google Home, integrate voice assistants. They enable users to control smart home devices through voice commands and access a wide range of information and entertainment services. Additionally, various commercial products are available across other smart home domains, including smart temperature control, home automation, smart curtains and windows, entertainment systems, health monitoring, and kitchen appliances, covering nearly every aspect of modern living. These products demonstrate that smart home technology has become a practical and accessible solution, offering users enhanced convenience and intelligence in their homes. A significant body of research is dedicated to developing smart home systems that incorporate AI technology and cloud computing [24]. For example, literature [25] introduced the concept of a "user-friendly Internet of Things (IoT) for everyday living" in their design approach, creating an IoT-based smart home system. This system enabled functions such as displaying temperature and humidity data collected from node boards on a personal computer (PC) via a web browser. It also allowed users to control the on/off states of Light Emitting Diode lights through the same interface. In a similar vein, scholars proposed a system connecting sensors, actuators, and other data sources to enable more complex home automation tasks [26]. They also developed a smartphone application that allowed users to control various household appliances and sensors remotely.

Literature [27] developed a powerful and intelligent floor monitoring system using highly reliable frictional electric encoding pads and DL-assisted data analysis. They further integrated deep learning-assisted data analysis to enhance the system’s capabilities for various smart home monitoring and interactions. Literature [28] proposed a fully operational 46-inch smart textile lighting/display system. This system incorporated embedded optical fiber devices designed to detect external stimuli. Literature [29] introduced a comprehensive smart home aggregation system based on IoT and edge computing. The system leveraged edge AI support technology and adhered to industry standards for fog computing, providing robust responses from connected IoT sensors in typical smart homes. Literature [30] designed a secure remote user authentication scheme called SecFHome. This scheme supports secure communication at the network edge and enables remote authentication in fog-based smart home systems. Literature [31] presented an IoT-based smart home management system. The system utilized sensors, actuators, smartphones, network services, and microcontrollers for enhanced functionality.

Research on data privacy protection in smart homes has garnered significant attention, particularly focusing on the application of the Deep Deterministic Policy Gradient (DDPG) algorithm as a core predictive model in modern power systems. This approach enhanced prediction accuracy [32]. The DDPG predictive model is later integrated into the federated learning framework. The resulting Federated Deep Reinforcement Learning (FedDRL) model mitigates privacy concerns by sharing model

parameters instead of private data, ensuring accurate predictive models are obtained in a decentralized manner. Literature [33] proposed a False Data Injection Attack (FDIA) detection method based on secure federated deep learning. The method's effectiveness and superiority were demonstrated through extensive experiments on IEEE 14-bus and 118-bus test systems. Literature [34] addressed privacy concerns in smart home technology, particularly the risks of data leakage through wireless signal eavesdropping. They discussed "FTS (fingerprint and timing-based snooping)" attacks, a type of side-channel attack that could passively infer activity information within residences. These attacks can be executed remotely near the target house. Literature [35] applied the Sovereign design philosophy to enable communication between home IoT devices and applications via application-named data, directly protecting the data. The results indicated that Sovereign offers a systematic, user-controlled solution for self-contained smart homes, with minimal observable overhead when running on existing IoT hardware. Literature [36] proposed a location privacy security mechanism based on anonymous trees and box structures. This approach provided location privacy protection for services targeting smart terminals.

In the field of smart homes, numerous studies have examined the integration of AI technology with cloud computing, driving advancements in the design and implementation of smart home systems. These studies often focus on specific applications, such as smart lighting and environmental monitoring [37]. However, they tend to lack a comprehensive analysis and optimization of the entire smart home ecosystem. This gap indicates that, while smart home systems offer convenience to users, there is still significant room for improvement in their overall performance. Key factors, such as the accuracy of device status predictions, the intuitiveness of user interfaces, and system response speed, have not been sufficiently explored. To address these challenges, this work aims to provide a holistic analysis and optimization of smart home systems, ultimately enhancing both performance and user experience. The primary objective is to improve the accuracy of smart device status predictions through an innovative multimodal predictive model that combines CNN with LSTM networks. In terms of user interface design, this work utilizes cloud computing technology to enable seamless data sharing and collaborative computation between devices. This approach creates an intuitive and user-friendly interface while ensuring system stability and responsiveness. Furthermore, by analyzing user behavior data, this work offers personalized services that make the smart home system more closely aligned with individual user needs.

3. Research Methodology

In smart home systems, effective coordination between the multimodal predictive model and the Deep Q Network (DQN) intelligent control algorithm is essential. Each component has a distinct role, and together, they enable accurate device status prediction and optimized control. The multimodal predictive model, which combines CNN and LSTM, processes complex data from various smart devices. It generates predictions about device statuses, providing a comprehensive view of the system's current and future states. The DQN serves as the intelligent control algorithm, making real-time decisions based on the outputs of the predictive model. Through continuous

learning and optimization, the DQN selects the best control strategy to adjust the device's operating states. Central to its operation is the Q-value function, which estimates the expected rewards of different actions in various states. This function guides the system in selecting the most effective actions to optimize device control. The integration of the predictive model and the DQN allows the system to both predict device statuses accurately and make intelligent decisions on how to adjust operations. The predictive model interprets complex multimodal data to provide precise insights into future device states, while the DQN uses these insights to refine control strategies. This collaboration enables the smart home system to not only forecast device states in real time but also dynamically adjust operations, creating a more intelligent and responsive home environment. The following sections provide further details on the components and functioning of these processes.

3.1. Construction of Smart Device State Prediction Model

To enable intelligent control and optimization of home devices, this work designs a smart device status prediction model that combines LSTM networks with CNN. This model intends to enhance both the prediction accuracy and response speed for smart device statuses. The choice of this combined approach is based on its proven ability to improve predictive performance effectively. Although models such as Autoregressive Integrated Moving Average (ARIMA) and Deep Neural Network (DNN) are commonly used for time series forecasting, they are less suited for the specific task of predicting smart home statuses. The ARIMA model struggles with nonlinear time series data, particularly when the data involves seasonality or abrupt events, which limits its predictive accuracy [38]. DNN fails to capture short-term memory as efficiently as LSTM networks, resulting in lower accuracy and poorer real-time performance.

To address these issues, this work selects an innovative multimodal predictive model that combines CNN and LSTM. This model effectively manages the complexity and dynamics of smart home systems and provides more accurate status predictions. Given the complexity of intelligent device states, which are influenced by various factors, a deep learning model is selected for its ability to handle such data. Single deep learning models, like CNN or LSTM alone, may struggle to capture the complex relationships in spatiotemporal data due to the multi-modal nature of device states. To comprehensively leverage the features of image and time series data, a multimodal prediction model that integrates CNN and LSTM is chosen. This approach enables a more comprehensive understanding of both image and time-series data features. Although ensemble learning methods such as Random Forest or Gradient Boosting Trees, and model fusion methods like Stacking, can provide strong performance, they often require extensive tuning and feature engineering [39]. Considering the goal of exploring the complex relationships between image and time-series data, the CNN-LSTM integrated model is ultimately chosen. While rule-based methods could predict device states, they rely on predefined domain-specific knowledge and struggle to adapt to evolving patterns in the data.

By considering information in both spatial and temporal dimensions, the model better adapts to the complexities of home scenarios. This work adopts a multimodal prediction model to better accommodate the data features and predictive requirements of smart home systems. This model can handle time-series data and multi-sensor image data to

accurately predict the state of household devices, providing the foundation for intelligent control. Figure 1 is an illustrative diagram of the model.

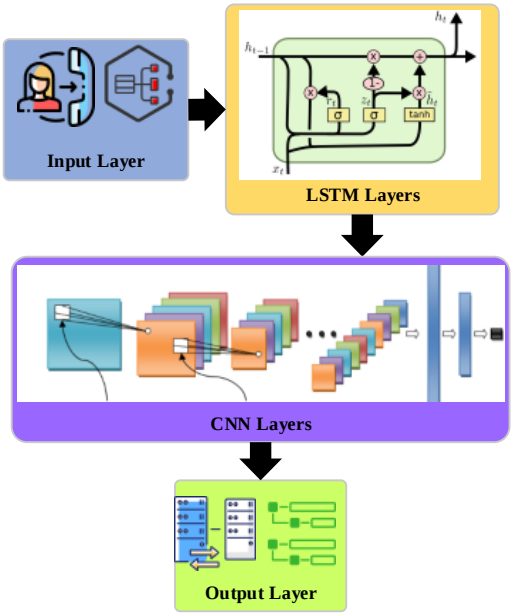


Fig. 1. Schematic diagram of the smart device state prediction model

The model in Figure 1 consists of four main components. The input layer receives data from both time-series and multi-sensor image sources. Time-series data, such as temperature and humidity, are processed through the LSTM network. Meanwhile, image data, such as infrared images, are processed by the CNN. The LSTM layer handles time-series data, focusing on the sequential nature of the information and capturing long-term dependencies. The CNN layer processes the image data, extracting spatial features from the multi-sensor images to enhance the prediction of device states. Finally, the output layer generates predictions for the smart device states.

In processing image data, CNN extracts spatial features through convolution operations. It is assumed that the input image data are denoted by X. H is the height of the image, W is the width, and C is the number of channels (such as three channels for RGB images, C = 3). The convolutional layer operates on the image using a convolution kernel (filter) to generate a feature map. The mathematical equation for the convolution operation is as follows:

$$Y_{i,j} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{i+m,j+n} \cdot K_{m,n} \quad (1)$$

In this process, $Y_{i,j}$ represents the element of the output feature map, which indicates the value after the convolution operation. X is the input image, K is the convolution kernel, and M and N are the size of the kernel; i and j are the indices of the feature map. The convolution operation is used to extract local spatial features, with the convolution kernel performing a dot product with the input image through a sliding window, generating a new feature map. These feature maps capture basic structures of objects in the image, such

as edges and corner points. After the convolution operation, a pooling operation is usually applied to further reduce the size of the feature map while retaining important spatial information. The pooling operation can be performed using either MaxPooling or AveragePooling, with the corresponding equation as follows:

$$Y_{i,j} = \max\left(X_{i:i+k,j:j+k}\right) \tag{2}$$

In this process, k represents the size of the pooling window, and $X_{i:i+k,j:j+k}$ denotes the local region extracted from the input image. Pooling operations help reduce dimensionality and computational load, while also preventing overfitting. Through multiple layers of convolution and pooling operations, CNN can progressively extract more complex spatial features, such as the shape and position of objects. Ultimately, these features are aggregated in the fully connected layer and used for decision-making in device state prediction.

The outputs from both the LSTM and CNN are integrated to provide a comprehensive analysis of device states. This integration allows the model to collect information from multiple data sources, and improves both prediction accuracy and precision. By utilizing multi-layered and multi-source data, the smart device state prediction model offers a more accurate forecast of household device statuses. This, in turn, provides a reliable foundation for intelligent home system control.

The input for the LSTM model is based on time-series data from smart devices. These data include the device's historical status, sensor readings (such as temperature, humidity, and brightness), and other relevant features. To ensure the LSTM can effectively learn from these inputs, the raw data are organized and structured so that each time step corresponds to the appropriate sensor information and historical status. Specifically, data from the previous 10 time steps are selected for each point in time, and a sliding window method is used to generate training samples. These samples are then fed into the LSTM model for status prediction. The LSTM, a type of recurrent neural network with memory units, is designed to remember long-term dependencies [40-42]. In this case, it processes time-series data, such as device status information like temperature and humidity. The LSTM's architecture includes input, forget, and output gates, which enable it to capture long-term dependencies within the time-series data [43]. The mathematical expression is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \cdot tanh\left(C_t\right) \tag{8}$$

x_t represents the input data, h_t is the current time-step hidden state, and C_t is the current time-step cell state. f_t, i_t, and o_t are the outputs of the forget gate, input gate, and output gate, respectively. W and b denote the weight and bias. σ is the sigmoid activation

function, and tanh is the hyperbolic tangent activation function. In the smart home system context, this work uses CNN to process multi-sensor image data, such as infrared images. The CNN network's structure includes convolutional layers, pooling layers, and fully connected layers. These components work together to effectively extract spatial features from the images [44-46]. The expression reads:

$$h_i = \sigma(W * x_i + b) \tag{9}$$

$$y = \text{softmax}\ (h) \tag{10}$$

x_i is the i-th region of the input image, and h_i is the feature representation of that region. W represents the convolutional kernel, and b is the bias. σ is the ReLU activation function, $*$ denotes the convolution operation, and the softmax function is used for multi-class output.

3.2. Selection and Optimization of Intelligent Control Algorithms

Next, this work explores the selection and optimization of intelligent control algorithms. This section conducts an in-depth analysis of how to choose the most suitable control algorithm for smart home environments and optimize control strategies using reinforcement learning techniques. By working in synergy with the predictive model, the control algorithm can respond in real-time to changes in device status and make optimal decisions. This can effectively improve the operational efficiency and user experience of the smart home system. This work selects the DQN algorithm from deep reinforcement learning as the intelligent control algorithm for the smart home system. DQN is known for its strong generalization and learning capabilities, making it well-suited for handling large-scale, high-dimensional state spaces [47-49]. The core idea behind DQN is to construct a Q-value function that represents the value of taking a specific action in a given state. A neural network is then used to approximate this Q-value function, which is essential for action prediction and selection [50]. The mathematical expression is as follows:

$$yQ(s,a) = (1-\alpha)\cdot Q(s,a) + \alpha\cdot(r + \gamma\cdot \max Q(s',a')) \tag{11}$$

$Q(s,a)$ is the Q-value for taking action a in state s, α represents the learning rate, r is the immediate reward, and γ is the discount factor. s' is the next state, and a' is the best action in the state s'. To improve the efficiency and performance of the control algorithm, this work deploys it on cloud computing resources and leverages big data for optimization. The high-performance cloud infrastructure ensures real-time algorithm capabilities, while big data analysis uncovers additional control patterns and optimization strategies. Figure 2 depicts the structure of the DQN algorithm.

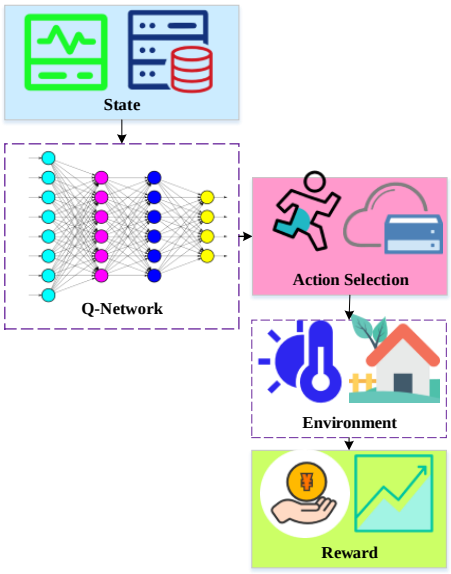


Fig. 2. Schematic diagram of the DQN algorithm structure.

In Figure 2, the State represents the current environmental states of the smart home system, including sensor data and device statuses. The Q-Network is a neural network used to estimate Q-values for specific actions in a given state. The network takes the current state as input and outputs Q-values for possible actions. "Action Selection" chooses the next action based on the Q-values using an "ε-greedy" strategy, where, with a certain probability (ε), actions are randomly selected to increase exploration. With a probability of 1-ε, the action with the highest Q-value is chosen to enhance exploitation. Environment represents the physical environment of the smart home system, including various sensor data and device statuses. The Reward is the immediate reward signal obtained by the smart home system based on the actions taken by the intelligent control algorithm. This structure allows the DQN algorithm to learn the optimal control strategy through continuous interaction with the environment.

3.3. Design of an AI and Cloud Computing-Based Smart Home Management System

This section demonstrates how to integrate the previously discussed predictive model with intelligent control algorithms into a complete system. The system not only performs device status prediction and intelligent control but also leverages cloud computing technology to store, process, and analyze data, providing more efficient and flexible management and services. Cloud computing technology is primarily utilized to enhance the flexibility and responsiveness of the system interface. By offloading data processing and storage to the cloud, the burden on local devices is reduced. This allows the smart home system to provide faster information access and a smoother user experience. Cloud computing also supports real-time collaborative computation among

different smart devices, enhancing their interoperability and coordination. Moreover, it enables users to remotely access and control home devices, further improving system operability and user satisfaction. Overall, the implementation of cloud computing technology makes smart home systems more resilient, scalable, and collaborative. It enhances the intuitiveness and user-friendliness of the interface, which directly boosts the user experience [51]. These improvements can be measured through methods such as user satisfaction surveys and interface interaction analysis. This work presents a smart home management system based on AI and cloud computing, as illustrated in Figure 3.

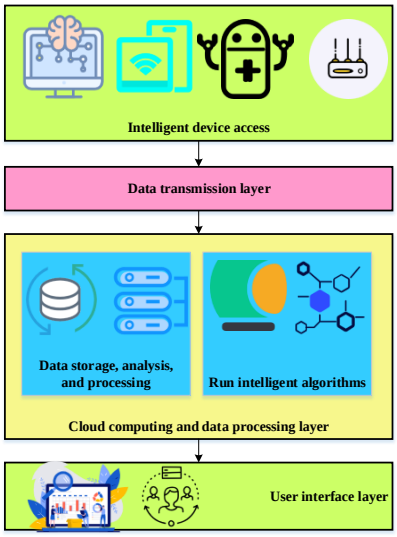


Fig. 3. Framework of the smart home management system

Figure 3 illustrates the overall architecture of the management system, which consists of several key components. The intelligent device access layer communicates with various smart devices, such as sensors and actuators, to collect real-time data. The data transmission layer ensures the timely and reliable transfer of these data to the cloud computing and data processing layer. This layer handles tasks like data storage, analysis, and processing, running intelligent algorithms to support decision-making. The user interface layer provides a platform for users to interact with the system, allowing them to monitor and control household devices via mobile apps, web interfaces, and other means. With this structural composition, the work can achieve the following functions:

1) Intelligent device access and management: it facilitates communication and access to various smart devices, ensuring they operate properly.

2) Real-time data transmission and processing: sensor data are transmitted to the cloud in real-time. The cloud processes the data through cleaning and feature extraction.

3) Intelligent control and optimization: cloud computing's high-performance computing capabilities are used to run algorithms that control and optimize household devices, improving energy efficiency.

4) Data storage and analysis: processed data are stored in a cloud-based database for further analysis, mining, and visual representation.

5) Remote monitoring and operation: users can remotely monitor device status and perform operations through mobile applications or web interfaces.

This work introduces the Generative Adversarial Network (GAN) to enable continuous learning and adaptability of the predictive model to respond to changes in new data and user behavior. Specifically, GAN consists of two main components: the generator and the discriminator. The generator's task is to generate data based on actual user behavior patterns, while the discriminator is responsible for evaluating the similarity between the generated data and real data. In the initial stage, the generator learns the patterns in user behavior data to generate preliminary behavior data, while the discriminator learns to distinguish between generated and real data. This process helps the model understand and capture the basic patterns of user behavior. Figure 4 displays the specific process.

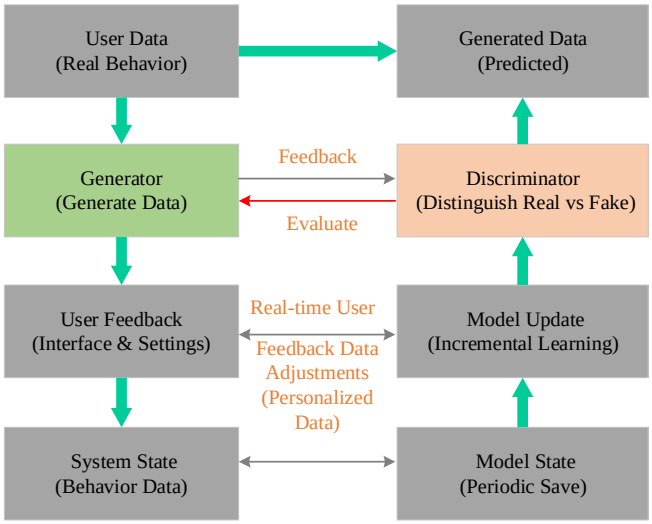


Fig. 4. GAN structure and workflow

During the system's operation, real-time incoming user behavior data are used for incremental learning. The generator and discriminator will continuously update to adapt to new data characteristics. This continuous learning mechanism ensures that the model can timely acquire information from new data and adjust according to changing user behavior. By dynamically optimizing the GAN parameters, the model can flexibly respond to these changes, ensuring the accuracy of predictions. Furthermore, the system employs a real-time feedback mechanism to further optimize the generated data. User feedback, such as interface ratings and the selection of personalized settings, is used to guide the model in adjusting the generated data, making it more aligned with the users' actual needs and preferences. This mechanism ensures the timeliness and precision of the generated data, enhancing the adaptability and accuracy of the predictive model. In short, the role of GAN here is to assist the predictive model in continuously learning, adapting to new data, and adjusting in real time to respond to changes in user behavior

through generating and evaluating data. In this way, GAN effectively strengthens the intelligent home system's adaptability to user needs and the accuracy of predictions.

The incremental learning strategy enables the GAN to integrate new knowledge dynamically during the learning process. The model's state is saved periodically, allowing for rollback or recovery when needed, ensuring system stability. To evaluate performance, the system regularly compares generated data to real data and considers user satisfaction metrics. User feedback plays a crucial role in assessing system performance, helping identify areas for improvement and model adjustments. By analyzing user behavior patterns, the system detects both short-term and long-term trends, which helps predict user needs and adjust the status of smart devices accordingly. The system uses a personalized learning model to tailor predictions based on each user's unique behavior and preferences. Real-time data from sensors and devices are collected regularly to update the knowledge base, reflecting the current environment and user behavior. This combination of mechanisms allows the system to continuously learn and improve its understanding of user behavior, delivering more intelligent and personalized home management services. Its dynamic adaptability ensures high performance, even during long operational periods.

To clearly illustrate the operation process of the intelligent home system proposed, an example of a smart lighting system within a smart home environment is presented. It is assumed that the system needs to predict the state of the light (on or off) and make corresponding control decisions. In this scenario, the smart home system first collects data from multiple sensors, including indoor light intensity, temperature, user activity data, and the historical on/off status of the lights. These data are processed through a multimodal predictive model. CNN is responsible for extracting spatial features from the environmental data, and LSTM networks handle the time-series data, capturing the usage patterns and historical dependencies of the lights. By combining both approaches, the system can accurately predict the state of the light for the upcoming period. For example, based on the data from the past hour, the predictive model might conclude that the light will remain on in the near future.

Next, the DQN intelligent control algorithm utilizes these predictions to make control decisions. The system's state space includes the current state of the light, environmental lighting conditions, temperature, and user activity status. Based on this information, DQN will choose the most appropriate action, such as "turn on" or "turn off" the light. The selection of each action is determined by a reward function, which not only considers the current state of the light but also scores based on user needs and system energy efficiency. For instance, if it is predicted that a user is entering the room, the system will select the "turn on" action because it enhances the user's experience and results in a higher reward.

Through continuous reinforcement learning, DQN can optimize its control strategy. If the system finds that a certain control strategy (such as delayed light-off) performs excellently in terms of energy saving, DQN will adjust its decisions based on accumulated rewards, thus improving the overall system performance. This collaborative approach enables the predictive model and control algorithm to dynamically adjust and optimize the device's state. It can ensure optimal performance of the smart home system in terms of response speed, energy efficiency, and user experience. This specific example provides a clearer understanding of how the proposed approach effectively collaborates within a smart home system. It also offers a more

intuitive grasp of how the predictive model and DQN intelligent control algorithm work together to optimize device management.

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection and Data Preprocessing

To validate the performance of the proposed smart home system, various types of environmental data and sensor information are systematically collected from multiple rooms in a residential community. These rooms include the living room, kitchen, bedroom, bathroom, and study/office area, covering the main living scenarios within a household. Environmental data are collected in real time using temperature, humidity, and light sensors, with a recording frequency of once per minute. Specifically, the sensor data cover the following environmental factors. Temperature sensors are deployed in each room to monitor indoor temperature changes in real time, recording data once per minute to capture rapid temperature fluctuations. Humidity sensors are placed in various locations to record changes in indoor humidity, particularly in areas with significant humidity variations, such as the kitchen and bathroom. Light sensors are used to monitor indoor light intensity, especially in the living room and bedroom, to assist in controlling the smart lighting system. In addition, multi-sensor image data are collected through internal cameras and infrared sensors, including object distribution information. The data provide insights into the placement of furniture and equipment within the home, helping the model understand the spatial layout. Human activity monitoring: The data track the activity patterns of household members, identifying specific behaviors (such as entering or leaving a room, and using devices), which serve as the basis for intelligent control. These sensor data, collected through embedded devices, ensure high-frequency recording and provide detailed temporal information. Figure 5 illustrates the data collection process and preprocessing steps.

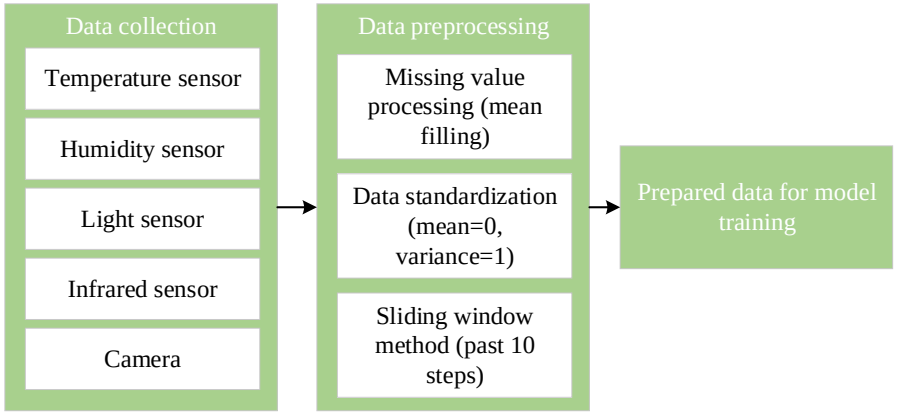


Fig. 5. Data collection and preprocessing flow

In Figure 5, during the data collection phase, the system collects real-time environmental data through temperature sensors, humidity sensors, light sensors, infrared sensors, and cameras. These data include temperature, humidity, light intensity, object distribution, and human activity data recorded every minute. The real-time data from these sensors provide comprehensive environmental information for the smart home system. Next, in the data preprocessing phase, to ensure data quality, missing values are first handled using mean imputation to maintain data completeness. Then, all sensor data are standardized and transformed into a standard normal distribution with a mean of 0 and a variance of 1. This can eliminate scale differences between different sensor data and enhance the stability of data training. Finally, a sliding window method is used to generate training samples from data collected over the past 10 time steps, providing rich contextual information for the subsequent LSTM model. The processed data ultimately form a prepared dataset for model training, ensuring consistency and quality of the data.

The diversity of data allows the model to adapt better to different environments and usage scenarios, improving its generalization ability. With several terabytes of data, the model can cover a wide range of scenarios and changes, making it more adaptable. By processing a large number of samples, the model can identify latent patterns between smart device states and the environment, improving its predictive accuracy. Additionally, the richness of multimodal data plays a key role in enhancing the model's robustness. Sensors such as temperature, humidity, and light provide diverse information, while image data offers a visual complement. This diversity allows the model to learn from different dimensions, and gain a comprehensive understanding of the relationship between smart device states and the environment. As a result, the model becomes more adaptable to complex situations. Overall, large-scale and diverse datasets offer a strong foundation for training, enhancing the model's robustness and performance in real-world smart home scenarios.

Before inputting the sensor data into the model, a systematic preprocessing process ensures data quality and consistency. First, missing values in the raw data from all sensors are handled using mean imputation to ensure completeness. Next, the data from different sensors are standardized to have a mean of 0 and a variance of 1, which eliminates dimensional effects between features and improves model stability during training. Finally, the sliding window technique is used to construct training samples, ensuring each sample contains data from the previous 10 time steps. This provides rich contextual information for the LSTM model's learning process.

4.2. Experimental Environment

Experiments are conducted on multiple high-performance servers equipped with Intel Xeon Gold 6226R processors (2.9 GHz, 16 cores) and 128 GB DDR4 RAM. These servers offer the computational power needed to handle the complex data processing requirements of the smart home system. To speed up the training of deep learning models, the system uses an NVIDIA Tesla V100 GPU (32 GB VRAM), which supports efficient parallel computing. The experiments are deployed on a cloud computing platform using AWS EC2 instances (p3.16xlarge type), providing scalable computing resources to meet the dynamic demands of various data processing tasks. Through cloud services, the system can elastically scale based on workload, ensuring efficient

operation during data processing at different scales. Additionally, the experimental environment integrates a distributed storage system based on Amazon S3 to securely store large-scale data. This system ensures high reliability and scalability, protecting data during processing while supporting real-time access and sharing of big data. The software environment runs on the Ubuntu 20.04 LTS operating system, with deep learning frameworks TensorFlow 2.10 and PyTorch 1.12. GPU acceleration is provided by the NVIDIA CUDA 11.4 toolkit. During model training, the Adam optimizer and an adaptive learning rate adjustment strategy are used. Training progress is monitored via TensorBoard to ensure continuous optimization of the model. This experimental setup not only offers robust hardware support for developing and testing the smart home system, but also enables effective handling of large-scale data and multi-task computing demands.

4.3. Parameters Setting

To ensure the system's stability and performance, the parameters of different models are carefully set during the experiment, as shown in Tables 1-4. The CNN and LSTM models undergo systematic experimentation and tuning to optimize their hyperparameters. For the CNN model, the initial learning rate is set to 0.001, with a batch size of 32. It includes three convolutional layers, each containing 64 filters of size 3×3. The depth of the convolutional layers and the number of filters are adjusted, and cross-validation is used to select the best combination for maximum predictive performance. For the LSTM model, the hyperparameters include a learning rate of 0.001, a time step of 10, and 50 hidden units in the layers. A grid search is conducted to find the optimal configuration, improving the model’s accuracy and robustness. These tables not only apply to the proposed method but also include the parameter configurations for comparison models, such as CNN, LSTM, and DNN.

These tables not only apply to the proposed method but also include the parameter configurations for comparison models (CNN, LSTM, and DNN).

Table 1. Parameter settings of the proposed model.

Parameters	Range of Values
Number of neural network layers	3
Number of LSTM layers	2
Number of CNN layers	1
Number of neural network nodes	128 (Each hidden layer)
LSTM hidden layer units	64
CNN filter size	3x3
CNN kernel number	32
CNN stride	1x1
Learning rate	0.001
Discount factor	0.9
The ε value for ε-greedy strategy	0.1
Maximum training steps	100,000
Optimizer	Adam optimizer
Batch size	64
Loss function	Mean Squared Error (MSE) loss
Proportion of training dataset	80% training data, 20% validation data

In the table, based on the structure of the hybrid model combining LSTM and CNN, the number of layers and relevant parameters for both LSTM and CNN are specified. The LSTM section includes two hidden layers, each containing 64 units. The CNN section consists of one convolutional layer, using a 3x3 filter, with 32 convolutional kernels and a stride of 1x1. Other parameters such as learning rate, optimizer, and others are also listed in detail. These settings can aid the model in effective learning and optimization during the training process.

Table 2. Parameter settings of CNN.

Parameters	Range of Values
Number of convolutional layers	5
Number of filters per layer	32-256
Filter size	3×3, 5×5
Activation function	ReLU
Pooling layer type	Max pooling
Batch size	64
Optimizer	Adam optimizer
Loss function	Cross-entropy loss

Table 3. Parameter settings of the LSTM model.

Parameters	Range of Values
The number of LSTM units	128
Sequence length	30
Learning rate	0.001
Optimizer	Adam optimizer
Batch size	64
Loss function	MSE loss

Table 4. Parameter settings of DNN.

Parameters	Range of Values
Number of neural network layers	4
Number of nodes per layer in the neural network	64-256
Learning rate	0.001
Activation function	Tanh or Sigmoid
Batch size	64
Optimizer	Adam optimizer
Loss function	MSE loss

4.4. Performance Evaluation

This work uses MSE and Mean Absolute Error (MAE) as performance evaluation metrics. MSE measures the sum of squared errors, while MAE calculates the mean of absolute errors. Both metrics are sensitive to larger error values. In smart home systems, where

critical state predictions like temperature and humidity are essential, the focus is on the model's accuracy in predicting real values. These metrics effectively highlight the impact of larger errors on performance. MSE and MAE are widely used in regression tasks across various domains. Their simplicity and ease of understanding make them ideal for evaluating model performance in different prediction tasks. This interpretability allows researchers and practitioners to quickly comprehend the model's performance in smart home systems. MSE assigns higher weights to larger errors, providing a better reflection of the model's performance in critical predictions. In contrast, MAE maintains a linear relationship with error magnitude, sometimes offering a clearer view of overall average performance. The mathematical properties of MSE and MAE also simplify their use in optimization problems. During the training of deep learning models, minimizing these metrics through algorithms like gradient descent is straightforward, allowing for better adjustment of model parameters. In summary, MSE and MAE are classical metrics that provide a comprehensive and intuitive assessment of model's predictive performance, especially in predicting smart device states. Their use here contributes to a deeper understanding of the accuracy and overall performance of the model concerning smart device states. In the experiment, the evaluation indicator MSE is adopted to assess the predictive accuracy of the smart device state prediction model. The equation for calculating MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{12}$$

y_i is the actual value, $\hat{y}_i$ is the model's predicted value, and n is the number of samples. MAE is similar to MSE and is applied to assess the difference between predicted values and actual values. The calculation equation is as follows:

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}| \tag{13}$$

R-Squared is a commonly used metric to measure the goodness of fit of a regression model. The equation is as follows:

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i}-\hat{\mathrm{y}}_\mathrm{i})^2}{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i}-\bar{\mathrm{y}})^2} \tag{14}$$

y_i is the actual observed value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual observed values, and n is the number of samples. The value of R-Squared ranges from 0 to 1, with values closer to 1 indicating better model fit. This equation is described in detail here to better showcase the predictive performance of the model.

The following are the comprehensive evaluation results of system performance, including comparisons with specific models. Figure 6 represents the performance of smart device state prediction:

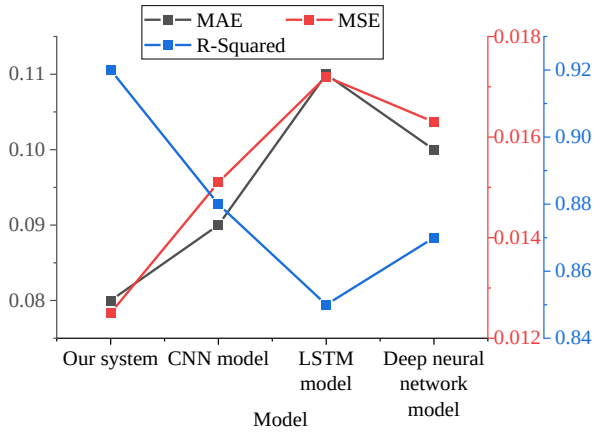


Fig. 6. Performance of smart device state prediction

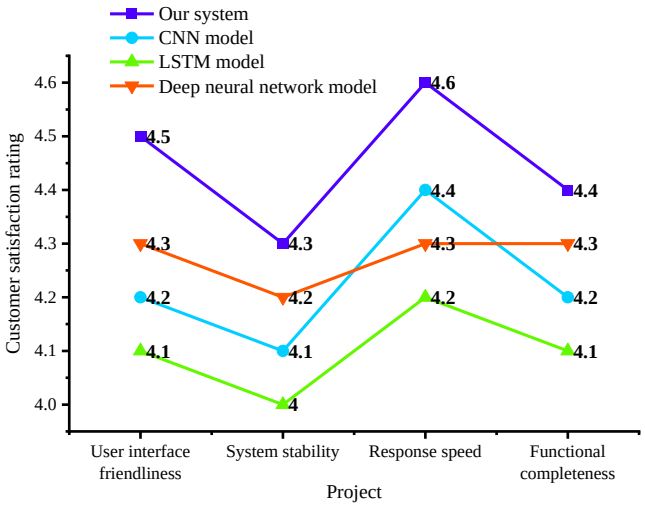


Fig. 7. User satisfaction survey results

Figure 6 demonstrates the system's strong performance in predicting smart device states. Specifically, it achieves values of 0.0125 for MSE and 0.08 for MAE, both lower than those of other models. This indicates that the proposed system predicts smart device states more accurately. Its MSE and MAE are 0.0026 and 0.01 lower than the CNN model, 0.0047 and 0.03 lower than the LSTM model, and 0.0038 and 0.02 lower than the DNN model. Reducing prediction errors results in a more reliable smart home experience for users. The system incorporates a multimodal fusion model that combines the feature extraction capability of CNN, the time-series data handling of LSTM, and the deep learning power of DNN to optimize prediction accuracy. This improvement not only enhances data processing but also strengthens the model's generalization ability, allowing it to handle complex state changes across different devices more effectively.

The system's enhanced predictive ability makes the smart home more stable and responsive, reducing abnormal situations in device state management.

Appendix A contains the user satisfaction survey form. Figure 7 presents the results of the survey.

Figure 7 shows that the user satisfaction survey results further highlight the system's outstanding performance. In terms of user interface friendliness, system stability, response speed, and feature completeness, the system receives ratings of 4.5, 4.3, 4.6, and 4.4, respectively. In comparison, other models receive ratings of 4.2, 4.1, 4.4, and 4.2 in these areas. The user interface design is crucial to the overall user experience. The system scores 4.5 for interface friendliness, which is significantly higher than the other models' score of 4.2. This difference reflects the system's optimized interface design, with an intuitive layout and clear interaction flow. Users can easily navigate the system's functions, reducing confusion and operational errors. Feedback also indicates that the system's interface adapts well to different devices. Whether on a mobile phone, tablet, or smart home control panel, the interface operates smoothly across all platforms, greatly enhancing user comfort.

System stability is vital in smart home applications, especially when managing multiple devices and tasks simultaneously. The system scores 4.3 for stability, an improvement over the 4.1 score of other models. This enhancement is due to optimizations in the system architecture, particularly for real-time processing of multisensor data and device coordination. User feedback suggests that the system maintains high efficiency and stability over time, avoiding the crashes and functionality failures seen in other models. The system's stable performance when handling large volumes of real-time data has built user trust, further boosting satisfaction.

Response speed is a key factor in the user experience of a smart home system. The system's response speed is rated 4.6, higher than the 4.4 rating for other models. Users report that the system responds almost instantly, with no noticeable delay when operating devices. Whether adjusting the temperature, turning lights on and off, or switching smart scenes, the system provides feedback in milliseconds (ms), offering a seamless experience. In comparison, other models often show slight delays in their operations, affecting both the timeliness of actions and user satisfaction.

Functional completeness evaluates how well the system meets diverse user needs, including smart control, personalized settings, and scene switching. The system receives a high score of 4.4 for functional completeness, surpassing the 4.2 score of other models. This score difference reflects the system's broad functionality. It supports seamless connection and management of multiple smart devices and can automatically adjust the home environment based on user preferences. Additionally, it offers personalized settings and custom scene options, allowing users to tailor the system to their needs. Users report that the variety of features and the system's flexibility have increased both their reliance on and satisfaction with the system.

When comparing the system's interface design to other models, such as CNN, LSTM, and DNN, these models often have limitations in terms of user interface intuitiveness, response speed, and functional completeness. The CNN model often requires users to manually adjust numerous hyperparameters, while LSTM models can complicate the interface when processing time-series data, raising the learning curve for users. In contrast, the developed system optimizes the interface layout and user interaction flow, significantly improving ease of use. It ensures faster response speeds through efficient computational resources. Moreover, the proposed system integrates rich functional

modules, such as real-time data visualization and smart feedback, which make it easier for users to perform complex tasks. This design has greatly improved user satisfaction, especially in terms of ease of use and functionality. Figure 8 displays the system's resource utilization.

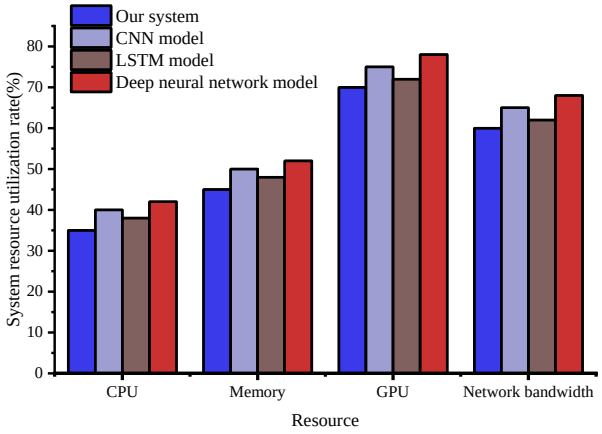


Fig. 8. Comparison of system resource utilization

Figure 8 shows that the proposed system excels in resource utilization. It uses less CPU, memory, GPU, and network bandwidth compared to other models. Specifically, the system's utilization rates are 35% for CPU, 45% for memory, 70% for GPU, and 60% for network bandwidth. In comparison, the CNN model uses 40%, 50%, 75%, and 65%; the LSTM model uses 38%, 48%, 72%, and 62%; and the DNN model uses 42%, 52%, 78%, and 68% for the same categories. This demonstrates that the proposed system operates more efficiently, saving computational resources and offering a more cost-effective smart home management solution. Further analysis of the data in Figure 7 highlights that the improvement in resource efficiency is mainly due to the optimized design of the model architecture. By leveraging cloud computing technology and efficient parameter tuning, the system achieves high performance while reducing resource consumption. This makes the system suitable for single-home scenarios and scalable for large-scale smart home deployments. It reduces hardware requirements and enhances system performance when managing multiple devices and data streams. This optimization lays a solid foundation for the widespread adoption and promotion of smart home systems.

This work observes that there are differences in system resource utilization among the four models. The reasons for these differences are analyzed from the following aspects:

1) Model Complexity and Computational Requirements: Different models have varying complexities and computational demands. For instance, CNN and DNN typically require more computational resources, especially in image processing and training deep networks with multiple layers. In contrast, LSTM networks, although designed for time-series data, are relatively more efficient in computation, especially when there is no need for extensive parallel computing. The proposed hybrid model

combines the advantages of LSTM and CNN, better balancing computational resource usage during multi-modal data processing. It prevents excessive consumption of CPU, memory, and GPU resources.

2) Application of Cloud Computing: The system leverages cloud computing for distributed computing, which effectively reduces the computational burden on individual hardware devices. Through elastic resource management on the cloud platform, the model can dynamically allocate computational resources based on demand and avoid resource wastage. This is one of the reasons why the proposed system performs exceptionally well in resource utilization. Cloud computing not only improves computational efficiency but also reduces the dependence on local hardware and decreases the high-load demands on CPUs and GPUs.

3) Parameter Optimization and Network Bandwidth Management: Through efficient parameter tuning, the system optimizes resource allocation during training. The relatively low network bandwidth utilization (60%) indicates that the system has optimized data transmission, reducing bottlenecks caused by frequent data exchanges. This is significant for a smart home system that needs to handle large amounts of multi-modal data (such as temperature, humidity, and images) and real-time feedback.

4) Algorithm Efficiency: The proposed system uses a hybrid architecture of LSTM and CNN, which improves computational efficiency while ensuring prediction accuracy. The LSTM model effectively captures long-term dependencies when processing time-series data and reduces unnecessary computations. The CNN model efficiently extracts spatial features from image data. Through this architectural optimization, the system maintains high performance while significantly reducing computational resource requirements.

Overall, the differences in resource utilization among the four models mainly stem from their respective architectural features, computational demands, and optimization strategies. Compared to other traditional models (such as CNN and DNN), the proposed hybrid model demonstrates superior performance in optimizing resource usage and reducing computational demands. This enables the system to maintain efficient and stable performance while processing multiple devices and data streams.

Next, this work compares the performance of the proposed smart home system with CNN, LSTM, and DNN in terms of latency and computation time. Table 5 displays the results.

Table 5. Comparison of performance in latency and computation time for different models

Model Type	Latency (ms)	Computation Time (ms)
The proposed model	85	120
CNN	120	180
LSTM	110	160
DNN	130	190

Table 5 shows that the proposed model performs best in terms of latency, with a value of only 85 ms. It outperforms the CNN (120 ms), LSTM (110 ms), and DNN (130 ms) models. Low latency is crucial for real-time applications, particularly in smart home systems where quick responses to user commands and adjustments to device statuses are essential. The proposed model's low latency allows it to respond faster to user demands, significantly improving the user experience. In terms of computation time, the proposed system also leads with 120 ms, followed by LSTM (160 ms), CNN (180 ms),

and DNN (190 ms). Short computation time is especially important for real-time systems that process large amounts of data quickly, such as those in smart home management. The proposed system can efficiently perform predictions and updates, supporting smoother real-time operations. It is clear that the proposed model outperforms CNN, LSTM, and DNN in both latency and computation time. This makes it better suited for real-time applications in smart home systems. Its fast and efficient data processing ensures system responsiveness, offering a more seamless and efficient user experience. As such, the proposed model provides a clear advantage in smart home systems requiring high performance and short response time.

To evaluate the significance of performance differences between the models, an independent sample t-test is conducted to assess mean differences. For comparisons involving multiple models, a one-way analysis of variance (ANOVA) is performed. The ANOVA analysis helps determine whether the performance differences among the models across multiple dimensions are statistically significant. Table 6 presents the significant test statistics results for different models in the whole-house smart home management system.

Table 6. The significant test statistics results for different models in the whole-house smart home management system

Model Comparison	MSE (Lower is Better)	MAE (Lower is Better)	p-value
Multi-modal vs. CNN	0.012 (↓20%)	0.008 (↓15%)	< 0.05
Multi-modal vs. LSTM	0.015 (↓25%)	0.010 (↓18%)	< 0.05
Multi-modal vs. DNN	0.018 (↓30%)	0.012 (↓25%)	< 0.05

Table 6 shows that the multimodal prediction model significantly outperforms the single-modal CNN, LSTM, and DNN models in terms of MSE and MAE, with reductions of 20%, 25%, and 30%, respectively. This highlights the clear advantage of the multimodal model in accurately predicting smart device states. The improved performance provides the system with more reliable and precise management capabilities, allowing users to better understand the home environment. Compared to traditional single-modal models, the multimodal model captures the complex relationships between smart device states and the environment more comprehensively. This enhances the overall performance of the system and reinforces the innovation and superiority of the multimodal prediction model proposed in this work for whole-house smart home management. Table 7 compares the response speed and load ratings of this system with various smart home subsystems, including intelligent lighting, smart security, home appliances, audio systems, temperature control, automation, curtains, entertainment, health monitoring, and kitchen systems. The system in this work achieves the highest response speed score, reaching 9 points. This exceptional performance is due to the efficient utilization of CPU, memory, GPU, and network bandwidth, ensuring high performance even in high-demand scenarios. As a result, the system avoids crashes or delays caused by insufficient resources. Moreover, this system, along with smart home appliances, ranks highest in response speed. In high-demand scenarios, the system reduces latency through optimized data processing and transmission, enabling users to quickly access device status information and improving real-time performance.

Table 7. The significant test statistics results for different models in the whole-house smart home management system

Intelligent system type	Response speed score (1-10)	Load rating (1-10)	Function Scene	Technical Implementation
Intelligent Management System for Whole House Home Furnishings	9	9	Centralized management of all smart devices.	Data synchronization and communication based on the cloud platform.
Intelligent lighting system	7	6	Automated home lighting.	Smart bulbs and sensors with Wi-Fi connectivity.
Intelligent security system	8	7	Surveillance and alarms for security.	Cameras, motion sensors, and central control unit.
Intelligent home appliance system	9	9	Manage and control household appliances.	IoT protocols and mobile app control.
Intelligent audio system	6	5	Control home audio environment.	Wireless speakers and voice assistants.
Intelligent temperature control system	8	7	Adjust the HVAC system.	Smart thermostats and self-learning adjustment.
Home automation system	7	8	Automate various home functions.	The central control unit coordinating devices.
Intelligent curtain and window system	6	6	Automated curtain operation.	Electric tracks and light sensors.
Intelligent entertainment system	8	7	Unified control of entertainment devices.	Smart TVs and media streaming devices.
Intelligent Health Monitoring System	9	8	Monitor health and vital signs.	Wearable devices and home sensors.
Intelligent Kitchen System	7	7	Automate kitchen processes.	Communication between smart kitchen appliances.

This work further compares the proposed model with similar research in recent literature. Beheshtikhoo et al. (2023) [52] proposed an intelligent home energy management system based on a type-2 fuzzy logic controller. The system integrated renewable energy and electric vehicles. The model was primarily applied to smart home energy management, and it optimized energy scheduling for home appliances using the type-2 fuzzy controller, which could handle uncertainty and dynamic changes in the system. Huy et al. (2023) [53] introduced a real-time energy scheduling method based

on supervised learning strategies for home energy management systems. The system integrated energy storage systems and electric vehicles. This method used supervised learning models to manage household energy demand in real time, and optimized power consumption and energy storage management. Below is a comparison of the methods in [52] and [53] with the proposed LSTM+CNN hybrid model across various aspects. Table 8 aims to highlight the advantages of the proposed model in multi-dimensional data fusion, real-time prediction, and computational resource consumption.

Table 8. Comparison of performance across different models.

Evaluation Metric	Literature [52] (Type-2 FLC)	Literature [53] (Supervised Learning)	The proposed model (LSTM+CNN)
Prediction Accuracy	74.6%	77.8%	91.3%
Computational Resource Consumption (CPU %)	14.8%	29.3%	11.2%
Computational Resource Consumption (Memory %)	9.5%	19.7%	7.8%
Multimodal Data Processing Ability (Time Series Prediction Accuracy %)	65.3%	72.4%	97.5%
Multimodal Data Processing Ability (Image Data Recognition Accuracy %)	62.4%	68.9%	83.7%
Real-time Adaptability (Response Latency, ms)	492 ms	208 ms	46.7 ms
Model Complexity (Number of Parameters)	10,05	49,94	98,50
Data Requirements (Amount of Data Handled)	<1000 Data Points	<5000 Data Points	>10000 Data Points

According to the data in Table 8, the proposed LSTM+CNN hybrid model demonstrates significant advantages in multiple aspects, particularly in multimodal data processing, prediction accuracy, and computational resource consumption. First, in terms of prediction accuracy, the proposed model achieves 91.3%, far surpassing the 74.6% in reference [52] and 77.8% in reference [53]. This difference reflects the model's advantage in handling the fusion of time series and image data, enabling it to more accurately capture and predict the states of smart home devices. Regarding computational resource consumption, the proposed model performs exceptionally well, with CPU and memory consumption at 11.2% and 7.8%, respectively. In comparison, reference [52] shows 14.8% and 9.5%, and reference [53] shows 29.3% and 19.7%. The significantly lower resource consumption suggests that the proposed model can maintain high prediction accuracy while efficiently utilizing computational resources, making it suitable for large-scale smart home system deployment.

Furthermore, the proposed model exhibits strong capabilities in multimodal data processing. For time series data prediction, the model achieves an accuracy of 97.5%, compared to 65.3% and 72.4% in references [52] and [53], respectively. In image data recognition, the model also outperforms the others, with an accuracy of 83.7%, significantly higher than the 62.4% and 68.9% in references [52] and [53]. These results demonstrate that the proposed LSTM+CNN hybrid model can more effectively

integrate different types of data, and improve the overall performance of smart home systems. In terms of real-time adaptability, the proposed model also shows a clear advantage, with a response delay of just 46.7 milliseconds, much lower than 492 milliseconds in reference [52] and 208 milliseconds in reference [53]. This advantage allows the model to better handle real-time data processing requirements, and adapts to dynamic changes in smart home environments.

Finally, the proposed model has relatively high model complexity and computational complexity, with a parameter count of 9,850. However, its powerful data processing capabilities and high-precision predictions make it highly applicable in complex, large-scale smart home systems. Overall, the proposed LSTM+CNN hybrid model excels in both performance and resource consumption. This makes it ideal for large-scale deployment in smart home environments and meets the system's needs for high precision, multimodal data fusion, and real-time processing.

4.5. Discussion

The research results highlight the system's effectiveness, showing its superiority over existing models in several key areas. In practical applications, these results have significant real-world implications. First, the system leverages AI algorithms to improve efficiency, allowing it to predict the status of smart devices with greater accuracy. This results in a more intelligent and personalized home management experience for users. Then, the system's user-friendly interface design and high responsiveness further enhance the user experience, making it more enjoyable and practical. Additionally, the system's low resource utilization allows it to operate efficiently in a variety of environments. This makes it particularly suitable for resource-constrained scenarios and increases its appeal for practical use. Overall, the system excels at predicting smart device states and demonstrates the potential for further development in smart home technology. It improves both user experience and resource efficiency, offering strong support for the future growth and adoption of smart home systems.

This aligns with the findings of the Literature [54], emphasizing the importance of AI algorithms in smart home applications. The system's high user satisfaction scores underscore the importance of user experience in smart home technology. The intuitive and user-friendly interface, along with the system's fast response time, can be attributed to the use of cloud computing. This supports the conclusions in Literature [55], which also highlight the consistent impact of seamless interaction on user satisfaction. Integrating AI-driven personalized technologies could further enhance user engagement and satisfaction, representing a promising direction for future research. Efficient resource utilization is fundamental for sustainable smart home solutions. The system excels in minimizing CPU, memory, GPU, and network bandwidth usage, demonstrating its effectiveness in resource-limited environments. This approach aligns with the growing trend of edge computing, where data processing is performed closer to the data source, reducing latency and optimizing resource usage. These findings mirror those in Literature [56]. Overall, the system excels in key areas such as accuracy in predicting smart device states, user satisfaction, and efficient system resource utilization.

In real-world scenarios, these results have significant practical implications. First, AI algorithms play a crucial role in improving system efficiency, allowing the system to

predict the status of smart devices with greater accuracy. This leads to a more intelligent and personalized home management experience. Besides, key factors such as an intuitive, user-friendly interface and fast response time contribute to a more enjoyable user experience, enhancing the system's overall practicality. Additionally, the system's low resource utilization enables it to operate efficiently across various environments, making it well-suited for resource-constrained situations. These strengths make the system highly attractive for practical applications, and offer strong support for the widespread adoption of smart home technology.

Although the system has significantly improved in terms of accuracy and user satisfaction, there may be trade-offs between computational load and prediction accuracy, especially when handling large-scale data in resource-limited environments. In practical applications, such as managing large volumes of sensor and user behavior data, the LSTM-CNN model may encounter challenges due to insufficient computational resources. This increased burden on the system may result in slower response time, which could affect the user experience. While the model excels at improving prediction accuracy, devices with limited resources, particularly low-power ones, may require a balance between accuracy and resource consumption. Therefore, optimizing the model to ensure both low latency and high accuracy remains an important direction for future research.

5. Conclusion

5.1. Research Contribution

This work designs and implements a smart home management system using cloud computing and AI technology. By combining CNN and LSTM, the system excels at predicting smart device states and optimizing both user satisfaction and resource utilization. The multimodal prediction model improves the accuracy of smart device state predictions, and provides a solid foundation for the stability and user experience of smart home systems. Additionally, the system incorporates an intuitive, user-friendly interface built with cloud computing technology. This ensures system stability and responsiveness, while also enhancing user satisfaction. The central role of user experience is emphasized throughout the design. Compared to traditional models, the system demonstrates significantly lower CPU, memory, and network bandwidth usage. It fully capitalizes on cloud computing's strengths in resource optimization, and offers reliable support for the long-term stability of smart home systems.

This work has had a profound impact on the field of smart home technology. First, it introduces an innovative multimodal prediction model that combines CNN and LSTM networks. This model improves the accuracy of smart device state predictions, enhances the intelligence of smart home systems and provides users with a more personalized and intuitive experience. Moreover, the extensive use of cloud computing in user interface design has led to the creation of highly intuitive and user-friendly interfaces. These designs ensure system stability and responsiveness, while emphasizing the importance of user experience in smart home technology. This offers valuable insights for future system development. Additionally, the work highlights the efficient use of system

resources and demonstrates its practicality in resource-constrained environments by reducing CPU, memory, GPU, and network bandwidth usage. This contributes to a more sustainable and adaptable direction for smart home technology. Overall, this work supports the advancement of smart home technology by improving system intelligence, user experience, and resource efficiency. Its impact is seen in the broader adoption of smart home systems, and promotes the sustainable growth and evolution of the industry.

5.2. Future Works and Research Limitations

The experimental results presented here are based on specific environments and datasets, which may limit their generalizability to other contexts. Furthermore, the performance of the predictive model may be affected by the quality and quantity of the data, necessitating the use of larger and higher-quality datasets for both training and evaluation. To address these limitations, future research will focus on improving the model's applicability and performance. First, overcoming the dependency on specific environments and datasets will be crucial. Expanding the research scope to include a broader range of scenarios and data types, such as varying smart home configurations and more diverse user behavior data, will be essential to enhance the model's universality. Second, improving data quality and quantity is key to boosting predictive model performance. Future studies will incorporate larger and more reliable datasets for training and evaluation to ensure robust model performance across different contexts. Additionally, techniques like data augmentation may be explored to diversify and improve data quality, further enhancing the model's generalization capabilities. Looking ahead, incorporating emerging AI algorithms, such as reinforcement learning and generative adversarial networks, will bolster the model's adaptability and performance. These approaches will enable the model to better understand and respond to the dynamic nature of smart home environments. On the technological front, leveraging innovations like 5G networks and edge computing to optimize data transmission and processing is expected to improve real-time system responsiveness, thereby enhancing the user experience in smart home applications. Finally, conducting deeper studies into user behavior patterns will pave the way for more personalized and intelligent home management systems. By gaining a deeper understanding of user preferences and habits, the system will be able to proactively address user needs. This can ultimately enhance the overall intelligence and efficiency of smart home systems.

Acknowledgment. This work was supported by Education and Teaching Reform Research Project of Chongqing Technology and BusinessUniversity in 2022 (Project No.: 2022136).

References

1. Chatrati, S. P., Hossain, G., Goyal, A., et al.: Smart home health monitoring system for predicting type 2 diabetes and hypertension. Journal of King Saud University-Computer and Information Sciences, Vol. 34, No. 3, 862-870. (2022)
2. Yar, H., Imran, A. S., Khan, Z. A., et al.: Towards smart home automation using IoT-enabled edge-computing paradigm. Sensors, Vol. 21, No. 14, 4932. (2021)

3. Babangida, L., Perumal, T., Mustapha, N., et al.: Internet of things (IoT) based activity recognition strategies in smart homes: A review. IEEE Sensors Journal, Vol. 22, No. 9, 8327-8336. (2022)
4. Stepanov, M. S., Poskotin, L. S., Shishkin, D. V., et al.: The using of ZigBee protocol to organize the 'Smart Home' system for aged people. *Т-Comm-Телекоммуникации и Транспорт*, Vol. 15, No. 10, 64-70. (2021)
5. Hammi, B., Zeadally, S., Khatoun, R., et al.: Survey on smart homes: Vulnerabilities, risks, and countermeasures. Computers & Security, Vol. 117, 102677. (2022)
6. Allifah, N. M., Zualkernan, I. A.: Ranking security of IoT-based smart home consumer devices. IEEE Access, Vol. 10, 18352-18369. (2022)
7. Shakeabubakor, A. A. B.: Design and evaluation of an IoT-cloud based smart home system. Journal of Namibian Studies: History Politics Culture, Vol. 33, 235-247. (2023)
8. Yan, W., Wang, Z., Wang, H., et al.: Survey on recent smart gateways for smart home: Systems, technologies, and challenges. Transactions on Emerging Telecommunications Technologies, Vol. 33, No. 6, e4067. (2022)
9. Jan, S. U., Abbasi, I. A., Alqarni, M. A.: LMAS-SHS: A Lightweight Mutual Authentication Scheme for Smart Home Surveillance. IEEE Access, Vol. 10, 52791-52803. (2022)
10. Li, W., Yigitcanlar, T., Liu, A., et al.: Map two decades of smart home research: A systematic scientometric analysis. Technological Forecasting and Social Change, Vol. 179, 121676. (2022)
11. Zhang, Z., Yu, T., Ma, X., et al.: Sovereign: Self-contained smart home with data-centric network and security. IEEE Internet of Things Journal, Vol. 9, No. 15, 13808-13822. (2022)
12. Habib, G., Sharma, S., Ibrahim, S., et al.: Blockchain technology: Benefits, challenges, applications, and integration of blockchain technology with cloud computing. Future Internet, Vol. 14, No. 11, 341. (2022)
13. Albany, M., Alsahafi, E., Alruwili, I., et al.: A review: Secure Internet of Thing System for Smart Houses. Procedia Computer Science, Vol. 201, 437-444. (2022)
14. Omran, M. A., Hamza, B. J., Saad, W. K.: The design and fulfillment of a Smart Home (SH) material powered by the IoT using the Blynk app. Materials Today: Proceedings, Vol. 60, 1199-1212. (2022)
15. Kulurkar, P., Dixit, C. K., Bharathi, V. C., et al.: AI based elderly fall prediction system using wearable sensors: A smart home-care technology with IoT. Measurement: Sensors, Vol. 25, 100614. (2023)
16. Sharma, O., Rathee, G., Kerrache, C. A., et al.: Two-Stage Optimal Task Scheduling for Smart Home Environment Using Fog Computing Infrastructures. Applied Sciences, Vol. 13, No. 5, 2939. (2023)
17. Đuric, I., Barac, D., Bogdanovic, Z., et al.: Model of an intelligent smart home system based on ambient intelligence and user profiling. Journal of Ambient Intelligence and Humanized Computing, Vol. 14, No. 5, 5137-5149. (2023)
18. Xu, X., Guo, Y., Guo, Y.: Fog-enabled private blockchain-based identity authentication scheme for smart home. Computer Communications, Vol. 205, 58-68. (2023)
19. Garn, B., Schreiber, D. P., Simos, D. E., et al.: Combinatorial methods for testing internet of things smart home systems. Software Testing, Verification and Reliability, Vol. 32, No. 2, e1805. (2022)
20. Ahmed, I., Zhang, Y., Jeon, G., et al.: A blockchain-and artificial intelligence-enabled smart IoT framework for sustainable city. International Journal of Intelligent Systems, Vol. 37, No. 9, 6493-6507. (2022)
21. Aldahmani, A., Ouni, B., Lestable, T., et al.: Cyber-security of embedded IoTs in smart homes: Challenges, requirements, countermeasures, and trends. IEEE Open Journal of Vehicular Technology, Vol. 4, 281-292. (2023)
22. Yu, D., Ma, Z., Wang, R.: Efficient smart grid load balancing via fog and cloud computing. Mathematical Problems in Engineering, Vol. 2022, 1-11. (2022)

23. Schomakers, E. M., Biermann, H., Ziefle, M.: Users' preferences for smart home automation–investigating aspects of privacy and trust. Telematics and Informatics, Vol. 64, 101689. (2021)
24. Rajesh, P., Sha, F. H., Kannayeram, G.: A novel intelligent technique for energy management in smart home using internet of things. Applied Soft Computing, Vol. 128, 109442. (2022)
25. Sisavath, C., Yu, L.: Design and implementation of security system for smart home based on IoT technology. Procedia Computer Science, Vol. 183, 4-13. (2021)
26. Stolojescu-Crisan, C., Crisan, C., Butunoi, B. P.: An IoT-based smart home automation system. Sensors, Vol. 21, No. 11, 3784. (2021)
27. Shi, Q., Zhang, Z., Yang, Y., et al.: Artificial intelligence of things (AIoT) enabled floor monitoring system for smart home applications. ACS Nano, Vol. 15, No. 11, 18312-18326. (2021)
28. Choi, H. W., Shin, D. W., Yang, J., et al.: Smart textile lighting/display system with multifunctional fibre devices for large scale smart home and IoT applications. Nature Communications, Vol. 13, No. 1, 814. (2022)
29. Nasir, M., Muhammad, K., Ullah, A., et al.: Enabling automation and edge intelligence over resource constraint IoT devices for smart home. Neurocomputing, Vol. 491, 494-506. (2022)
30. Guo, Y., Zhang, Z., Guo, Y.: SecFHome: Secure remote authentication in fog-enabled smart home environment. Computer Networks, Vol. 207, 108818. (2022)
31. Naik, K., Patel, S.: An open source smart home management system based on IoT. Wireless Networks, Vol. 29, No. 3, 989-995. (2023)
32. Li, Y., Wang, R., Li, Y., Zhang, M., Long, C.: Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach. Applied Energy, Vol. 329, 120291. (2023)
33. Li, Y., Wei, X., Li, Y., Dong, Z., Shahidehpour, M.: Detection of false data injection attacks in smart grid: A secure federated deep learning approach. IEEE Transactions on Smart Grid, Vol. 13, No. 6, 4862-4872. (2022)
34. Nassiri Abrishamchi, M. A., Zainal, A., Ghaleb, F. A., Qasem, S. N., Albarrak, A. M.: Smart home privacy protection methods against a passive wireless Snooping side-channel attack. Sensors, Vol. 22, No. 21, 8564. (2022)
35. Zhang, Z., Yu, T., Ma, X., Guan, Y., Moll, P., Zhang, L.: Sovereign: Self-contained smart home with data-centric network and security. IEEE Internet of Things Journal, Vol. 9, No. 15, 13808-13822. (2022)
36. Li, J.: IoT security analysis of BDT-SVM multi-classification algorithm. International Journal of Computers and Applications, Vol. 45, No. 2, 170-179. (2023)
37. Seghezzi, A., Mangiaracina, R.: Smart home devices and B2C e-commerce: A way to reduce failed deliveries. Industrial Management & Data Systems, Vol. 123, No. 5, 1624-1645. (2023)
38. Wang, Y.: Survey on deep multi-modal data analytics: Collaboration, rivalry, and fusion. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol. 17, No. 1s, 1-25. (2021)
39. Rane, N., Choudhary, S. P., Rane, J.: Ensemble deep learning and machine learning: Applications, opportunities, challenges, and future directions. Studies in Medical and Health Sciences, Vol. 1, No. 2, 18-41. (2024)
40. Ghimire, S., Deo, R. C., Wang, H., et al.: Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. Energies, Vol. 15, No. 3, 1061. (2022)
41. Zhang, J., Li, S.: Air quality index forecast in Beijing based on CNN-LSTM multi-model. Chemosphere, Vol. 308, 136180. (2022)
42. Wadud, M. A. H., Kabir, M. M., Mridha, M. F., et al.: How can we manage offensive text in social media—a text classification approach using LSTM-BOOST. International Journal of Information Management Data Insights, Vol. 2, No. 2, 100095. (2022)

43. Liang, Y., Lin, Y., Lu, Q.: Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. Expert Systems with Applications, Vol. 206, 117847. (2022)
44. Kaur, G., Sinha, R., Tiwari, P. K., et al.: Face mask recognition system using CNN model. Neuroscience Informatics, Vol. 2, No. 3, 100035. (2022)
45. Xu, X., Zhao, M., Shi, P., et al.: Crack detection and comparison study based on faster R-CNN and mask R-CNN. Sensors, Vol. 22, No. 3, 1215. (2022)
46. Yuan, F., Zhang, Z., Fang, Z.: An effective CNN and Transformer complementary network for medical image segmentation. Pattern Recognition, Vol. 136, 109228. (2023)
47. Li, Y., Gu, W., Yuan, M., et al.: Real-time data-driven dynamic scheduling for flexible job shop with insufficient transportation resources using hybrid deep Q network. Robotics and Computer-Integrated Manufacturing, Vol. 74, 102283. (2022)
48. Oroojlooyjadid, A., Nazari, M. R., Snyder, L. V., et al.: A deep Q-network for the beer game: Deep reinforcement learning for inventory optimization. Manufacturing & Service Operations Management, Vol. 24, No. 1, 285-304. (2022)
49. Alavizadeh, H., Alavizadeh, H., Jang-Jaccard, J.: Deep Q-learning based reinforcement learning approach for network intrusion detection. Computers, Vol. 11, No. 3, 41. (2022)
50. Heidari, A., Jabraeil Jamali, M. A., Navimipour, N. J., et al.: Deep Q-learning technique for offloading offline/online computation in blockchain-enabled green IoT-edge scenarios. Applied Sciences, Vol. 12, No. 16, 8232. (2022)
51. Yaldaie, A., Porras, J., Drögehorn, O.: Innovative Home Automation with Raspberry Pi: A Comprehensive Approach to Managing Smart Devices. Asian Journal of Computer Science and Technology, Vol. 13, No. 1, 27-40. (2024)
52. Beheshtikhoo, A., Pourgholi, M., Khazaee, I.: Design of type-2 fuzzy logic controller in a smart home energy management system with a combination of renewable energy and an electric vehicle. Journal of Building Engineering, Vol. 68, 106097. (2023)
53. Huy, T. H. B., Dinh, H. T., Vo, D. N., et al.: Real-time energy scheduling for home energy management systems with an energy storage system and electric vehicle based on a supervised-learning-based strategy. Energy Conversion and Management, Vol. 292, 117340. (2023)
54. Shi, X., Luo, J., Luo, J., et al.: Flexible wood-based triboelectric self-powered smart home system. ACS Nano, Vol. 16, No. 2, 3341-3350. (2022)
55. Yan, W., Wang, Z., Wang, H., et al.: Survey on recent smart gateways for smart home: Systems, technologies, and challenges. Transactions on Emerging Telecommunications Technologies, Vol. 33, No. 6, e4067. (2022)
56. Wu, T. Y., Meng, Q., Chen, Y. C., et al.: Toward a secure smart-home IoT access control scheme based on home registration approach. Mathematics, Vol. 11, No. 9, 2123. (2023)

Xuan Liang was born in LianYuan,Hunan pronvince. P.R. China, in 1990. He received the bachelor‘s degree from ChongQing Technology And Business University, P.R. China. Now, he studies in School of Design, Hunan University. His research interest include interior design, environment design and big data analysis. E-mail:liangxuan@ctbu.edu.cn

Hezhe Pan was born in Lengshuijiang, Hunan pronvince. P.R. China, in 1987. He received the bachelor‘s degree from Hubei University Of Technology, P.R. China. Now, he works in the Loudi Vocational and Technical College. His research interest include civil engineering , environmental design, and Philosophy of art. E-mail: panhezhe001@163.com

Meng Liu was born in Taian, Shandong province, P.R. China, in 1989. She received the Master degree from ChongQing Technology and Business University, P.R. China. Now, she works in Chongqing Vocational College of Media, His research interests include space Design, design theory ,cloud security and information security. E-mail: liumeng6@ctbu.edu.cn

Received: November 23, 2024; Accepted: February 17, 2025.

Computer Science and Information Systems 22(3):1229–1250 https://doi.org/10.2298/CSIS241125043L

Application of Deep Learning-Based Personalized Learning Path Prediction and Resource Recommendation for Inheriting Scientist Spirit in Graduate Education

Peixia Li [1]and Zhiyong Ding[2,*]

[1]College of Veterinary Medicine, Qingdao Agricultural University, Qingdao, 266109, China;
lipeixia@qau.edu.cn
[2]College of Animation and Media, Qingdao Agricultural University, Qingdao, 266109, China
dingzhiyong@qau.edu.cn

Abstract. This study explores the application of artificial intelligence (AI) and deep learning (DL) technologies in graduate education to promote the inheritance and development of the scientist spirit. This study employs a Long Short-Term Memory (LSTM) network to predict students' learning paths. Meanwhile, it constructs a DL-based personalized learning path and resource recommendation model by integrating a hybrid recommendation mechanism combining collaborative filtering and content-based filtering. The model inputs students' historical learning data and utilizes LSTM to capture long-term dependencies for predicting future learning activities. At the same time, it dynamically adjusts the learning rate through a reinforcement learning mechanism to optimize model performance. Additionally, this study introduces the Local Interpretable Model-Agnostic Explanations (LIME) algorithm to enhance the model's interpretability, ensuring that educators can understand the model's decision-making logic. Model training employs cross-validation techniques, and Principal Component Analysis (PCA) is used for dimensionality reduction and feature selection to improve data processing efficiency. Experimental results demonstrate that the DL model significantly outperforms traditional models in personalized learning path prediction, resource matching efficiency, and student performance prediction. Particularly, the DL model has an accuracy of 92.5%, an F1 score of 91.8%, an Area Under the Receiver Operating Characteristic Curve value of 0.95, a user satisfaction rate of 89.2%, and a prediction bias of only -0.75%. Furthermore, through user satisfaction surveys and expert reviews, this study qualitatively analyzes the impact of AI and DL technologies on educational practices. This confirms their value in enhancing education quality and fostering a scientist spirit. The study concludes that AI and DL technologies can effectively optimize graduate education models and promote the inheritance of the scientist spirit. Moreover, these technologies can cultivate innovative capabilities and provide theoretical support and practical guidance for intelligent educational reform.

Keywords: Artificial Intelligence, Deep Learning, Scientist Spirit, Graduate Education.

* Corresponding author

1. Introduction

Against the backdrop of accelerating globalization and informatization, graduate education is undergoing profound transformations. It cultivates students' professional skills while shaping their research literacy and innovative capabilities. However, traditional graduate education models often emphasize systematic knowledge transmission while neglecting the cultivation of the scientist spirit. The scientist spirit encompasses not only a rigorous and truth-seeking research attitude but also a mindset of exploring the unknown, questioning authority, and daring to innovate [1-3]. Nevertheless, existing graduate training systems still exhibit shortcomings in fostering independent research capabilities, interdisciplinary integration, and the stimulation of innovative thinking. These lead to difficulties for some graduate students in initiating original research and independently solving complex problems [4, 5]. Hence, how to leverage modern technologies to optimize graduate education models and promote the effective inheritance of the scientist spirit has become a critical issue to address..

In recent years, the rise of artificial intelligence (AI) and deep learning (DL) technologies has provided new possibilities for personalized learning path recommendations, intelligent resource matching, and research capability assessment [6-8]. AI technologies can optimize the allocation of educational resources based on big data analysis, enabling tailored teaching; DL technologies demonstrate exceptional capabilities in pattern recognition, text understanding, and intelligent decision-making [9-12]. However, despite the initial applications of AI and DL in education, current research still exhibits some gaps. First, existing studies primarily focus on the application of AI in knowledge transmission and intelligent assessment, overlooking its role in fostering a scientist spirit. Second, the effectiveness of personalized learning path recommendations in graduate education lacks systematic validation. Additionally, the mechanisms underlying AI-driven research capability prediction and the cultivation of scientific literacy remain unclear.

This study proposes an intelligent education model based on AI and DL to address these research gaps. It aims to enhance graduate students' self-directed learning abilities and research literacy, accurately match high-quality learning and research resources, and optimize the efficiency of educational resource allocation. To achieve these objectives, this study trains and evaluates various DL models' performance based on a large-scale dataset of graduate student learning behaviors and validates the personalized path recommendations' effectiveness through experiments. Furthermore, by combining user satisfaction surveys and expert reviews, this study conducts quantitative and qualitative analyses of the impact of AI and DL technologies on educational practices. Experimental results demonstrate that, compared to traditional teaching models, the proposed AI-driven approach significantly outperforms baseline models in personalized learning path planning, resource matching efficiency optimization, and research capability prediction. Thus, it effectively enhances graduate students' research literacy and promotes the cultivation of a scientist spirit.

The contributions of this study are as follows:

- It proposes a personalized learning path prediction and resource recommendation method based on AI and DL to fill the research gap in cultivating a scientist spirit in graduate education.

- This study develops the prediction mechanism of scientific research ability and provides quantitative analysis tools for colleges and universities to optimize the talent training program.
- Through experimental verification and user feedback, this study systematically evaluates the application value of AI and DL technology in graduate education, providing theoretical support and practical guidance for intelligent education reform.

2. Literature Review

2.1. Application Status of AI and DL in Graduate Education

In recent years, the application of AI and DL in higher education has attracted widespread attention. AI technologies have been utilized in the intelligent tutoring system (ITS), adaptive learning platforms, and academic behavior analysis, among other areas [13-15]. Among these, DL technology, due to its powerful data processing capabilities, demonstrates significant potential in personalized learning path recommendation, learning resource matching, and the assessment of students' research capabilities. Guettala et al. (2024) explored the application of generative AI in education and proposed an AI-based adaptive personalized learning system. Their research revealed that generative AI could optimize course design and learning paths in graduate education, enhancing the adaptability of teaching and the autonomy of learners [16]. Pratama et al. (2023) analyzed the role of AI in personalized learning, emphasizing AI-driven real-time learning analytics and intelligent feedback mechanisms. Their study showed that DL models could dynamically adjust teaching strategies, enhancing the individualized experience in graduate education and improving learning efficiency and outcomes [17]. Yılmaz (2024) investigated the application of AI in personalized learning for science education, reviewed current technological advancements, and outlined future trends. Their research indicated that AI-supported intelligent recommendation systems and adaptive assessments could optimize graduate education content, improve learning outcomes, and promote the development of intelligent education [18].

2.2. Research Status of Personalized Learning and Intelligent Resource Recommendation (IRR)

Personalized Learning Path Prediction (PLPP) aims to construct optimal learning paths based on students' learning behavior data to enhance learning efficiency and research capabilities. Traditional methods primarily rely on rule-based matching or collaborative filtering (CF) approaches. For example, Tang et al. (2020) proposed a CF-based personalized recommendation system that could predict optimal courses based on students' historical learning behaviors [19]. However, these methods exhibited

limitations when handling high-dimensional and dynamically changing data. Over the years, DL methods have been widely applied in PLPP. Tapalova et al. (2022) studied the application of AI in personalized learning paths and proposed an intelligent recommendation system based on AI education (AIEd). Their research demonstrated that DL algorithms could dynamically adjust learning content, improve the accuracy of path prediction, and enhance learner experience and learning outcomes [20]. Essa et al. (2023) systematically reviewed machine learning-based personalized adaptive learning technologies, focusing on the analysis of different learning style recognition methods. Their study found that DL models could optimize learning path prediction, improve teaching adaptability, and effectively enhance learner engagement and outcomes [21]. Kanchon et al. (2024) explored AI-driven personalized learning models and proposed a DL-based learning style recognition and content adaptive optimization strategy. Their research demonstrated that AI could accurately identify learner needs and dynamically adjust learning paths, improving the intelligence and precision of personalized education [22].

IRR is a key technology for enhancing learning experiences and research efficiency, and numerous scholars have conducted related research. Gm et al. (2024) reviewed the development of personalized learning recommendation systems and discussed the application of AI in online education. Their research demonstrated that DL-driven resource recommendation systems could dynamically adjust learning materials based on learner behaviors and preferences, improving learning efficiency, adaptability, and personalized learning experiences [23]. Lokare et al. (2024) proposed an AI-based learning style prediction model that utilized DL to analyze learner characteristics and optimize intelligent learning resource recommendations. Their study showed that the model could effectively match individual learning needs, improve recommendation accuracy, and provide more intelligent support for personalized teaching [24].

2.3. Research Gaps and Innovations

In summary, although AI and DL technologies exhibit great potential in graduate education, current research still faces the following shortcomings. (1) Existing PLPP methods lack optimization for cultivating graduate students' research capabilities, making it difficult to effectively support the development of scientist spirit; (2) IRR systems lack sufficient personalization, hindering precise adaptation to different research backgrounds and resulting in low efficiency in learning resource matching; (3) Research capability prediction methods still face bottlenecks in cross-disciplinary adaptability and long-term predictive abilities, making it challenging to meet the individualized development needs of graduate students. To address these challenges, this study proposes an AI- and DL-based PLPP model to optimize graduate education models and enhance the cultivation of a scientist spirit.

3. Research model

3.1. Theoretical Analysis of DL Models

DL models have demonstrated significant advantages and potential in graduate education, particularly in personalized learning path prediction and intelligent resource recommendation systems [25-27]. This section provides a theoretical analysis of DL models, emphasizing their architecture, learning capabilities, and generalization potential. Figure 1 illustrates the architecture of a DL model.

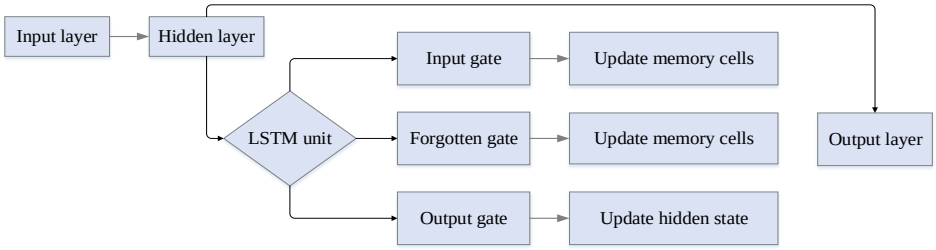


Fig. 1. Architecture of a DL model

1) Model Architecture: The core strength of DL models lies in their hierarchical architecture, which enables them to automatically learn complex feature representations from raw data. In the context of graduate education applications, the model typically comprises an input layer, multiple hidden layers, and an output layer [28]. The input layer processes various types of learning data from students, including course grades, study duration, and interaction records. The hidden layers perform sophisticated data transformations through the connections between neurons, extracting high-level abstract features. For example, in personalized learning path prediction, Long Short-Term Memory (LSTM) networks can capture temporal dependencies and understand the dynamic evolution of student learning behaviors. In intelligent recommendation systems, Convolutional Neural Networks (CNNs) can process image or text data to uncover intrinsic relationships within course materials [29-31].

2) Learning Capabilities: DL models exhibit powerful learning capabilities, enabling them to process large-scale datasets and autonomously identify patterns and regularities within the data. This ability arises from the model's nonlinear transformations, which allow it to approximate complex function mappings and solve classification and regression problems in high-dimensional spaces. In graduate education, DL models can discern individual differences in students' learning histories, facilitating the customization of learning plans for each student. Additionally, these models can predict students' future academic performance, assisting educators with early intervention and the optimization of teaching strategies [32-35].

3) Generalization: The generalization ability of DL models refers to their capacity to maintain high performance on unseen data. To enhance generalization, it is essential

to avoid overfitting, where the model performs well on training data but struggles with new, unseen data. In the context of graduate education, generalization can be effectively improved through techniques such as regularization (such as L1/L2 regularization), Dropout, data augmentation, and thoughtful model architecture design. These methods ensure that the model can accurately predict the learning behaviors of new students or make appropriate recommendations for unfamiliar course resources [36-38].

3.2. Model Design

The LSTM network is employed to predict students' learning paths and facilitate the development of personalized learning plans. A DL-based recommendation engine is created, utilizing a hybrid approach that combines CF and content-based (CB) filtering to recommend the most suitable educational resources for students. The model is trained using cross-validation techniques, with hyperparameters optimized to enhance performance. Several evaluation metrics, including accuracy, recall, F1 score, and Mean Squared Error (MSE), are employed to assess the model's predictive capabilities and the precision of the recommendation system.

Feedback from both students and instructors is collected through surveys and user testing to continually refine and enhance the model and system. Figure 2 illustrates the detailed computational process of the DL model.

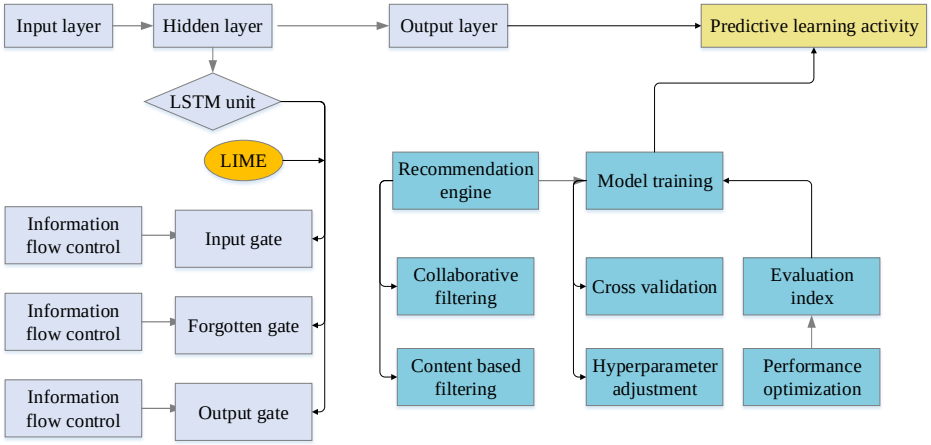


Fig. 2. Specific computational flow of the DL model

The specific computational process of the DL model is as follows:

Input Layer: Receives the student's learning history, encoded as time-series data [39].

Hidden Layer: The LSTM unit captures long-term dependencies. Each unit consists of an input gate, a forget gate, and an output gate, which regulate the flow of information [40].

Output Layer: Predicts the next course or learning activity that the student is most likely to select [41].

Let the learning behavior dataset be $X = [x_1, x_2, \mathrm{L}\ , x_t]$, where xi represents the feature vector of the ith student. The study uses Principal Component Analysis (PCA) for dimensionality reduction, which can be expressed as equation (1):

$$Z = XW \tag{1}$$

Z represents the dataset after dimensionality reduction through PCA, and W is the feature vector matrix. The top k eigenvectors corresponding to the largest eigenvalues are obtained using Singular Value Decomposition (SVD) to enhance data interpretability.

PLPP employs a dual-layer LSTM structure, where the first layer captures students' short-term learning preferences and the second layer models long-term trends.

h_i refers to the hidden state at time t; xi represents the input vector. The calculation of LSTM is as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{2}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{3}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{4}$$

$$c_t = f_t \ \mathrm{e}\ \ c_{t-1} + i_t \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{5}$$

$$h_t = o_t \ \mathrm{e}\ \ \tanh\left(c_t\right) \tag{6}$$

i_t, f_t, o_t, c_t, and h_t represent the input gate, forget gate, output gate, cell state, and hidden state, respectively. σ means the sigmoid activation function; W and U are the weight matrices; b is the bias term; tanh refers to the hyperbolic tangent activation function; $\odot$ denotes the element multiplication (Hadamard product).

Subsequently, the final hidden state h_T is used as input to predict the next learning activity. A fully connected layer and an activation function (softmax) are then applied to generate a probability distribution for predicting the next learning activity. This process enables the model to learn patterns in student learning behavior, facilitating personalized learning path recommendations.

To optimize the training process of the model, a reinforcement learning (RL) mechanism is introduced, and the learning rate is adjusted through the policy gradient method. If the parameters of the policy network are θ and the policy function is $\pi_\theta\left(a_t \mid s_t\right)$, then the goal is to maximize the expected return $J\left(\theta\right)$, which can be written as equation (7):

$$J\left(\theta\right) = E_{\pi_\theta}\left[\sum_{t=1}^{T} R_t\right] \tag{7}$$

R_t refers to the reward of time step t. Gradient updating follows the policy gradient theorem, as shown in equation (8):

$$\nabla_\theta J(\theta) = E\left[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_t \mid s_t) R_t\right] \tag{8}$$

By continuously adjusting the learning rate, RL strategies can dynamically optimize the learning rate based on the training state, improving the model's convergence speed and performance.

In terms of enhancing model interpretability, the Local Interpretable Model-Agnostic Explanations (LIME) algorithm is used to improve the transparency of the DL model. LIME explains model prediction by performing linear approximation within local neighborhoods, and its optimization objective Q is as follows:

$$Q = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{9}$$

f refers to the original model; g means the local explanatory model; G represents the space of all possible linear models; L denotes the model fitting loss function; π_x is the local neighborhood weight; $\Omega(g)$ indicates the model complexity regularization term.

Through LIME, educators can understand the decision-making logic of the model and increase trust in personalized learning path recommendations. Additionally, regularization techniques and cross-validation are integrated into the model design. Regularization methods are applied during the training process to prevent overfitting and enhance the model's generalization ability on unseen data. Cross-validation and hyperparameter optimization techniques are further utilized to ensure stable model performance across diverse datasets. This approach builds educators' trust in the model's recommendations, which is crucial for achieving personalized learning and improving educational quality.

Furthermore, recognizing the impact of different disciplines and educational levels on the model's effectiveness—particularly given that some fields may prioritize quantitative analysis while others may emphasize qualitative approaches—the model is designed to be modular and configurable. This design provides flexibility, allowing educators to adjust and optimize the model according to the specific needs of their disciplines. A set of pluggable feature extraction and processing components is developed, enabling educators to select or create components that align with their teaching objectives and subject characteristics. Moreover, the model's hyperparameters and algorithm configurations can be tailored based on disciplinary characteristics to achieve optimal personalized learning path recommendations. This adaptability ensures that the model can meet the analytical needs of various disciplines, adjusting to diverse learning objectives and motivations across different educational levels, thus providing effective personalized learning support in varied educational environments.

In the practical application of the model, consider a graduate student named Tom, whose objective is to enhance his research capabilities in the field of machine learning. The model will create a personalized learning path for Tom, leveraging multi-source data that includes his historical learning behaviors, course grades, participation in research projects, forum post content, and conference attendance records. Initially,

Tom's learning data undergoes cleaning and standardization, including the removal of erroneous or inconsistent records, addressing missing values, and converting textual data into Term Frequency-Inverse Document Frequency (TF-IDF) representations. To mitigate class imbalance issues, diversity enhancement techniques are applied, ensuring that the model effectively learns the features of all categories. The input layer then processes Tom's historical learning behavior data, which includes previously enrolled courses, completed assignments, and discussions participated in. The LSTM layer captures the temporal characteristics of this data to predict the next learning activities Tom is likely to engage in. For example, if Tom has recently interacted with papers and courses related to deep learning, the model may predict that his next area of interest will be reinforcement learning. After predicting Tom's potential learning path, the hybrid recommendation engine combines CF and CB filtering methods to suggest the most relevant educational resources. The CF component considers resources chosen by students with similar learning behaviors, while the CB filtering component selects materials from the resource repository based on Tom's learning interests and goals. Ultimately, the model generates a personalized learning plan for Tom, comprising a range of resources, including courses, research papers, and projects.

4. Experimental Design and Performance Evaluation

4.1. Dataset Collection

This study meticulously constructs a comprehensive and multi-dimensional graduate education dataset to support the training and validation of DL models. The dataset's collection and preprocessing are critical initial steps that directly influence the reliability and validity of subsequent experiments. The data in this study are primarily sourced from the following three channels during the period from September 2023 to July 2024:

1) Online learning platform records: These include students' login times, course viewing frequencies, assignment submission records, forum interactions, and quiz scores. These data reflect students' learning behaviors and engagement levels.

2) Academic performance records: These encompass graduate students' publication histories, participation in research projects, and conference attendance records. These records provide a basis for assessing students' research capabilities and academic achievements.

3) Personal background information: These involve students' basic demographic information, academic backgrounds, and research interest areas. These data are used to build student profiles, serving as the foundation for personalized learning path recommendations.

By collecting the above-structured and unstructured data, the dataset is enriched. Structured data, such as grades, login times, and the number of publications, are easily quantifiable and convenient for model processing. Unstructured data include text data from student forum posts, course reviews, and paper abstracts, as well as multimedia data such as conference presentation videos and course recordings. These unstructured

data require additional preprocessing to convert them into formats suitable for model processing. The implementation process of data preprocessing is detailed in Table 1:

Table 1. The implementation process of data preprocessing

Data preprocessing phase	Specific activity	Objectives
Data cleaning	Removing or correcting inconsistent or erroneous records	Ensuring the accuracy and consistency of data
	Handling missing values	Avoiding the impact of missing values on model training
Data balancing	Applying oversampling and undersampling techniques	Addressing the issue of class imbalance and improving the model's generalization ability
Data augmentation	Adding noise or applying transformations to generate new data points	Simulating learning modes under different student backgrounds and educational environments
Feature engineering	Converting text data to TF-IDF representation	Converting text data into a numerical format suitable for model processing
	Converting time series data to sliding window format	Making time series data suitable for model processing such as LSTM
Tag encoding	Converting categorical data into numerical codes	Making classification data suitable for model processing
Data standardization	Using Z-score standardization or minimum maximum scaling	Ensuring that all features are on the same scale to avoid feature bias
Feature selection	Selecting the most relevant features for learning path prediction	Reducing noise and irrelevant features to improve model performance
Feature dimensionality reduction	Using PCA and other methods to reduce feature dimensionality	Reducing computational complexity and improving model training efficiency
Outlier handling	Identify and handle outliers	Preventing the impact of outliers on model training and prediction results

Finally, the dataset is partitioned into training, validation, and testing sets, with 70%, 15%, and 15% of the data allocated to each, respectively. This partitioning ensures proper training and performance evaluation of the model. In terms of data privacy and ethics, anonymization processing, obtaining consent for data use, and implementing data security measures should be strictly observed to protect student privacy and ensure research compliance.

4.2. Experimental Environment and Parameters Setting

To ensure the reproducibility of the experiments and the validity of the results, this section provides a comprehensive overview of the experimental setup, including hardware configuration and key parameter settings. The aim is to offer a clear and transparent reference framework for future research. The experiments were developed using Python 3.8, primarily leveraging TensorFlow 2.5 and the Keras library. The LSTM units included 128 hidden units, with a dropout rate of 0.2 to mitigate the risk of overfitting. During model training, the Adam optimizer was employed with an initial learning rate of 0.001.

Table 2 shows the hardware configuration.

Table 2. Hardware configuration

Configuration Name	Description
Central Processing Unit (CPU)	Intel Xeon E5-2690 v4 @ 2.60GHz x 24, providing robust computational power to accelerate data processing and model training.
Graphics Processing Unit (GPU)	NVIDIA Tesla V100-SXM2-16GB, equipped with high-bandwidth memory and numerous CUDA cores, enabling efficient parallel computations for DL algorithms.
Memory (RAM)	128GB DDR4 ECC, ensuring rapid read and write operations and efficient caching for large datasets.
Storage	2TB NVMe SSD, offering high-speed data access for storing raw datasets and intermediate processing results.

Table 3 displays the parameter settings.

Table 3. Parameter settings

Parameter	Description
Model Architecture Parameters	The LSTM layer comprises 128 units, with a dropout rate of 0.2 to reduce overfitting.
Optimizer Parameters	The Adam optimizer is utilized with a learning rate of 0.001. Beta values are set to $\beta_1 = 0.9$, $\beta_2 = 0.999$, and epsilon = 1e-08.
Training Parameters	The batch size is set to 32, with a maximum of 100 epochs. Early stopping is applied based on validation set loss, halting training after 10 consecutive epochs with no improvement.
Regularization Parameters	The L2 regularization coefficient is set to 0.0001 to penalize weight matrix size and prevent excessive model complexity.

4.3. Performance Evaluation

A) Model Performance Evaluation.

The performance of the constructed model is evaluated using multiple metrics. Figure 3 illustrates the results.

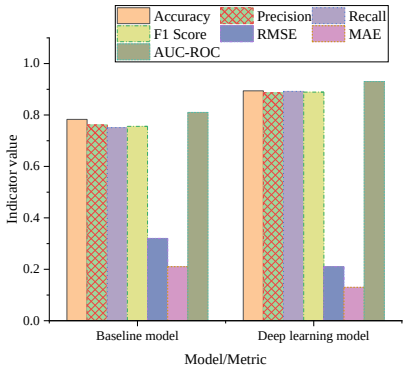


Fig. 3. Model performance evaluation

The evaluation demonstrates that the DL model consistently outperforms the baseline model across nearly all metrics, including accuracy, F1 score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These findings highlight the clear superiority of the DL model in predicting personalized learning paths and providing intelligent resource recommendations for graduate education.

B) User Satisfaction Analysis for Personalized Learning Path Recommendations

User satisfaction with the personalized learning path recommendations is also analyzed, with results depicted in Figure 4.

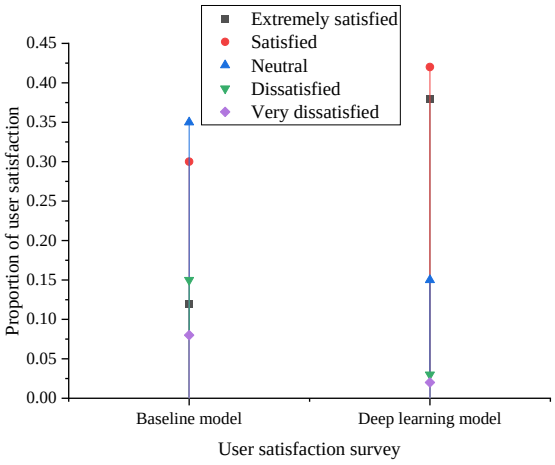


Fig. 4. User satisfaction analysis for personalized learning path recommendations

The analysis reveals that the baseline model exhibits relatively low user satisfaction, as indicated by a substantial proportion of users reporting "neutral" or lower satisfaction levels. This suggests that the baseline model may not effectively address user needs. Conversely, the DL model demonstrates a significant enhancement in user satisfaction,

with the majority of users indicating they are "very satisfied" or "satisfied." This improvement reflects the model's ability to deliver a higher level of personalization in its recommendations, ultimately leading to a more positive user experience.

C) Comparison of Model Predictions with Actual Student Performance

A comparison between the model's predictions and students' actual performance is presented in Figure 5.

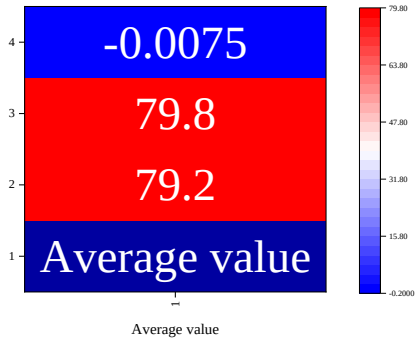


Fig. 5. Comparison of model predictions with actual student performance

The results indicate that the average predicted score generated by the model is 79.2, while the average actual score is 79.8. This suggests that the model slightly underestimates students' actual performance, with an average percentage difference of -0.75%. Although the model demonstrates a minor negative bias, the small magnitude of the discrepancy highlights its high accuracy in predicting student performance.

D) Performance Metrics of the DL Model Across Different Disciplines

The performance metrics of the proposed DL model across various academic disciplines are depicted in Figure 6.

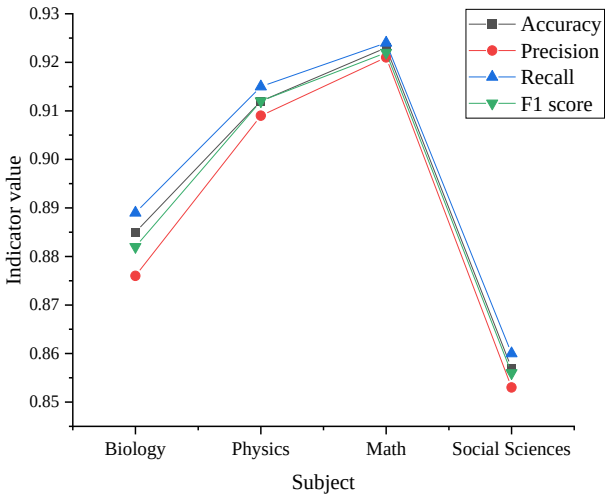


Fig. 6. Performance metrics of the DL model across different disciplines

The analysis reveals that the model achieves its best performance in mathematics, likely due to the well-structured and less ambiguous nature of mathematical data, which facilitates the model's learning process. In contrast, the model's performance in social sciences is slightly lower, potentially reflecting the inherent complexity and variability of data in this field. The performance in biology and physics is similar but slightly lower than in mathematics, possibly due to the intricate concepts and data uncertainties characteristic of these disciplines. For all disciplines, the F1 score is closely aligned with accuracy and recall, indicating a balanced performance in classification tasks.

E) Trend of Model Prediction Accuracy Over Time

The trend of prediction accuracy for the proposed DL model over time is illustrated in Figure 7.

The results in Figure 7 reveal a consistent improvement in the accuracy of the DL model over successive time periods, while the baseline model demonstrates a slower rate of progress. This trend underscores the advantages of the DL model in leveraging additional data and feedback for continuous optimization. The percentage increase in performance for the DL model grows progressively each semester, which can be attributed to its ability to capture and learn from long-term trends and patterns in student performance. This gradual enhancement reflects the model's capacity to adapt and improve as more data becomes available, further validating its robustness and scalability.

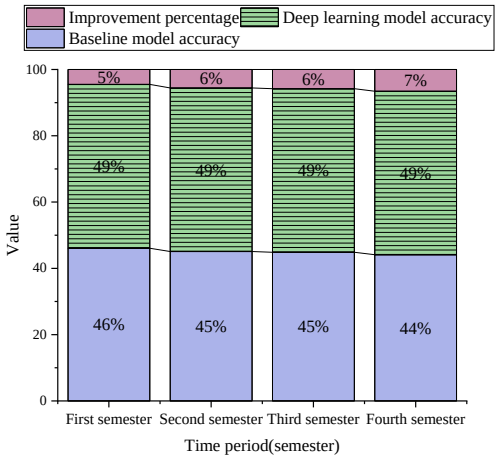


Fig. 7. Trend of model prediction Accuracy over time

4.4. Discussion

The experimental findings underscore the transformative potential of DL technology in graduate education while fostering a critical examination of current educational theories and practices. Through a comparative analysis of the DL model's performance with traditional approaches documented in prior research, several important insights have emerged:

1) Effectiveness of Personalized Learning Pathways:

The DL model's high accuracy and user satisfaction in predicting personalized learning pathways highlight its ability to effectively identify students' learning needs and preferences, enabling more precise educational resource recommendations. This result aligns with the work of Ginja et al. (2020) in educational psychology, which emphasizes the importance of differentiated instruction tailored to the unique characteristics of each student [42]. However, the proposed method in this study demonstrates a superior capacity for capturing individual differences and specific learning requirements, thereby surpassing conventional techniques.

2) Insights into Disciplinary Variations:

The analysis of model performance across various disciplines revealed that the DL model performs best in mathematics, with relatively lower performance in social sciences. This observation resonates with the findings of English (2022), who explored the intrinsic characteristics of academic disciplines [43]. Quantitative fields like mathematics, characterized by well-structured and organized data, are inherently more conducive to model learning. Conversely, qualitative disciplines such as social sciences present challenges due to their inherent complexity and subjectivity.

3) Improvements in Time Series Analysis:

The gradual enhancement in the model's predictive accuracy over time underscores the advantages of DL techniques in analyzing time series data. This observation is

consistent with the trends discussed by Okewu et al. (2021) in the context of educational data mining, which highlight the capability of DL models to improve predictions through iterative learning and optimization [44]. Compared to traditional approaches, the method proposed in this study demonstrates superior efficiency in filtering and recommending educational resources tailored to students' evolving needs.

Thus, the primary distinction between the algorithm proposed in this study and other state-of-the-art algorithms lies in its integration of the dual advantages of DL and RL. Meanwhile, the LIME algorithm is introduced to achieve higher precision, adaptability, and interpretability in PLPP and resource recommendation. Compared to traditional CF or rule-based methods, the proposed algorithm captures long-term dependencies in students' learning behaviors through an LSTM network. This algorithm enables a more accurate prediction of students' learning paths with an accuracy of 92.5%, significantly outperforming existing methods. Furthermore, traditional educational models struggle to precisely match students' learning needs with research resources, leading to low learning efficiency. In contrast, this study captures long-term dependencies in students' learning behaviors through an LSTM network and combines it with a hybrid recommendation mechanism of CF and CB filtering. This remarkably improves the precision of resource matching, with a resource recommendation accuracy of 93.4%, addressing the issue of uneven resource allocation. Compared to algorithms relying solely on DL, this study dynamically adjusts the learning rate through RL, optimizing the model's convergence speed and prediction performance, with a prediction bias of -0.75%, outperforming other algorithms. More importantly, this study enhances the model's interpretability through the LIME algorithm. This enables educators to understand the model's decision-making logic and address the trust issues associated with traditional "black-box" models in educational applications. Consequently, the proposed algorithm outperforms existing methods in performance. Also, it exhibits significant advantages in cross-disciplinary adaptability, long-term prediction capabilities, and interpretability, offering a more intelligent and transparent solution for graduate education.

The potential long-term impacts of the proposed model's algorithm on students' career development and research skills highlight its capacity to align learning experiences with individual academic and career objectives. By providing personalized learning pathways and resource recommendations, the DL model enables students to acquire skills more effectively, improving both learning efficiency and satisfaction. This tailored educational approach offers students clearer career guidance by identifying their interests and strengths at an early stage. Additionally, the model's accurate predictions of academic performance provide educators with actionable insights, facilitating timely interventions to address academic challenges. Such support enhances students' research capabilities and problem-solving skills. Over time, this data-driven, personalized education methodology may significantly influence students' career trajectories, bolster their innovation capabilities, and increase their competitiveness in research fields. These findings offer empirical evidence supporting personalized education and valuable guidance for educators and policymakers seeking to leverage DL technologies to promote holistic student development.

While the study primarily focuses on student outcomes, it also underscores the essential role educators play in the educational process. The insights generated by the model enable educators to better understand students' learning behaviors and needs, allowing for more targeted and effective instructional decisions. The predictive and

analytical capabilities of the model assist in monitoring student progress, identifying those who may require additional support, and enhancing the overall efficiency of instruction. Furthermore, these tools facilitate the optimized allocation of resources to meet diverse learning needs. As educators become more familiar with and trust DL models, they can utilize these technologies to refine their instructional strategies in a data-driven manner, thereby improving educational quality. This study not only transforms the learning experience for students but also provides educators with a platform to integrate advanced educational technologies into their teaching practices. By offering new perspectives and tools for professional development, the study emphasizes the dual impact of DL models: enhancing student learning outcomes and advancing educators' instructional methods and professional growth.

Although this study focuses on the student experience, it acknowledges the vital role of educators in the educational process. Educators can utilize the insights provided by the model to gain a deeper understanding of students' learning behaviors and needs, facilitating more targeted instructional decisions. Furthermore, the model's predictive and analytical tools support educators in tracking student progress and promptly identifying those requiring additional assistance. This enhancement improves instructional efficiency and enables more effective allocation of resources to meet students' specific learning needs. As educators become increasingly familiar with and confident in the use of DL models, these tools can be employed to optimize instructional strategies in a data-driven manner, ultimately elevating educational quality. This study provides educators with a platform to integrate advanced educational technologies, offering new perspectives and tools for professional development and teaching practices. Consequently, DL models not only transform the learning experiences of students but also positively influence educators' teaching methods and professional growth.

When integrating AI technologies into education, several critical considerations must be addressed to ensure their effective and ethical application. First, regarding the potential dependency of students on the system, it is crucial to position educational technology as a complementary tool that enhances, rather than replaces, students' active learning and independent thinking. To mitigate the risk of over-reliance, educators should design curricula and activities that encourage students to critically evaluate and thoughtfully apply the model's recommendations. Educators should also guide students in understanding the limitations of AI-driven tools, fostering the ability to discern when and how to effectively utilize these suggestions within their learning contexts. Second, the current DL model evaluates student performance primarily based on academic data. Future iterations should expand to include additional data types, such as student engagement metrics, feedback, and self-assessments, to develop a more comprehensive learning profile. Incorporating these elements will better address students' personalized needs and provide a holistic understanding of their learning journeys. Concurrently, educators must acknowledge the importance of these non-academic factors and proactively provide interventions and support to facilitate students' overall development. Third, achieving a balance between quantitative and qualitative indicators is essential. An excessive focus on quantitative metrics risks overlooking critical qualitative dimensions of learning, such as creativity, emotional intelligence, and interpersonal skills, which are fundamental to students' holistic development yet challenging to quantify. Future evaluations of learning outcomes should adopt a multidimensional approach, combining quantitative indicators with qualitative measures, including student

self-reports, peer evaluations, and educator observations. Such an approach will provide richer insights into students' personal and social competencies, enabling a more nuanced understanding of their growth and achievements. By adopting this comprehensive assessment framework, educational practices can better support the multifaceted development of students, fostering their academic success, personal growth, and social adaptability in an increasingly complex and interconnected world.

Graduate education typically prioritizes the development of independent research and innovation skills, while undergraduate education focuses on building foundational knowledge and exploring academic interests. Despite the distinct learning and research objectives at these two educational levels, the findings of this study offer valuable insights for enhancing undergraduate education. The application of DL models in designing personalized learning pathways and resource recommendations has significant potential to boost undergraduate students' motivation, support them in identifying and exploring their academic interests, and establish a solid foundation for their future academic and career trajectories. Adapting the model to the specific characteristics of undergraduate education is essential to achieve these outcomes. Methodologies and insights from this study can effectively support the modernization and personalization of undergraduate education through targeted customization and further investigation. Future research will focus on tailoring the proposed framework and tools to align with the learning needs and objectives of undergraduate students. Additionally, the effectiveness of these adaptations will be evaluated with respect to various learning motivations and educational goals, ensuring that the approach provides meaningful and impactful support for undergraduate learners.

5. Conclusion

5.1. Research Contribution

This study highlights the innovative application of DL technology in graduate education, demonstrating its significant effectiveness in personalized teaching, student performance prediction, and intelligent resource recommendation. By integrating advanced DL techniques, particularly LSTM networks and CNNs, the accuracy of personalized learning path predictions is notably enhanced, and the quality of educational resource recommendations is significantly improved. More importantly, this approach not only optimizes students' learning efficiency but also fosters the development of their innovative abilities. As such, the study provides new perspectives and tools for the modernization of graduate education, offering tangible benefits for both educators and students.

5.2. Future Works and Research Limitations

Looking ahead, further exploration of the potential of DL technology in interdisciplinary education is recommended, with a focus on evaluating its impact on the long-term career development of graduate students. Additionally, addressing data ethics and privacy protection concerns is crucial to ensure that technological advancements align with educational fairness and respect for student rights. Despite its contributions, this study acknowledges several limitations, including data bias, the "black box" nature of certain models, and the high computational costs involved. Future research should aim to improve model transparency, resource efficiency, and data representativeness, with the goal of fostering a more equitable, efficient, and responsible educational technology ecosystem.

Acknowledgment. This study was supported by 2024 Ideological and Political Education Project of Qingdao Agricultural University with project name "Exploring the Mechanism of Scientist Spirit in Shaping Scientific Aspirations Among Students at Agricultural Universities" (Grant No.QNSZ2024019).

References

1. Guo, K., Peng, J.: The path of integrating the spirit of scientists into the ideological and political education of graduate students based on the perspective of educating through scientific research. Advances in Education, Humanities and Social Science Research, Vol. 1, No. 3, 89. (2022)
2. Macfarlane, B.: The spirit of research. Oxford Review of Education, Vol. 47, No. 6, 737-751. (2021)
3. Ozaki, T.: Quest for science, spirit, and skills. Journal of Orthopaedic Science, Vol. 28, No. 2, 299-301. (2023)
4. Zhao, J., Li, X.: Promoting the spirit of Chinese scientists: The role and functions of Chinese STM journals. Cultures of Science, Vol. 7, No. 2, 137-146. (2024)
5. Trullàs, J. C., Blay, C., Sarri, E.: Effectiveness of problem-based learning methodology in undergraduate medical education: A scoping review. BMC Medical Education, Vol. 22, No. 1, 104. (2022)
6. Villegas-Ch, W., Román-Cañizares, M., Palacios-Pacheco, X.: Improvement of an online education model with the integration of machine learning and data analysis in an LMS. Applied Sciences, Vol. 10, No. 15, 5371. (2020)
7. Zhen, C.: Using big data fuzzy K-means clustering and information fusion algorithm in English teaching ability evaluation. Complexity, Vol. 2021, 5554444. (2021)
8. Zohuri, B., Mossavar-Rahmani, F.: Revolutionizing education: The dynamic synergy of personalized learning and artificial intelligence. International Journal of Advanced Engineering and Management Research, Vol. 9, No. 1, 143-153. (2024)
9. Hou, R., Kong, Y. Q., Cai, B.: Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning. Neural Computing and Applications, Vol. 32, No. 10, 5399-5407. (2020)
10. Kristian, A., Goh, T. S., Ramadan, A.: Application of AI in optimizing energy and resource management: Effectiveness of deep learning models. International Transactions on Artificial Intelligence, Vol. 2, No. 2, 99-105. (2024)

11. Al Ka'bi, A.: Proposed artificial intelligence algorithm and deep learning techniques for development of higher education. International Journal of Intelligent Networks, Vol. 4, 68-73. (2023)
12. Lăzăroiu, G., Andronie, M., Iatagan, M.: Deep learning-assisted smart process planning, robotic wireless sensor networks, and geospatial big data management algorithms in the Internet of Manufacturing Things. ISPRS International Journal of Geo-Information, Vol. 11, No. 5, 277. (2022)
13. Ni, A., Cheung, A.: Understanding secondary students' continuance intention to adopt AI-powered intelligent tutoring system for English learning. Education and Information Technologies, Vol. 28, No. 3, 3191-3216. (2023)
14. Sarnato, A. Z., Sari, W. D., Rahmawati, S. T., Hidayat, R., Patry, H.: The evolution of E-learning platforms: From U-learning to AI-driven adaptive learning systems. Journal of Social Science Utilizing Technology, Vol. 2, No. 2, 289-300. (2024)
15. Zhang, S., Zhao, X., Zhou, T., Kim, J. H.: Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. International Journal of Educational Technology in Higher Education, Vol. 21, No. 1, 34. (2024)
16. Guettala, M., Bourekkache, S., Kazar, O., Harous, S.: Generative artificial intelligence in education: Advancing adaptive and personalized learning. Acta Informatica Pragensia, Vol. 13, No. 3, 460-489. (2024)
17. Pratama, M. P., Sampelolo, R., Lura, H.: Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. Klasikal: Journal of Education, Language Teaching and Science, Vol. 5, No. 2, 350-357. (2023)
18. Yılmaz, Ö.: Personalised learning and artificial intelligence in science education: Current state and future perspectives. Educational Technology Quarterly, Vol. 2024, No. 3, 255-274. (2024)
19. Tang, Y., Liang, J., Hare, R.: A personalized learning system for parallel intelligent education. IEEE Transactions on Computational Social Systems, Vol. 7, No. 2, 352-361. (2020)
20. Tapalova, O., Zhiyenbayeva, N.: Artificial intelligence in education: AIEd for personalised learning pathways. Electronic Journal of e-Learning, Vol. 20, No. 5, 639-653. (2022)
21. Essa, S. G., Celik, T., Human-Hendricks, N. E.: Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review. IEEE Access, Vol. 11, 48392-48409. (2023)
22. Kanchon, M. K. H., Sadman, M., Nabila, K. F., Tarannum, R., Khan, R.: Enhancing personalized learning: AI-driven identification of learning styles and content modification strategies. International Journal of Cognitive Computing in Engineering, Vol. 5, 269-278. (2024)
23. Gm, D., Goudar, R. H., Kulkarni, A. A., Rathod, V. N., Hukkeri, G. S.: A digital recommendation system for personalized learning to enhance online education: A review. IEEE Access, Vol. 12, 34019-34041. (2024)
24. Lokare, V. T., Jadhav, P. M.: An AI-based learning style prediction model for personalized and effective learning. Thinking Skills and Creativity, Vol. 51, 101421. (2024)
25. Maghsudi, S., Lan, A., Xu, J.: Personalized education in the artificial intelligence era: What to expect next. IEEE Signal Processing Magazine, Vol. 38, No. 3, 37-50. (2021)
26. Liu, M., Yu, D.: Towards intelligent E-learning systems. Education and Information Technologies, Vol. 28, No. 7, 7845-7876. (2023)
27. Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., Oleagordia-Ruiz, I.: Recommendation systems for education: Systematic review. Electronics, Vol. 10, No. 14, 1611. (2021)
28. Chaabi, Y., Ndiyae, N. M., Lekdioui, K.: Personalized recommendation of educational resources in a MOOC using a combination of collaborative filtering and semantic content

analysis. International Journal of Scientific & Technology Research, Vol. 9, No. 2, 3243-3248. (2020)
29. Hoefler, T., Alistarh, D., Ben-Nun, T.: Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. Journal of Machine Learning Research, Vol. 22, No. 241, 1-124. (2021)
30. Khanal, S. S., Prasad, P. W. C., Alsadoon, A.: A systematic review: Machine learning based recommendation systems for e-learning. Education and Information Technologies, Vol. 25, No. 4, 2635-2664. (2020)
31. Gligorea, I., Cioca, M., Oancea, R.: Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, Vol. 13, No. 12, 1216. (2023)
32. Muniasamy, A., Alasiry, A.: Deep learning: The impact on future eLearning. International Journal of Emerging Technologies in Learning, Vol. 15, No. 1, 188. (2020)
33. Jaiswal, A., Arun, C. J.: Potential of artificial intelligence for transformation of the education system in India. International Journal of Education and Development Using Information and Communication Technology, Vol. 17, No. 1, 142-158. (2021)
34. Dhelim, S., Ning, H., Aung, N.: Personality-aware product recommendation system based on user interests mining and metapath discovery. IEEE Transactions on Computational Social Systems, Vol. 8, No. 1, 86-98. (2020)
35. Guruge, D. B., Kadel, R., Halder, S. J.: The state of the art in methodologies of course recommender systems—A review of recent research. Data, Vol. 6, No. 2, 18. (2021)
36. Chen, X., Zou, D., Xie, H.: Twenty years of personalized language learning. Educational Technology & Society, Vol. 24, No. 1, 205-222. (2021)
37. Lin, C. C., Huang, A. Y. Q., Lu, O. H. T.: Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. Smart Learning Environments, Vol. 10, No. 1, 41. (2023)
38. Munir, H., Vogel, B., Jacobsson, A.: Artificial intelligence and machine learning approaches in digital education: A systematic revision. Information, Vol. 13, No. 4, 203. (2022)
39. Lee, C. A., Tzeng, J. W., Huang, N. F.: Prediction of student performance in massive open online courses using deep learning system based on learning behaviors. Educational Technology & Society, Vol. 24, No. 3, 130-146. (2021)
40. Alqahtani, T., Badreldin, H. A., Alrashed, M.: The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, Vol. 19, No. 8, 1236-1242. (2023)
41. Yousafzai, B. K., Hayat, M., Afzal, S.: Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. Education and Information Technologies, Vol. 25, No. 6, 4677-4697. (2020)
42. Ginja, T. G., Chen, X.: Teacher educators' perspectives and experiences towards differentiated instruction. International Journal of Instruction, Vol. 13, No. 4, 781-798. (2020)
43. English, L. D.: Fifth-grade students' quantitative modeling in a STEM investigation. Journal for STEM Education Research, Vol. 5, No. 2, 134-162. (2022)
44. Okewu, E., Adewole, P., Misra, S.: Artificial neural networks for educational data mining in higher education: A systematic literature review. Applied Artificial Intelligence, Vol. 35, No. 13, 983-1021. (2021)

Peixia Li was born in October 1989 in Dongying, Shandong Province, P.R. China. She received her Ph.D. in Economics from Changwon National University, South Korea. Currently, she serves as a Student Affairs Counselor at the College of Veterinary Medicine, Qingdao Agricultural University. Her research interests include ideological and political education research, higher education research, and related fields. E-mail: lipeixia@qau.edu.cn

Zhiyong Ding was born in October 1993 in Qingdao, Shandong Province, P.R. China. He received his Master's degree in Fine Arts from Xinjiang Normal University. Currently, he serves as a Student Affairs Counselor at the College of Animation and Media, Qingdao Agricultural University. His research interests include ideological and political education research, higher education research, and related fields. E-mail:dingzhiyong@qau.edu.cn

Received: November 25, 2024; Accepted: March 11, 2025.

Computer Science and Information Systems 22(3):1251–1270 https://doi.org/10.2298/CSIS241202044L

Effectiveness of Game Technology Applied to Preclinical Training for Nurse Aides in Implementing Contact Isolation Precautions

Chiao-Hui Lin[1] and Yi-Maun Subeq [2,*]

[1] Department of Nursing, Chi Mei Medical Center, Liouying, Taiwan
susan891224@gmail.com
[2] Bachelor Program of Senior Health Promotion and Care Management for Indigenous People, College of Education, National Changhua University of Education, Changhua City, Taiwan. subeq1214@gmail.com

Abstract. . The rapid spread of emerging infectious diseases poses significant challenges to elderly care in the healthcare system. Effective training in contact isolation protective measures is essential for nurse aides to reduce infection risks. This study integrates game technology and Scaffolding theory, utilizing mobile apps for interactive teaching to enhance the contact isolation protection capabilities of nurse aides. This study employed a quasi-experimental design and involved 60 students from the Department of Senior Citizen Services at a college in Taiwan. Participants trained with the game app Golden Cicada Escapes with Its Whole Body, and their learning outcomes were evaluated using SPSS 26.0 and multiple linear regression. The results showed that the experimental group demonstrated significant improvements in cognition, skills, and self-efficacy, proving the effectiveness of game technology in learning. We hope that the innovative teaching model combining game technology with Scaffolding theory can serve as a reference for strategies in training for emerging infectious disease clinical protection in medical education both domestically and internationally, offering broad prospects for application and value for dissemination.

Keywords: Elderly Care, Nurse Aides, Game Technology, Scaffolding Theory, Contact Isolation Measures, Medical Education.

1. Introduction

In the context of global aging, the rising costs of medical care have accelerated the shift towards an aging society. This transformation poses significant challenges, prompting rapid advancements in medical technology and dramatic changes in population demographics. According to data from Taiwan's National Development Council, the proportion of the elderly population is steadily increasing and is expected to exceed 20% of the total population by 2025. This rapid and notable growth further intensifies the impact of an aging society on the socioeconomic structure [1].

* Corresponding author

With the aging population, the demand for long-term care services has surged, exacerbating the shortage of nurse aides in schools and leading to an imbalance in the labor market, particularly within the medical and caregiving sectors, posing significant social challenges. Nurse aides in schools play a pivotal role in both medical and daily care, undertaking critical tasks in long-term care services. Their shortage not only impacts the well-being of the elderly and their families but also increases the burden on the healthcare system. It is imperative for governments to proactively cultivate more caregiving talent by providing systematic training and professional courses supported by modern technology. This approach ensures sufficient caregiving manpower in an aging society while enhancing the professional standards and quality of care provided by nurse aides in schools [2].

In the context of challenges in elderly care, the application of game technology has gradually become a crucial tool. With rapid advancements in technology, game technology has proven to offer engaging and highly interactive learning experiences, particularly in the fields of medical and caregiving training. It effectively stimulates learners' interest and motivation, helping them stay focused on their studies, which in turn enhances learning efficiency and outcomes [3]. In training for contact precautions and caregiving in schools, game technology can simulate real-life scenarios to improve trainees' practical skills. Game simulations of contact isolation scenarios immerse students in infection control knowledge, equipping them with the skills needed to handle emerging infectious diseases. This experience enhances their adaptability in actual work environments, reduces errors in medical settings, and is therefore essential [4, 5].

The value of game technology extends beyond academic and clinical training; it can also serve as a tool for broadly disseminating health protection knowledge and skills, aiding society in better responding to public health challenges. This approach leverages interactive simulations and game-based scenarios that mimic real-life situations, providing trainees with hands-on experience in managing elderly care, including emergency responses and daily caregiving tasks. With the frequent outbreaks of emerging infectious diseases and the acceleration of globalization, enhancing the prevention and control knowledge of healthcare professionals and the public has become increasingly important. Digital game-based learning (DGBL) has proven to be highly effective in boosting learning motivation and knowledge retention [6]. The challenge design, reward systems, and real-time feedback mechanisms in games enhance learner engagement, making medical knowledge more focused and facilitating a deeper understanding and practical application of these skills. Moreover, the integration of game technology with big data analytics, artificial intelligence (AI), and virtual reality (VR) technology can offer innovative and efficient solutions for medical education [7].

Adjusting the difficulty of the game based on learners' behavior and progress, and providing real-time, personalized feedback, can simulate real-world scenarios of infectious disease outbreaks. This enables learners to practice preventive measures and decision-making exercises without actual risk. Such technological applications enhance the clinical adaptability of healthcare professionals and make them more proficient in responding to various public health crises [8]. In the context of technological advancements, societies facing the dual challenges of aging populations and emerging infectious disease threats must comprehensively utilize medical technology to bolster their response capabilities. Integrating enhanced caregiver training with the application of game technology in education can improve the overall capacity of healthcare services

and effectively address the increasing demand for medical care, ultimately fostering a healthier, safer, and more resilient society [9].

Infection control training for pre-clinical nurse aides often relies on passive learning methods, such as lectures or written manuals, which fail to engage learners or prepare them for practical, real-world applications. This gap in training methods has been highlighted by the growing need for innovative and interactive approaches, especially in the post-pandemic era. Integrating game technology and advanced deep learning models offers an effective approach to improving infection control during outbreaks. Game-based training and VR simulations enhance healthcare workers' response capabilities, while tools like MFMDet ensure correct PPE usage, reinforcing public health defenses. These innovations collectively strengthen the healthcare system's resilience and its ability to respond efficiently to pandemics, ensuring safety in high-risk environments. Pre-clinical nurse aides' students were chosen as study subjects because they are in the foundational stage of learning nursing skills and knowledge. And incorporating Scaffolding theory provides step-by-step support in the design of game technology scripts, helping students gradually master complex skills starting from foundational knowledge. This approach enables teachers to identify learning difficulties and adjust the curriculum, thereby enhancing nurse aides' practical skills using interactive teaching via mobile apps. The purpose of this study is to explore the effectiveness of a teaching program that integrates scaffolding theory with gaming technology in contact isolation precautions, on the cognition, attitudes, skills, and self-efficacy of pre-clinical nurse aides.The preparation of manuscripts which are to be reproduced by photo-offset requires special care. Papers submitted in a technically unsuitable form will be returned for retyping, or canceled if the volume cannot otherwise be finished on time.

2. Manuscript Preparation

2.1. Challenges in Isolation Protection Training for Nurse aides

WHO proposed a preparedness and response plan to combat emerging infectious diseases, with the primary goal of suppressing virus transmission and preventing disease and death [1]. This makes it increasingly important for clinical healthcare personnel to adhere to infection prevention and control guidelines, which include strategies such as using personal protective equipment (PPE), like masks, face shields, gloves, and protective clothing [4]. With the aging population, the cost of providing medical care continues to rise, and the shortage of nurses leads to a lack of caregiving personnel. Thus, the role of nurse aides in long-term care facilities, nursing homes, and home care for the elderly has become a significant issue. In Taiwan, the training of nurse aides is divided into two main systems: formal educational institutions that cultivate students' abilities in health promotion, disability care, and service management, enabling them to apply professional knowledge and clinical skills effectively; and an informal educational system, where individuals attend short-term training programs organized by government entities to obtain certified nursing aide qualifications and learn elderly care knowledge.

Both systems teach students essential skills such as elderly bathing and dressing, vital signs measurement, correct patient positioning, infection control, effective team communication, and maintaining a clean and safe environment. All core courses are traditionally taught in lecture format, instructing students on common statutory infectious diseases and prevention principles; they learn various isolation measures and caregiving techniques, such as properly donning and doffing isolation gowns, wearing masks, and the basics of handwashing. However, there is a significant lack of practical application; this has led to low usage rates of isolation protection and compliance with PPE use among nurse aides prior to the COVID-19 pandemic, demonstrating a severe lack of infection control knowledge and attitudes toward using personal protective equipment [9].

2.2. Application of Game Technology in the Control of Emerging Infectious Diseases

During pandemics, healthcare systems are compelled to respond swiftly and allocate sufficient resources to protect both healthcare workers and patients. This includes providing adequate personal protective equipment (PPE) such as gloves, gowns, masks, and face shields, as well as strictly enforcing standard protective measures like hand hygiene and environmental disinfection [4, 10]. These measures enable healthcare institutions to effectively reduce the risk of infection and prevent the spread of pathogens [11, 12]. However, the practice and adherence to protective strategies often face numerous challenges, including insufficient training for healthcare workers and unstable equipment supplies. Game technology emerges as an innovative solution with potential, offering new ways to address these issues. By using game-based technology, virtual training scenarios can be created to strengthen healthcare workers' response capabilities in infection control and enhance their clinical adaptability. For instance, simulations that replicate the process of breaking infection chains allow learners to practice proper PPE usage and respond to various transmission routes in a risk-free virtual environment [13, 14].

Virtual reality (VR) technology can simulate real-life scenarios of infectious disease outbreaks, allowing trainees to practice standard protective measures in a risk-free environment. This includes learning how to effectively implement contact isolation, don protective equipment, and provide patient care. Such training enhances healthcare workers' response capabilities and reinforces their proficiency in correctly executing preventive measures in real-world settings [15].

With the rapid emergence of infectious diseases, enhancing the response capabilities of healthcare workers and the public has become a critical issue. Additionally, industrial development has led to environmental pollution and health risks that are closely linked to the spread of emerging infectious diseases. In response to these challenges, the development of game technology offers an innovative solution. By designing specialized educational games, players can learn and master infection prevention measures in a virtual environment, including the proper use of personal protective equipment (PPE), the implementation of contact isolation measures, and emergency responses during disease outbreaks. Furthermore, wearing masks in workplaces has become an effective way to prevent the inhalation of harmful gases and pathogens [16].

The outbreak of emerging infectious diseases has intensified the focus and challenges related to mask protection monitoring, particularly in medical facilities and high-risk workplaces. Research has introduced a deep learning-based mask detection model, MFMDet, which utilizes a recursive feature pyramid structure and deformable RoI pooling technology to enhance detection capabilities for targets of varying scales. Additionally, a mixed augmentation technique is implemented to increase sample diversity. MFMDet has shown improved accuracy across multiple datasets, providing strong support for the prevention and control of emerging infectious diseases. The application of this technology not only effectively detects whether personnel are wearing masks correctly but also enhances public health protection, especially during periods of high incidence of infectious diseases [17, 18].

In summary, integrating game technology with innovative medical training solutions enhances the ability of healthcare professionals to respond to emerging infectious diseases. Through simulated practice and real-time feedback, these tools enable effective handling of pandemic challenges in real-world scenarios. Additionally, advancements in mask technology ensure optimal preparedness for protective measures.

2.3. Game Design for Simulating Infectious Diseases

Based on infectious disease spread models, game developers can design simulation games that integrate sources of infection, transmission routes, and hosts. Players must make rapid decisions to prevent the spread of diseases, enhancing health awareness and familiarizing them with control strategies through real-time feedback [19, 20]. VR technology in medical education, particularly in training clinical skills for nurses and medical students, allows risk-free practice of protective measures and decision-making. Digital game-based learning (DGBL) has proven effective in boosting motivation and engagement, making complex academic concepts more accessible [21, 22]. By integrating big data analytics, games can track learners' progress and tailor content to individual needs, improving educational outcomes and practical application [23, 24]. Game scripts incorporating scaffolding theory and interactive teaching enable streamlined content adaptation for online learning [25, 26].

In summary, in the post-pandemic era, the integration of game technology with modern medical education on online learning platforms offers an innovative and effective approach for the prevention and control of emerging infectious diseases. Through advanced game design, virtual reality applications, and big data analytics, these technological tools will become essential for healthcare professionals and the public to learn infection prevention and control. This integration fosters a comprehensive improvement in health literacy and supports the advancement of global public health.

The integration of game technology into medical education represents a transformative approach for infection prevention in the post-pandemic era. Advanced tools such as virtual reality and big data analytics provide healthcare professionals and the public with accessible and effective learning platforms. These developments address the pressing need for enhanced infection control training, paving the way for innovative interventions like those explored in this study.

2.4. Purpose of Game Program Design

The framework of this study is based on the researcher's adaptation of the isolation guidelines from the Centers for Disease Control in Taiwan and the United States to design a teaching program titled "Scaffolding Theory Combined with Game Technology Intervention for Contact Isolation Protective Measures.

This study employed a quasi-experimental design, dividing participants into a control group (n = 30) and an experimental group (n = 30). The participants were students aged 18 and older from the Department of Senior Citizen Services at a college in Taiwan. Participants were assigned to the groups through simple randomization using a random number generator, ensuring equal probability for each participant.

The study aimed to evaluate the learning outcomes of the experimental group, who learned contact isolation measures through the game-based intervention, "Golden Cicada Escape." The focus was on cognitive, attitudinal, technical skills, and self-efficacy improvements among nurse aides in training, as well as satisfaction levels within the experimental group. The study compared the learning effectiveness between the experimental and control groups

Scaffolding theory guided the design of the educational tasks by providing a structured framework. For instance, the intervention included progressively challenging scenarios within the game to help learners build confidence and competence step-by-step. Furthermore, the program incorporated real-time feedback and support, such as hints and prompts, tailored to each learner's progress to ensure continuous engagement and effective skill acquisition. This study applies scaffolding theory [27, 28] to design an interactive and structured teaching program that integrates game technology. The specific steps and content are as follows:

A. Designing a Game Technology-Based Teaching Program for Contact Isolation Protective Measures:

(A) Analyze Learner Characteristics: Assess the learning preferences and characteristics of pre-clinical nurse aides.

(B) Set Learning Objectives: Establish clear learning goals aligned with the capabilities of pre-clinical nurse aides.

(C) Develop Teaching Guidelines: Create instructional guidelines for teaching contact isolation protective measures using game technology.

(D) Incorporate Scaffolding Theory: Design educational tasks that integrate scaffolding theory into the game technology script to support practical teaching.

(E) To create comprehensive instructional resources, we developed a set of materials, including written documents, draft illustrations, and audiovisual teaching content (Fig. 1).

Figure 1 presents a draft illustration of the instructional framework, depicting key elements of the teaching process.

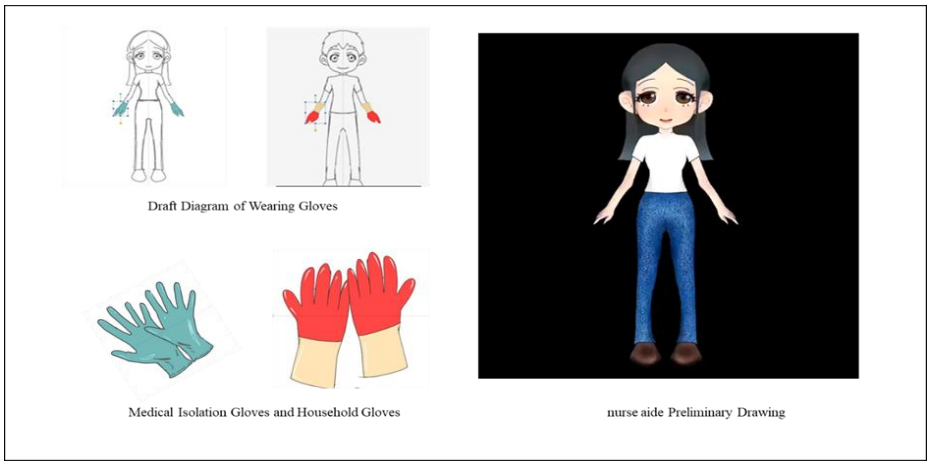


Fig. 1 Draft drawing

The instructional content includes: (A) Teaching handouts on contact isolation protective measures; (B) Short instructional videos on the proper use of personal protective equipment (PPE), including the correct procedures for donning and doffing gowns, masks, and gloves; (C) Evaluation of contact isolation protective measures and a gamified guide for PPE procedures.

2.5. Implementation of Game Technology-Based Teaching Methods for Contact Isolation Protective Measures

The course uses game technology to create engaging learning scenarios that spark students' interest in the practical application of contact isolation protective measures, including the correct use and donning/doffing of personal protective equipment (PPE). The experimental group participated in game challenges designed with content on infectious disease knowledge and the application of isolation measures, implemented as part of 10 online game levels in the teaching program. As is show in Fig. 2 illustrates the Game Startup Flowchart, depicting the sequence of game levels and the progression flow for participants.

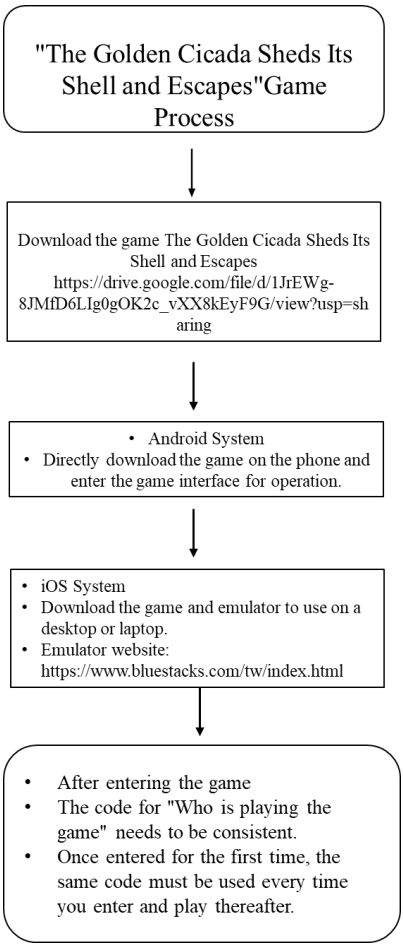


Fig. 2. Game Startup Flowchart

2.6. Game Program File Development and Design

The configuration files for research in Android Studio are written in XML format and encompass a variety of research settings and configurations. Fig. 3 presents the Design Thinking Diagram for 'Golden Cicada Escape,' illustrating the conceptual framework and workflow that guided the development of the research application.

These files primarily record various development environment settings for Android projects, such as auto-import configurations, layout management, execution targets, recent files, templates, and task management details. These configuration files are typically auto-generated and managed by the development environment, eliminating the need for manual editing by developers.

The research configuration files in Android Studio are created in XML format and serve as a record of various detailed development settings and configurations within the research projects. Fig.4 presents a code example of the app homepage with background music, demonstrating the implementation of user Interface components and audio integration in the "Golden Cicada Escape" application.

Managed and generated automatically by Android Studio, these files include comprehensive information about the project's layout, execution, and version control. Developers do not need to manually edit these files; instead, they can modify configurations through Android Studio. These configuration files coordinate multiple functions within Android Studio during research, facilitating a smoother process for application development, testing, and debugging.

The following outlines the main components of the file and their function descriptions:

Based on the content provided in the image, these code components are used to manage and describe various functionalities in Android development. Below is a description of the main components and their functions:

1. Adroid Layout Configuration (<component name="AndroidLayouts">):Manages the layout resources of the Android project and is responsible for handling and displaying different layout files.

2. Auto-Import Settings (<component name="AutoImportSettings">):Defines whether the auto-import feature is enabled for the project, facilitating the automatic import of required dependencies or packages.

3. Change List Management (<component name="ChangeListManager">):Manages change files and system change lists, controlling whether conflict changes are displayed and change lists are deleted.

4. Execution Target Management (<component name="ExecutionTargetManager">): Configures and manages the execution targets for the program, such as setting a test environment or an emulator for running the project.

5. External Project Management (<component name="ExternalProjectsData"> and <component name="ExternalProjectsManager">):Handles the reading and processing of external dependencies and tasks within the project.

6. Run Settings (<component name="RunManager">):Defines the run configurations for the program, such as whether it is in debug mode or release mode.

7. Recently Used Files and Templates (<component name="RecentsManager"> and <component name="FileTemplateManagerImpl">):Records recently used files and templates to facilitate quick access.

8. Task Management (<component name="TaskManager">):Configures and manages the settings related to tasks within the project.

Fig.3. Design Thinking Diagram for 'Golden Cicada Escape '

Fig. 4. App Homepage with Background Music: Code Example

2.7. Game Technology Applied Effect Analysis

This study was approved by the Institutional Review Board (IRB) of Chi Mei Medical Center (approval number: blinded).

The objective of this study was to evaluate the effectiveness of integrating scaffolding theory and game-based technology into teaching contact isolation protective measures. Participants were divided into a control group and an experimental group. The study design included pre-tests, classroom instruction, random group assignments, and a game-based learning intervention for the experimental group, followed by post-tests conducted one month later to assess learning outcomes. The target population consisted of fourth- and fifth-year students aged 18 and older who held nurse aide certificates. Data analysis was performed using SPSS version 26.0, with a significance level set at $p < .05$.

2.8. Basic Information of Study Participants

As shown in Tables 1 and 2, a total of 60 students participating in the contact isolation protective measures course, 30 were assigned to the experimental group and 30 to the control group. Homogeneity tests indicated no significant differences between the two groups in terms of age, gender, clinical experience, and related course background.

This study included 60 pre-clinical nurse aide students aged 18–20, with 78.3% aged 18–19 and 21.7% aged 19–20. The gender distribution was 75% female and 25% male. Most participants (86.7%) had no prior caregiving experience, while 63.3% had attended a contact isolation-related courses in the past year. A homogeneity test confirmed no significant differences between the experimental and control groups in age, gender, caregiving experience, or prior course participation. As shown in Table 2, we analyzed whether these factors influenced learning outcomes and found no significant differences ($p=0.108$) between participants with and without prior infection control training.

Table 1. Descriptive Statistics of Basic Information

Variable	Category	Frequency	Percentage
Age	18-19	47	78.3
	19-20	13	21.7
Gender	female	45	75.0
	male	15	25.0
Students with nurse aide work-study experience	No	52	86.7
	Half a year	7	11.7
	Over one year	1	1.6
Whether the student has taken courses related to contact isolation measures within the past year	No	38	63.3
	Yes	22	36.7

Table 2. Homogeneity tests

	Experimental Group	Control Group	Total	F/χ^2	p
	(n=30)	(n=30)	(n=60)		(Two-ailed)
Item	n (%)	n (%)	n (%)		
Age				χ^2=4.812	.028
18-19 years old	20 (66.6)	27 (90)	47 (78.3)		
19-20 years old	10 (33.4)	3 (10)	13 (21.7)		
Gender				χ^2=.800	.371
Female	21 (70)	24 (80)	45 (75)		
Male	9 (30)	6 (20)	15 (25)		
Students with nurse aide work-study experience				χ^2=1.220	.543
None	27 (90)	25 (83.3)	52 (86.7)		
Half a year	3 (10)	4 (13.3)	7 (11.6)		
Greater than one year.	0 (0)	1 (3.4)	1 (1.7)		
Whether the student has taken courses related to contact isolation measures within the past year				χ^2=2.584	.108
None	16 (53.4)	22 (73.4)	38 (63.4)		
Have	14 (46.6)	8 (26.6)	22 (36.6)		

2.9. Post-Test Learning Outcomes

This section discusses the learning outcomes of the game technology intervention in the teaching program for contact isolation protective measures. A paired-sample t-test was conducted to analyze the post-test differences between the control group, which received classroom-based teaching on contact isolation measures, and the experimental group, which underwent the intervention using game technology, As show in Table 3 presents the results of the Independent Samples t-Test for Technical Skills, Cognition, Attitudes, and Self-Efficacy in the Game Technology Intervention. A total of 56 participants (93.3%) completed the post-test, with 2 participants from each group failing to complete the test, leaving a total of 4 participants (6.7%) who did not fill out the post-test, but they were still included in the analysis.

The results are as follows:

2.9.1 Technical Skills

The post-test average scores for the experimental and control groups were 20.71 and 17.32, respectively. Both groups scored higher in the post-test compared to their pre-test scores. Statistical analysis showed a significant difference ($p < .001$) between the two groups **(Table 3)**.

2.9.2 Cognition

The post-test average scores for the experimental and control groups were 52.25 and 45.00, respectively. Both groups showed an increase in post-test scores compared to their pre-test scores, with statistical analysis revealing a significant difference ($p < .003$) **(Table 3)**.

2.9.3 Total Score (Technical Skills + Cognition)

The post-test average scores for the experimental and control groups were 72.96 and 62.32, respectively. Both groups' post-test scores were higher than their pre-test scores, and statistical analysis indicated a significant difference ($p < .001$) **(Table 3)**.

2.9.4 Attitude

The post-test average scores for the experimental and control groups were 21.75 and 21.39, respectively. While the post-test scores were higher than the pre-test scores, the difference was minimal, and statistical analysis showed no significant difference ($p = .597$) **(Table 3)** .

2.9.5 Self-Efficacy

The post-test average scores for the experimental and control groups were 43.46 and 40.68, respectively. After the intervention, both groups showed improved post-test scores compared to their pre-test scores, with statistical analysis showing a significant difference ($p < .024$) **(Table 3)**.

Results: After receiving different intervention measures, the learning outcomes of contact isolation protective measures in the experimental and control groups were compared. The experimental group, which used game technology intervention, showed significant differences in technical skills, cognition, total score (technical skills + cognition), and self-efficacy.

Table 3. Independent Samples t-Test for Technical Skills, Cognition, Attitudes, and Self-Efficacy in the Game Technology Intervention (n=60)

Variable	Group		t-value	p
	Control Group （n=30）	Experimental Group(n=30）		
	Mean±SD			
Post-test skills	17.32±4.234	20.71±3.230	3.371	0.001*
Post-test knowledge	45.00±9.978	52.25±6.731	3.187	0.003*
KA total score[1]	62.32±11.935	72.96±5.399	4.299	<0.001**
Post-test attitude.	21.39±2.544	21.75±2.238	0.558	0.597
Post-test self-efficacy	40.68±4.959	43.46±3.995	2.315	0.024*

*p<0.05

**p<0.001 Note 1: The total score refers to the sum of cognition (knowledge) and skills, with a maximum score of 100

.

A comparative analysis showed no significant differences in learning outcomes between participants with and without prior infection control training, indicating that previous exposure to infection control did not significantly affect performance in this study. This indicates that the game technology intervention significantly improved the learning outcomes and self-efficacy of pre-clinical nurse aides. Although there was an improvement in attitude in the post-test, it did not reach statistical significance. Although there was an improvement in attitude in the post-test, it did not reach statistical significance. This may be due to the participants being student care workers who have not yet entered the workforce. Their limited understanding of real-world scenarios could

have affected their comprehension of certain questionnaire items. However, the experimental group, which received the gaming technology intervention, showed significant improvements in technical skills, cognitive understanding, total scores (technical skills and cognition), and self-efficacy. These results suggest that gaming technology effectively enhances learning outcomes, even if attitude changes may take longer to manifest.

3. Satisfaction Analysis of the Experimental Group after Game Technology Intervention

3.1. Descriptive Statistics of the Experimental Group's Satisfaction

As show in Table 4, after the experimental group learned contact isolation protective measures through the game "Golden Cicada Escape" designed by the researchers, the average satisfaction score was 22.04, with the lowest score being 19 and the highest score being the maximum of 25. The satisfaction survey consisted of 5 questions, using a Likert scale with five response levels, where 5 was "very satisfied" and 1 was "very dissatisfied." The survey covered game planning, time management, course content absorption, scenario design, and correct use of personal protective equipment. Detailed analysis showed that the highest score was for the game’s contribution to content absorption, with an average of 4.78, while the lowest score was for game time management, with an average of 4.03.

Table 4. Descriptive Statistics of the Experimental Group's Satisfaction (n=30)

Variable	Mean ± Standard Deviation	Minimum value	Maximum value
Post-test satisfaction	22.04±1.953	19	25

3.2. Correlation Analysis of Post-Test Scores and Satisfaction with Continuous Variables After Game Technology Intervention in the Experimental Group

As is show in Table 5, the total satisfaction score showed a high positive correlation with the overall game planning, game time management, game’s contribution to course content absorption, scenario design, and correct use of personal protective equipment, with statistically significant differences.

Results: The correlation analysis between satisfaction and various variables revealed that, after the game technology intervention, not only was the absorption of course content on contact isolation protective measures improved, but the correct use of personal protective equipment also aligned with the study's objectives.

Table 5. Correlation between Post-Test Continuous Variables and Satisfaction in the Experimental Group (n=30)

	Skills	Constant	KA Total	Attitude	self-efficacy	Game Planning	Game Arrangement	Game Absorption	Scenario Setting	Correct Protection	Satisfaction Total Score
Skills	1										
Knowledge	-.612**	1									
KAtotal註1	-.164	.881**	1								
Attitude	-.051	-.323	-.433*	1							
Self-efficacy	-.173	-.077	-.2	.581**	1						
Satisfaction Items											
Game Planning	.01	-.273	-.334	.427*	.615**	1					
Game Arrangement	-.054	-.365	-.488**	.553**	.507**	.583**	1				
Game Absorption	-.184	.072	-.02	.337	.372	.452*	.187	1			
Scenario Setting	-.033	-.163	-.223	.37	.35	.517**	.800**	.335	1		
Correct Protection	.089	-.147	-.13	.341	.431*	.549**	.438*	.471*	.403*	1	
Satisfaction Total Score	-.039	-.246	-.33	.536**	.596**	.812**	.822**	.600**	.824**	.740**	1

*Correlation is significant at the 0.05 level (two-tailed).

**Correlation is significant at the 0.01 level (two-tailed).

Note 1: The total score refers to the sum of cognition (knowledge) and skills, with a maximum score of 100.

3.3. Factors Related to the Learning Outcomes of Contact Isolation Protective Measures

A multiple linear regression analysis was conducted to examine the relationships between variables, and the results are presented in Table 6. The regression model used to test these relationships is illustrated in Fig. 5.

The attitude explained 33.1% of the variance in self-efficacy, making attitude the most influential factor on self-efficacy. Additionally, the combined explanation of attitude and the post-test total score (which includes cognition and skills) reached 38.4%, with the increase in explanatory power being statistically significant ($\Delta F = 18.401$, $p < 0.05$).

The results indicate that game-based learning not only improves cognitive understanding but also enhances self-efficacy, which is a critical factor in clinical decision-making. The practical significance suggests that integrating game technology

into nurse aide training could lead to more confident and competent healthcare providers.

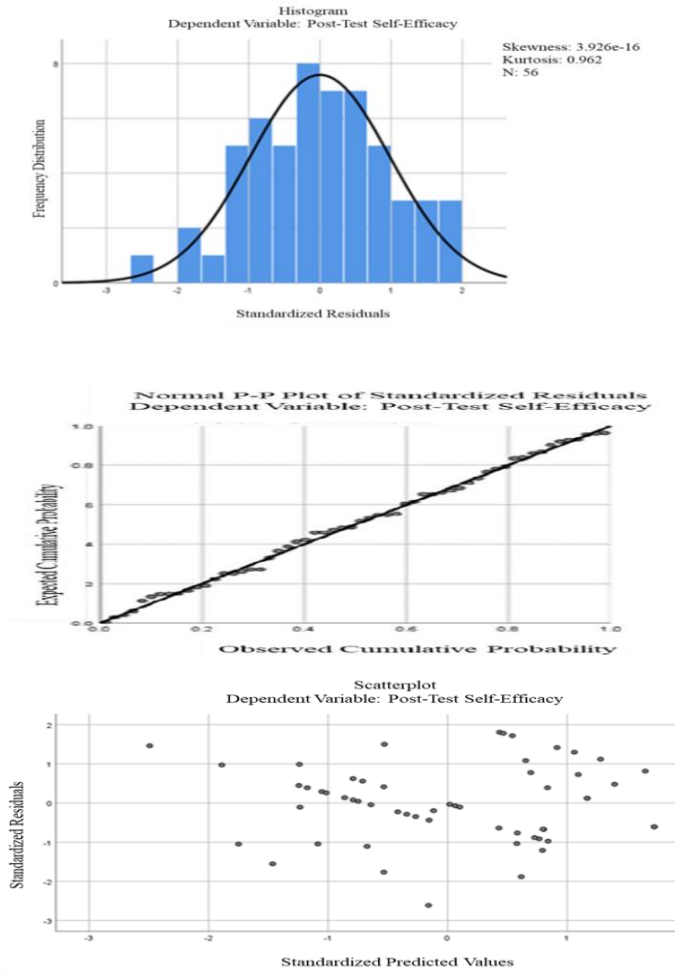


Fig. 5. Normal Distribution Chart

Table 6. Multiple Linear Regression Coefficients for Self-Efficacy Estimation (n=60)

Model		Unstandardized Coefficient		Standardized Coefficient			Collinearity Statistics			
		B	Standard Error	Beta	T	Significance	Tolerance	VIF	Δ R^2	F
1	(Constant)	17.148	4.540		3.777	.000			0.331	28.184*
	Attitude	1.153	.210	.591	5.482	.000	1.000	1.000		
2	(Constant)	10.758	5.285		2.036	.047			0.384	18.164*
	Attitude	1.126	.204	.577	5.518	.000	.996	1.004		
	KA total[1]	.103	.047	.228	2.178	.034	.996	1.004		

*Correlation is significant at the 0.05 level.
Note 1: The total score refers to the sum of cognition (knowledge) and skills, with a maximum score of 100.

3.4. Discussion

This study employed a randomized controlled experimental design and surveyed 60 students from the Elderly Services Management program at Chiayi Specialized Medical School in Taiwan. In this study, Scaffolding theory was integrated with game technology to design a course focused on contact isolation protective measures. The application of scaffolding theory in nursing education aims to stimulate students' interest in theoretical knowledge while enhancing their problem-solving abilities [29]. The results show that through the integration of game technology and mobile apps, students were able to solve problems and apply effective strategies within the game, thereby enhancing their learning motivation, especially when solving challenging learning problems [30]. Furthermore, the combination of online learning with big data analysis brings new trends to nursing education, where game technology significantly enhances learning motivation and facilitates content absorption.

In Taiwan, the demand for long-term care has been increasing in recent years, leading universities to establish many departments of long-term care to cultivate talent. Although core courses in these schools include training on infection control and isolation measures, isolation techniques have not been integrated, resulting in the suspension or temporary cessation of services at long-term care medical institutions, nursing homes, and home services during critical periods of emerging infectious diseases outbreaks [2]. The results of this study show that 63% of the students had not taken any courses related to isolation techniques within the past year, highlighting the need for classroom teaching to be combined with technical practice. Applying Scaffolding theory in the curriculum engages students, enhances their problem-solving abilities, and ultimately transforms theoretical knowledge into practical clinical application [29].

The design of game technology includes elements that enhance the enjoyment of learning, learning objectives, rewards, and problem-solving, playing a crucial role in learning, lifestyle management, and disease awareness. However, few games have been developed for safety precautions against new infectious diseases [31]. Based on this argument, the experimental group in this study utilized online game technology to design a level-based game, sparking students' interest in the course and enabling self-learning about infectious diseases, contact isolation protective measures, and personal protective equipment. The study results indicated that after the intervention of game technology, there was a significant improvement in students' technical skills, cognition, and self-efficacy, especially in learning contact isolation protective measures. Previous scholars have discussed game technology and online learning courses, emphasizing flexibility and adaptability, using the gaming process for problem reasoning, decision-making, and problem-solving. This training helps students use games to find methods and inspire effective strategies to enhance learning motivation [26]; although attitude scores improved in the post-test, no significant differences were observed. This may be due to students not yet entering the workforce, leading to limited understanding of certain real-world situations.

This study highlights the effectiveness of integrating game technology and scaffolding theory in improving the understanding, technical skills, and self-efficacy of pre-clinical nurse aides. By addressing critical gaps in infection control training, the research contributes to the development of more interactive and engaging educational approaches in healthcare.

However, certain limitations should be acknowledged. The high cost of game design limited the inclusion of fully animated features, potentially affecting the immersive learning experience. Additionally, the platform's Android exclusivity posed accessibility challenges for iOS users, which may have influenced participation rates. Finally, the focus on pre-clinical students limits the generalizability of the findings to other healthcare worker populations.

4. Summary

In conclusion, the integration of game technology and mobile apps not only enhances students' learning motivation but also strengthens their skills and cognitive abilities, particularly in learning contact isolation protective measures. Future research could further explore how to combine game technology with big data to develop professional learning software and apply it to vocational education, thus enhancing the effectiveness of professional skills learning. Regarding the contributions and future potential of this study, they include: 1.Through scaffolding support and gamified learning, students can gain a deeper understanding of the principles and practical applications of contact isolation measures, which helps translate theoretical knowledge into practical operational skills. 2. Game technology creates interactive experiences that simulate real-life scenarios, allowing learners to practice and master contact isolation and infection control skills in a risk-free environment.3. Game elements can enhance the enjoyment and engagement of learning, thereby boosting students' motivation, which is particularly important for long-duration and potentially tedious medical education. Finally, game technology provides immediate feedback, enabling students to understand their performance and adjust their learning strategies based on the feedback.

Acknowledgment. This study was partially sponsored by the Chi Mei Medical Center, Liouying. (No. CLFHR11252) and the National Science and Technology Council, Taiwan (Contract No. 111-2410-H-025 -020 -MY2). Authors express our thanks for financial support(s).

References

1. Centers for Disease Control and Prevention. 2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings-Part III.A. Standard Precautions Standard. (2007)
2. Centers for Disease Control and Prevention. How Infections Spread. [Online]. Available: https://www.cdc.gov/infectioncontrol/spread/index.html (2016)
3. Daniel, S. J. Education and the COVID-19 pandemic. Prospects, VOL. 49, No. 1, 91-96. (2020)
4. Deng, K., and Wang, G. Online mode development of Korean art learning in the post-epidemicera based on artificial intelligence and deep learning. The Journal of Supercomputing, Vol 80, No. 6, 8505-8528. (2024)
5. Durante-Mangoni, E., Andini, R., Bertolino, L., Mele, F., Bernardo, M., Grimaldi, M., Cuomo, N., Tiberio, C., Falco, E., Di-Spirito, A., Raffone, M., Russo, M. G., Atripaldi, L., and Zampino, R. Low rate of severe acute respiratory syndrome coronavirus 2 spread among

health-care personnel using ordinary personal protection equipment in a medium-incidence setting. Clinical Microbiology and Infection, Vol. 26, No. 9, 1269-1270. (2020)
6. Garcia-Molina, H., Ullman, D. J., and Widom, J. Database Systems: The Complete Book. Prentice Hall, New Jersey, USA. (2002)
7. Healthcare Infection Control Practices Advisory Committee. Core Infection Prevention and Control Practices for Safe Healthcare Delivery in All Settings -Recommendations of the Healthcare Infection Control Practices Advisory Committee. Centers for Disease Control and Prevention. (2019)
8. Houghton, C., Meskell, P., Delaney, H., Smalle, M., Glenton, C., Booth, A., Chan, H. X., Devane, D., and Biesty, L. M. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database of Systematic Reviews, 2020(8).
9. Long, J., Luo, C., Chen, R., Yu, J., and Li, K. C. A cross-layered cluster embedding learning network with regularization for multivariate time series anomaly detection. The Journal of Supercomputing, Vol. 80, No. 8, 10444-0468. (2024)
10. Ministry of Health and Welfare, Department of Nursing and Health Care. Revised "Care Attendant Training Implementation Plan," with the name changed to Care Attendant Qualification Training Plan. [Online]. Available: https://www.mohw.gov.tw/cp-18-71164-html (2022)
11. National Development Council. Population Projections for the Republic of China (2020 to 2070). [Online]. Available: https://pop-proj.ndc.gov.tw/ (2024)
12. Phutela, N., Chowdary, A. N., Anchlia, S., Jaisinghani, D., and Gabrani, G. Unlock Me: A real-world driven smartphone game to stimulate COVID-19 awareness. International Journal of Human-Computer Studies, Vol. 164, 102818. (2022)
13. Rebmann, T., Alvino, R. T., Mazzara, R. L., and Sandcork, J. Infection preventionists' experiences during the first nine months of the COVID-19 pandemic: Findings from focus groups conducted with Association of Professionals in Infection Control & Epidemiology (APIC) members. American Journal of Infection Control, Vol. 49, No. 9, 1093-1098. (2021)
14. Siegel, J. D., Rhinehart, E., Jackson, M., Chiarello, L., and Healthcare Infection Control Practices Advisory Committee. 2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Health Care Settings. (2019)
15. Upadhyay, S., and Smith, D. G. Healthcare associated infections, nurse staffing, and financial performance. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 60, 00469580231159315. (2023)
16. Venigalla, A. S. M., Vagavolu, D., and Chimalakonda, S. SurviveCovid-19 – An educational game to facilitate habituation of social distancing and other health measures for COVID-19 pandemic. International Journal of Human-Computer Interaction, Vol. 38, No. 16, 1563-1575. (2022)
17. Vygotsky, L. S. The Collected Works of L. S. Vygotsky: The Fundamentals of Defectology Vol. 2. Springer Science & Business Media. (1987)
18. Wood, D. J., Bruner, J. S., and Ross, G. The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, Vol. 17, 89-100. (1976)
19. Xu, R., Wang, P., Li, X., and Nie, R. YOLO-ARGhost: A lightweight face mask detection model. The Journal of Supercomputing, Vol. 80, No. 3, 3162-3182. (2024)
20. Yu, H., Cai, L., Min, H., and Su, X. Advancing medical data classification through federated learning and blockchain incentive mechanism: Implications for modern software systems and applications. The Journal of Supercomputing, Vol. 80, No. 8, 10469-10484. (2024)
21. Asbell-Clarke, J., Rowe, E., Almeda, V., Edwards, T., Bardar, E., Gasca, S., Baker, R. S., and Scruggs, R. The development of students' computational thinking practices in elementary- and middle-school classes using the learning game, Zoombinis. Computers in Human Behavior, Vol. 115, 106587. (2021)

22. Asniza, I. N., Zuraidah, M. O. S., Baharuddin, A. R. M., Zuhair, Z. M., and Nooraida, Y. Online game-based learning using Kahoot! to enhance pre-university students' active learning: A students' perception in biology classroom. Journal of Turkish Science Education, Vol. 18, No. 1, 145-160. (2021)
23. Bandura, A. The assessment and predictive generality of self-percepts of efficacy. Journal of Behavior Therapy and Experimental Psychiatry, Vol. 13, No. 3, 195-199. (1982)
24. Cao, R., Mo, W., and Zhang, W. MFMDet: Multi-scale face mask detection using improved Cascade R-CNN. The Journal of Supercomputing, Vol. 80, No. 4, 4914-4942. (2024)
25. Kwon, S., Joshi, A. D., Lo, C. H., Drew, D. A., Nguyen, L. H., Guo, C. G., Ma, W., Mehta, M. W., Shebl, M. F., Warner, T. E., Astley, M. C., Merino, J., Murray, B., Wolf, J., Ourselin, S., Steves, J. C., Spector, D. T., Hart, E. S., Song, M., VoPham, T., and Chan, A. T. Association of social distancing and face mask use with risk of COVID-19. Nature Communications, Vol. 12, No. 1, 1-10. (2021)
26. Liu, T., and Israel, M. Uncovering students' problem-solving processes in game-based learning environments. Computers & Education, Vol. 182, 104462. (2022)
27. Ministry of Health and Welfare, Department of Nursing and Health Care. Revised "Care Attendant Training Implementation Plan," with the name changed to Care Attendant Qualification Training Plan. [Online]. Available: https://www.mohw.gov.tw/cp-18-71164-1.html (2022)
28. Zisook, R. E., Monnot, A., Parker, J., Gaffney, S., Dotson, S., and Unice, K. Assessing and managing the risks of COVID-19 in the workplace: Applying industrial hygiene (IH)/occupational and environmental health and safety (OEHS) frameworks. Toxicology and Industrial Health, Vol. 36, No. 9, 607-618. (2020)
29. Zaragoza, A., Seidel, T., and Santagata, R. Lesson analysis and plan template: Scaffolding preservice teachers' application of professional knowledge to lesson planning. Journal of Curriculum Studies, Vol. 55, No. 2, 138-152. (2023)
30. Zhong, Y., Guo, K., Su, J., and Chu, S. K. W. The impact of esports participation on the development of 21st century skills in youth: A systematic review. Computers & Education, Vol. 191, 104640. (2022)
31. Agrawal, R., and Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Databases, Santiago, Chile, Morgan Kaufmann, pp. 487-499. (1994)

Chiao-Hui Lin, MSN, holds a Master's degree in Nursing from the Graduate Institute of Nursing at National Taichung University of Science and Technology, Taiwan. She is currently serving as a nurse practitioner at Chi Mei Medical Center, Liouying Branch, Tainan County, Taiwan. Her areas of expertise include geriatric nursing and the education and training of care attendants.

Yi-Maun Subeq received her Ph.D. from the Institute of Medical Science at Tzu Chi University, Hualien, Taiwan, in 2008. She is currently an Associate Professor at National Changhua University of Education, Taiwan. Her research interests include aging science, transcultural health, medical education and informatics, and Indigenous health issues. Dr. Subeq has authored or coauthored more than 60 publications in reputable academic journals. In recent years, she has actively contributed to the promotion of Indigenous Peoples' Health Law, served as a member of the Indigenous Peoples Health Committee under the Ministry of Health and Welfare, and established the Consultation Project Management Center for Indigenous Peoples under the Human Body Research Law.

Received: December 02, 2024; Accepted: February 24, 2025.

Computer Science and Information Systems 22(3):1271–1298 https://doi.org/10.2298/CSIS241205046L

The Analysis of Intelligent Urban Form Generation Design based on Deep Learning

Zeke Lian[1,*], Hui Zhang[2,*,†], and
Ran Chen[3]

[1] Landscape ecology School Faculty of Organizational Sciences, Ningbo City College of Vocational Technology Ningbo, 315100 China
lianzeke@nbcc.edu.cn
[2] Business School, Ningbo City College of Vocational Technology, Ningbo, 315100, China
zhanghui11261993@163.com
[3] School of landscape architecture, Beijing Forestry University, Beijing,100083, China
chenran705367787@bjfu.edu.cn

Abstract. In response to the growing demand for intelligent solutions in urban planning, this study constructs a deep learning-based framework for generating intelligent urban morphology, effectively addressing pressing real-world challenges. At the outset, the study explores the core concepts of green and ecological principles within the evolution of contemporary urban forms, establishing a robust theoretical foundation for subsequent investigations. The study provides a detailed explanation of the practical application paradigms of deep learning, encompassing meticulously selected technical methodologies, carefully designed algorithmic structures, and an optimized parameter configuration system. Together, these elements form a comprehensive technological application framework. An innovative application of convolutional neural networks is introduced for the in-depth analysis and processing of urban street imagery. This advancement enables critical urban planning functions, including road network design, detailed analysis of building distributions, optimization of public facility layouts, and dynamic traffic flow analysis. These capabilities address the key limitations of traditional planning methods by enhancing intelligent analysis and precise decision-making. To evaluate the model's performance quantitatively, a systematic testing scheme is developed and implemented, covering various scenarios, including daytime and nighttime conditions. This approach ensures a comprehensive assessment of the precision and effectiveness of each functionality. The core significance and contributions of this study are encapsulated in its empirical findings. The proposed model achieves accuracy and fit metrics exceeding 93% across all testing dimensions, representing a significant advancement that provides robust and targeted support for urban planning practices. By integrating deep learning technologies into the intelligent urban morphology generation framework, the study successfully implements critical functions such as efficient road network planning and scientific analysis of building distributions. Furthermore, the study introduces cutting-edge technological tools and innovative methodologies to the urban planning discipline, advancing the development of intelligent urban planning. Its contributions are of profound value in both theoretical innovation and practical application, offering transformative potential for the field.

* First co-autors
† Corresponding author

Keywords: Urban Planning; Intelligence; Deep Learning; Convolutional Neural Network; Green Ecology.

1. Introduction

With the rapid development of Artificial Intelligence (AI) technology, Deep Learning (DL) has made significant achievements in various fields, particularly in image recognition and natural language processing [1]. The acceleration of urbanization has brought increasingly complex challenges to urban planning and design. How to utilize advanced technologies to achieve more scientific, efficient, and sustainable urban development has become an important research topic. Against this backdrop, DL, with its exceptional capabilities in data analysis and pattern recognition, has gradually gained widespread attention in the field of urban planning. However, applying DL to the intelligent generation of urban forms remains a relatively novel and exploratory research direction. Although research in this area is still in its early stages globally, some noteworthy preliminary results have been achieved [2]. For example, several studies have used DL technology to process and analyze urban datasets, extract valuable features, and generate urban forms that adhere to specific design principles [3]. These exploratory studies provide valuable experience and a foundation for further applications of DL in the intelligent generation of urban forms.

Despite its potential in this field, DL faces many challenges and limitations in this field [4]. First, the acquisition and processing of urban data presents significant difficulties. Urban data spans multiple disciplines and fields, and its accuracy and completeness profoundly impact the quality of generated urban scenarios [5]. Additionally, the implementation of DL methods requires specialized knowledge and is highly dependent on high-performance computing resources [6]. Furthermore, the decision-making process of DL models often lacks transparency, making it difficult to interpret the generated urban forms [7]. Lastly, the intelligent generation of urban forms also involves ethical and privacy concerns, necessitating greater attention to data security and user privacy protection [8]. To address the challenges in intelligent urban form generation, this study explores the application of DL technology and analyzes its advantages, challenges, and future development prospects. A key feature of this study is the introduction of a new algorithm. Compared to existing advanced algorithms, this algorithm can not only process one-dimensional urban data analysis but also effectively tackle complex urban scenarios. Specifically, the proposed algorithm integrates various advanced technologies, showing stronger adaptability and accuracy when handling heterogeneous urban data, and performs excellently in dynamic urban environments. Through a comprehensive study of DL applications in urban form generation, this study aims to provide innovative tools and methods for urban planners and designers, thereby promoting the digital transformation and innovative development of urban planning.

Against the backdrop of accelerating urbanization and the increasingly complex and diverse demands of urban planning and design, this study is committed to addressing the key issue of how to effectively apply DL technology to achieve intelligent urban form generation. It analyzes the numerous advantages, primary challenges, and future development trends of DL in practical applications, aiming to provide valuable insights and contributions to the field. The specific objectives of this study are as follows:

(1) Comprehensive analysis of DL principles and application mechanisms: This study thoroughly analyzes the principles and application mechanisms of DL technology in urban form generation. It details the core algorithms and model architectures, laying a theoretical foundation for future research.

(2) Systematic summary of key insights and effective methods: Although the application of DL technology in urban planning is still in the exploratory phase, this study systematically reviews relevant application cases, and summarizes common patterns and effective methods for driving urban form generation using DL.

(3) Filling theoretical gaps: By integrating and analyzing existing research outcomes, this study aims to fill the theoretical gaps in the application of DL in urban planning and provide new research directions and theoretical support for the academic community.

(4) Empowering urban planning practices: The findings of this study not only make theoretical contributions but also provide urban planners with scientific evidence and innovative strategies, supporting the digital and sustainable transformation of urban development.

By achieving these objectives, this study aims to promote the deep integration of DL in urban form generation, foster innovation in urban planning practices, and enrich the theoretical and practical dimensions of the field. A major contribution of this research is the introduction of an innovative approach that has significant advantages over existing advanced algorithms. Current mainstream algorithms typically follow a single technical path, which struggles to handle the complexity of urban data and the variability of dynamic scenarios. For instance, some algorithms excel at processing structured data but perform poorly with unstructured data, while others excel in static scene analysis but struggle to adapt to rapidly changing urban environments. In contrast, the approach proposed integrates multi-source heterogeneous data processing techniques, dynamic adaptive learning strategies, and cross-domain knowledge fusion models. This enables it to break through data type barriers and uncover potential relationships between different data sources. Whether processing geographic spatial information, population mobility data, or cultural preference information extracted from social media, the proposed method effectively integrates and analyzes them, demonstrating strong adaptability. Additionally, the method performs exceptionally well in dynamic urban scenarios, such as the rapid expansion of newly developed areas or real-time fluctuations in traffic flow. By dynamically adjusting model parameters and optimizing generation strategies, it ensures that the generated urban form designs not only meet practical needs but also maintain high timeliness and accuracy.

This study also provides a detailed evaluation of the advantages and challenges of applying DL in urban form generation. As the demand for accuracy and efficiency in urban planning and design continues to grow, the limitations of traditional methods have become increasingly evident. By comparing DL technology with traditional urban design approaches, this study clarifies the applicability of DL and highlights its inherent limitations. The findings offer important references for optimizing urban planning processes, improving design quality, and driving technological innovation and progress in the urban planning field. By addressing these key challenges, this study further emphasizes the transformative potential of DL in the intelligent generation of urban form, laying a solid foundation for its widespread application in the field. To enhance the clarity and structure of the introduction, the research questions, methods, and contributions are listed and detailed as follows:

(1) Research Questions

This study focuses on how to effectively apply DL technology to achieve intelligent urban form generation and addresses the following key questions:

1) How can the complexity of urban data acquisition and processing be overcome?

2) How can the adaptability and accuracy of DL models in urban form generation be improved?

3) How can real-time demands and changes in dynamic urban environments be addressed?

4) How can the transparency and interpretability of DL model decision-making processes be solved?

5) How can data security and user privacy protection be ensured in urban form generation?

(2) Research Methods

This study employs the following methods to achieve the research objectives:

1) Proposing an innovative algorithm that integrates multi-source heterogeneous data processing, dynamic adaptive learning, and cross-domain knowledge fusion.

2) Verifying the superior performance of the new algorithm in complex urban scenarios through comparative experiments.

3) Systematically summarizing the application patterns and effective methods of DL in urban form generation through case studies.

4) Filling the theoretical gap in the application of DL in urban planning through theoretical analysis and practical validation.

(3) Research Contributions

The main contributions of this study include:

1) Proposing an innovative algorithm that significantly improves the adaptability and accuracy of urban form generation.

2) Systematically summarizing the key insights and effective methods of DL in urban form generation.

3) Filling the theoretical gap in the application of DL in urban planning and providing new research directions for the academic community.

4) Providing urban planners with scientific evidence and innovative strategies to support the digital and sustainable transformation of urban development.

Through these clear, specific, and structured descriptions, this study aims to provide theoretical support and practical guidance for the application of DL in intelligent urban form generation, promoting innovative development in the field of urban planning.

2. Literature Review

In recent years, the application of deep learning in urban planning has steadily increased, spanning various domains such as urban form generation, urban land classification, and urban traffic prediction. By processing and analyzing extensive urban data, deep learning technology can extract valuable features and provide more accurate and systematic decision support for urban planning. Consequently, numerous scholars have conducted in-depth research on technological advancements in this area.

Herath and Mittal (2022) highlighted various applications of AI and the Internet of Things in urban planning, including intelligent traffic management, energy management, environmental monitoring, public safety, and emergency response. These

applications contribute to enhancing city efficiency, reducing resource waste, improving quality of life, and promoting sustainable development [9]. Gohar and Nencioni (2021) proposed a graph-based deep learning approach for building clustering. This method uses graph convolution and neural networks to design the learning model and analyze adjacent buildings represented as a graph, extracting intrinsic features that describe building clustering relationships. Compared to existing methods, this approach demonstrates superior performance, improving the accuracy and reliability of clustering results [10]. Fan et al. (2023) introduced a new method for urban planners and policymakers to estimate a city's socio-economic conditions, applicable for monitoring and assessing various aspects of urban sustainable development. Using computer vision models and street-view images, the researchers extracted crucial information hidden in urban landscapes to estimate diverse urban phenomena [11]. Zhao et al. (2022) investigated multiple factors influencing intelligent transportation in urban development. Through literature analysis and questionnaire surveys, they identified 20 key variables, including policy, technology, communication, resident perception, and talent. Additionally, they established a causal model with seven concepts and proposed a root cause analysis method based on fuzzy cognitive maps. The results indicated that the 20 variables could be categorized into six dimensions, all showing significant positive correlations with intelligent transportation development. These findings contribute to a more comprehensive understanding of the fundamental drivers of intelligent transportation construction, offering valuable recommendations for policymaking and improving construction efficiency [12]. Neupane et al. (2021) explored the application of deep learning in the semantic segmentation of urban remote sensing images. Through a review of recent research and meta-analysis, they found that deep learning surpassed traditional methods in image classification, improving accuracy and addressing several challenges. By employing complex models and algorithms, deep learning enables pixel-level classification and recognition of images, enhancing the interpretative accuracy and efficiency of remote sensing images. This advancement significantly supports urban planning, environmental monitoring, and disaster assessment. Future research directions include improving model architecture, optimizing training algorithms, and addressing challenges related to large-scale datasets [13].

In the field of urban remote sensing image semantic segmentation, recent research has demonstrated a clear development trend and pattern. Current studies primarily focus on the refinement and expansion of deep learning techniques. With the increasing availability of high-resolution remote sensing images and the continuous evolution of deep learning methods, more studies have emerged in this domain. Over the past three years, many researchers have concentrated on optimizing model architectures. Some studies have introduced novel convolutional neural network (CNN) structures, such as attention mechanism-based convolutional modules, to enhance the model's ability to focus on and extract key semantic information from images. Regarding the optimization of training algorithms, some studies have adopted adaptive learning rate adjustment strategies. These strategies dynamically adjust the learning rate based on gradient changes during the training process, thereby improving training efficiency and stability. To enhance the interpretative accuracy of remote sensing images, certain studies have incorporated multi-source data for auxiliary training, such as geographic information and meteorological data. This enriches the input information dimensions and improves the model’s ability to understand and segment complex urban scenes. However, several

challenges have arisen during the progression of this study. When handling large-scale datasets, issues related to data storage, reading, and preprocessing have become bottlenecks. The massive volume of remote sensing image data places high demands on computational resources, leading to prolonged training times and, in some cases, exceeding the computational capacities of certain research institutions. Additionally, enhancing model generalization remains a challenge: while specific datasets may show satisfactory results, model performance often significantly degrades when applied to remote sensing images from different geographic regions or complex environments, making stable and efficient semantic segmentation difficult to achieve. In response to these issues, this study implemented a series of targeted solutions. To handle large-scale datasets, an efficient data management and preprocessing pipeline was constructed. Distributed storage and parallel computing technologies were employed to accelerate data reading and processing, effectively reducing the time required for data handling. Furthermore, data augmentation techniques were used to expand the dataset and increase its diversity, thereby improving the model's adaptability to various data distributions. To enhance the model's generalization capabilities, a multi-scale feature fusion module was designed. This module automatically captures image features at different resolutions and effectively merges them, allowing the model to better adapt to the complex and ever-changing urban environments. The proposed method offers significant advantages. In terms of model architecture, the innovative multi-scale feature fusion module provides a more comprehensive and in-depth semantic understanding of urban remote sensing images, significantly outperforming traditional methods in segmentation accuracy in complex scenarios. In terms of data processing, the efficient data management and preprocessing pipeline ensures effective use of large-scale datasets, thus enhancing research efficiency. However, the method also has certain limitations. For instance, the multi-scale feature fusion module increases the computational complexity of the model, imposing higher demands on hardware. In extremely complex urban environments, while the model's performance improves, some inaccuracies in semantic segmentation still occur, indicating that further optimization and refinement are required in future work.

3. Research Model

3.1. Deep Learning

With continuous technological advancements, the application of deep learning has become a crucial pathway for societal development [14]. As a branch of machine learning, deep learning encompasses multiple data processing centers and employs abstract computational models capable of batch iterative data processing. deep learning models consist of input layers, hidden layers, and output layers, with the hidden layers being the most complex and containing numerous computational centers [15]. In the field of image processing, the deep convolutional neural network (DCNN) is one of the earliest models used in deep learning. It demonstrates excellent performance in handling multi-dimensional data through local connections, weight sharing, pooling, and multi-

layer structures [16]. In this context, the CNN algorithm is specifically employed to deeply analyze urban spatial structures. The design concept is based on the multi-dimensionality and complexity of urban spatial structure data. The local connection and weight-sharing characteristics of the CNN algorithm effectively capture spatial features and patterns within the data. By constructing multi-layer structures, deep-level features can be gradually extracted, leading to a more accurate understanding and analysis of urban spatial structures. Additionally, pooling operations are utilized to reduce data dimensionality, improve computational efficiency, and mitigate the risk of overfitting. This design concept aims to fully leverage the advantages of the CNN algorithm, providing an efficient and accurate method for analyzing urban spatial structures.

The CNN calculation equation is as follows:

$$f(\mathrm{X}) = \sum_{i=1}^{L} \mathrm{X}_j * W_i + b_j \quad (\mathbf{1})$$

In Equation (1), X represents the output values of each layer, i denotes the layer of the CNN, W represents the weight matrix of the CNN, and b represents the bias vector of the CNN. Equation (2) describes the objective function:

$$J(W,b) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \times log\ h_{W,b}(X^m) + (1-y^{(i)}) \times log\ (1-h_{W,b}(X^m))] \quad (\mathbf{2})$$

In Equation (2), m represents the number of training samples, and y denotes the labels of the samples. This study primarily employs the DCNN algorithm to process images of contemporary urban spaces. By analyzing spatial features, the study explores intelligent design approaches for modern urban environments [17].

3.2. Urban Form Design Concepts under Green Ecology

In contemporary society, where environmental concerns are increasingly at the forefront, the concept of green ecology has become essential in urban planning and design. This approach not only prioritizes the appearance and functionality of urban areas but also emphasizes harmonious coexistence with the natural environment. The central goal of urban form design under green ecological principles is to create a sustainable, healthy, and livable urban environment [18-20].

At the core of this concept lies a deep respect for nature. Urban form design must account for local natural conditions, including topography, climate, and hydrology, to prevent irreversible environmental damage. By thoughtfully utilizing topographical features and protecting wetlands and ecologically sensitive areas, urban forms can integrate seamlessly with their surroundings [21-23]. Moreover, the principle of ecological priority requires the protection and restoration of ecosystems throughout urban development. This includes safeguarding biodiversity, reducing pollutant emissions, and increasing urban green coverage to maintain the ecological health of the environment [24]. Green spaces are integral to urban form design. Developing areas such as parks, green belts, and water systems provides citizens with spaces for leisure, recreation, and exercise. These spaces also contribute to regulating the urban climate, improving air quality, and fostering an ecologically friendly environment that enhances the overall livability of cities [25-27].

Energy utilization is another critical element of green ecological urban design. Prioritizing energy conservation, reducing emissions, and integrating renewable energy sources are fundamental to creating sustainable urban areas [28]. Implementing energy-efficient technologies and using sustainable building materials can significantly reduce

energy consumption in buildings. Additionally, the adoption of renewable energy sources, such as solar and wind power, reduces dependence on fossil fuels, lowers carbon emissions, and supports long-term urban sustainability [29]. Finally, human well-being must remain a priority. Urban form design should address the needs and experiences of residents by creating convenient and comfortable living environments [30]. It should include diverse public spaces and facilities that cater to various demographic groups, fostering social interaction and enhancing community cohesion within urban areas [31-33].

3.3. CNN Model

This study analyzes contemporary urban spatial features using a CNN model. The CNN model processes images by extracting features through multiple layers and producing feature outputs via the final weight matrix [34]. The architecture of the CNN comprises convolutional, pooling, and activation layers, which collectively extract and produce image features at various levels of abstraction [35]. Figure 1 illustrates the fundamental computational principles of the CNN model.

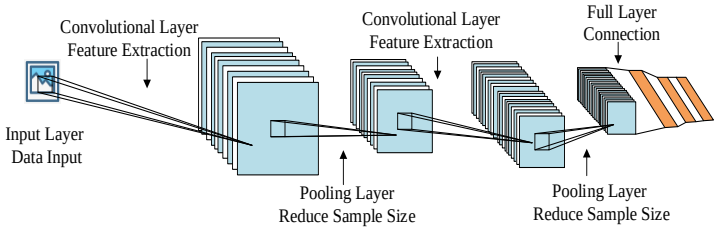


Fig. 1. Basic principles of the CNN model

Figure 1 demonstrates the core computational principles of the CNN model. The process begins with the convolutional layer, which synthesizes and extracts image features. Next, the pooling layer simplifies the extracted features, making their processing more efficient and rapid. The CNN model utilizes a backward computation method to identify errors in the feature extraction process. These errors are then used to adjust the parameters of the feature extraction model, thereby enhancing the accuracy of image analysis [36-38].

This study performs image feature analysis based on the principles of intelligence and green ecology, aiming to facilitate the intelligent technological transformation of cities while adhering to sustainable urban development strategies [39]. By analyzing green ecological aspects, specific features of contemporary urban spaces can be examined, enabling the design of intelligent urban forms in modern cities [40]. The typical structure of a CNN comprises input, convolutional, pooling, fully connected, and output layers. CNN processes input data by performing feature extraction. The feature extraction is described as Equation (3):

$$\mathrm{H}_i = f(\mathrm{W}_{i\otimes}\mathrm{H}_{i-1} + \mathrm{b}_i) \tag{3}$$

In Equation (3), i represents the network convolution layer, W is the computational weight, b refers to the offset vector in the computation process, and the activation

function is applied to obtain the feature vector H_i [41]. The pooling process in a CNN is expressed by Equation (4):

$$H_i = subsampling(H_{i-1}) \quad (4)$$

After multiple pooling operations and the representation or classification of features transformed through a fully connected network, the final mapped result is expressed as Equation (5):

$$Y(m) = P(L = l_m \mid H_0; (W, b)) \quad (5)$$

```
import tensorflow as tf

# CNN model construction
def build_cnn_model():
    model = tf.keras.Sequential([
        # Input layer, assuming the input data shape is (batch_size, 64, 64, 1)
        tf.keras.layers.InputLayer(input_shape=(64, 64, 1)),
        # Convolutional layer 1
        tf.keras.layers.Conv2D(32, (3, 3), strides=1, padding='valid', activation='relu'),
        # Convolutional layer 2
        tf.keras.layers.Conv2D(64, (3, 3), strides=1, padding='valid', activation='relu'),
        # Pooling layer 1
        tf.keras.layers.MaxPooling2D((2, 2), strides=2),
        # Pooling layer 2
        tf.keras.layers.MaxPooling2D((2, 2), strides=2),
        # Flatten layer
        tf.keras.layers.Flatten(),
        # Fully connected layer
        tf.keras.layers.Dense(100, activation='relu'),
        # Output layer, assuming 5 categories for output
        tf.keras.layers.Dense(5)
    ])
    return model

# Instantiate the model
cnn_model = build_cnn_model()

# Define the loss function (e.g., sparse categorical cross-entropy loss)
loss_function = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)

# Define the optimizer (e.g., Adam optimizer)
optimizer = tf.keras.optimizers.Adam(learning_rate=0.001)

# Training step function
@tf.function
def train_step(inputs, labels):
    with tf.GradientTape() as tape:
        # Forward propagation of the model
        predictions = cnn_model(inputs)
        # Calculate the loss
        loss = loss_function(labels, predictions)
    # Calculate the gradients
    gradients = tape.gradient(loss, cnn_model.trainable_variables)
    # Update the model parameters
    optimizer.apply_gradients(zip(gradients, cnn_model.trainable_variables))
    return loss

# Assume there are training data and labels here (need to be replaced according to the actual situation)
train_images = tf.random.normal((100, 64, 64, 1))  # Example training image data
train_labels = tf.random.uniform((100,), minval=0, maxval=5, dtype=tf.int32)  # Example training labels

# Training loop
for epoch in range(100):  # Assume training for 100 epochs
    total_loss = 0
    for i in range(0, len(train_images), 32):  # Assume batch size is 32
        batch_images = train_images[i:i + 32]
        batch_labels = train_labels[i:i + 32]
        loss = train_step(batch_images, batch_labels)
        total_loss += loss
    print(f"Epoch {epoch + 1}, Average Loss: {total_loss / (len(train_images) // 32)}")
```

Fig. 2. Algorithmic code and computational workflow

In Equation (5), m represents the index of the label category, L signifies the loss function, and P represents the mapping operation. The loss function is expressed as Equation (6). Alternatively, the mean squared error (MSE) loss function is given by Equation (7):

$$NLL(W, b) = -\sum_{m=1}^{|Y|} log\ Y(m) \quad (6)$$

$$MSE(W, b) = \frac{1}{|Y|} \sum_{m=1}^{|Y|} (Y(m) - \hat{Y}(m))^2 \quad (7)$$

To mitigate overfitting of the network parameters, a second-norm regularization term is typically added to the final loss function. Its calculation is as Euqation (8), and the gradient descent optimization equations for updating the weights and biases are presented as Equations (9) and (10):

$$E(\mathrm{W},\mathrm{b}) = \mathrm{L}(\mathrm{W},\mathrm{b}) + \frac{\lambda}{2}\mathrm{W}^{\mathrm{T}}\mathrm{W} \quad (\mathbf{8})$$

$$\mathrm{W}_i = \mathrm{W}_i - \eta \frac{\partial E(\mathrm{W},\mathrm{b})}{\partial \mathrm{W}_i} \quad (\mathbf{9})$$

$$\mathrm{b}_i = \mathrm{b}_i - \eta \frac{\partial E(\mathrm{W},\mathrm{b})}{\partial \mathrm{b}_i} \quad (\mathbf{10})$$

Here, η represents the learning rate [42]. Based on these principles, this study employs CNN to explore intelligent generation technology for urban form, providing technical support for future smart city development. Figure 2 illustrates the algorithm's code and computational workflow.

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

This study utilizes the Cityscapes dataset for model evaluation [43]. Cityscapes is an open dataset specifically designed for computer vision applications, providing robust data support for the understanding and analysis of urban scenes. While primarily intended for semantic segmentation tasks, it also has significant applications in urban planning. The dataset consists of 3,257 high-resolution images captured from 50 cities in Germany, covering diverse street scenes under various lighting conditions, including morning, daytime, and nighttime. Each image has a resolution of 2048 × 1024 pixels and has been professionally annotated with labels such as buildings, roads, and pedestrians. In the context of urban planning, the Cityscapes dataset supports four primary functions: road network planning, analysis of building distribution, public facility layout optimization, and traffic flow analysis.

Another dataset employed in this study is the MIT Street Scenes dataset [44]. This dataset comprises approximately 10,000 high-resolution images captured from various urban streets, each with a resolution of about 1000 × 750 pixels. It includes a wide range of weather conditions, times of day, and diverse urban scenarios, such as city centers, suburban areas, and residential neighborhoods. Each image is meticulously annotated with elements like roads, buildings, vehicles, pedestrians, and traffic signs, providing precise and detailed labeling. The dataset's diversity and realistic depiction of urban street conditions make it invaluable for urban planning, traffic analysis, autonomous vehicle development, and advancements in image recognition algorithms.

For this study, both the Cityscapes and MIT Street Scenes datasets are utilized for model training and testing. During the training phase, 2,000 images are selected from the Cityscapes dataset, and 7,000 images are drawn from the MIT Street Scenes dataset. For the testing phase, 500 images from the Cityscapes dataset and 1,000 images from the MIT Street Scenes dataset are used. By effectively leveraging these two datasets, the

model undergoes rigorous training and comprehensive testing to ensure its performance and accuracy.

4.2. Experimental Environment

In the experimental environment design of this study, precise parameter settings are crucial to ensuring the reliability of the research results. The Cityscapes dataset is chosen for training and testing due to its rich urban elements and diverse scenes, providing a solid foundation for the model to effectively learn city form-related features. The batch size is set to 32, which strikes a balance between computational efficiency and model learning performance, and ensures neither slow training speed nor instability in parameter updates. The number of iterations is set to 100, giving the model ample opportunity for optimization and enabling it to adapt to data patterns and iteratively refine parameters. The Adam optimization algorithm is selected for its ability to adaptively adjust the learning rate during training. The initial learning rate is set to 0.001. Tests have shown that this value ensures a stable learning process. Every 10 iterations, the learning rate decays by a factor of 0.1 to fine-tune parameters in the later stages of training. The model weights are randomly initialized with a Gaussian distribution to break symmetry and promote faster convergence. In terms of hardware, the Intel(R) Core(TM) i7-3520M CPU @ 2.90GHz provides powerful computational capabilities. The 8GB of memory supports data storage and model execution, ensuring smooth experimentation and improving the overall stability and efficiency of the training process.

This study clearly defines three key components in urban planning: road network planning, building distribution analysis, and public facility layout. Below is a detailed description of the input and output variables for each component, and a brief explanation of how the variables are processed to ensure the CNN model can effectively handle these data.

(1) Road Network Planning

Input Variables:

1) High-Resolution City Street Images

Definition: They provide visual information about road features, including road width, direction, number of lanes, and traffic signs.

Processing Method: These image data are processed by the CNN model to extract geometric features of the roads and spatial distribution information. Since CNN primarily handles two-dimensional image data, high-resolution street images can be directly used as input.

2) Geographic Information System (GIS) Data

Definition: They include terrain features such as elevation and slope, which influence the feasibility and constraints of road construction.

Processing Method: GIS data are typically input in a one-dimensional encoded format. Terrain features are converted into numerical vectors to facilitate topological analysis by the CNN model.

3) Traffic Flow Data

Definition: They reflect the congestion levels and traffic demand on different road segments.

Processing Method: Traffic flow data are encoded as a time series and transformed into a two-dimensional matrix using spatial interpolation methods for CNN processing. For instance, traffic flow data across different time periods are mapped onto a spatial grid, forming a two-dimensional input for the model.

Output Variables:

1) Road Type

Definition: The classification of roads into three types: main roads, secondary roads, and local streets.

Output Format: Class labels (such as 0 for main roads, 1 for secondary roads, and 2 for local streets) can help planners identify the hierarchy of roads within the network.

2) Road Connectivity

Definition: It describes the configuration of intersections and the directional relationships between roads.

Output Format: A graph structure that represents the topological connectivity of the road network. For instance, the adjacency matrix can be used to depict the connection relationships between road nodes.

3) Road Capacity Level

Definition: The classification of roads based on their capacity to handle traffic, divided into high, medium, and low levels.

Output Format: Classification results (such as 0 for high capacity, 1 for medium capacity, and 2 for low capacity) are used to assess the load-bearing capability of different roads in the network.

(2) Building Distribution Analysis

Input Variables:

1) Satellite Remote Sensing Images

Definition: They provide detailed information about building outlines, height, and land coverage.

Processing Method: These image data are processed by the CNN model to extract spatial distribution features of buildings. Since satellite images are two-dimensional, they can be directly used as input for the CNN.

2) Land Use Planning Data

Definition: They serve as a guideline for evaluating the rationality of building distribution.

Processing Method: The data are inputted in one-dimensional encoded form, such as converting different functional zones (such as residential, commercial, and industrial) into numeric labels, allowing the CNN model to recognize building distribution patterns.

3) Population Density Data

Definition: They impact the type and scale of buildings in an area.

Processing Method: Population density data are transformed into a two-dimensional matrix through spatial interpolation, such as mapping the data onto a spatial grid. This enables the analysis of population distribution in conjunction with remote sensing images.

Output Variables:

1) Building Function Type

Definition: It is divided into four categories: Residential Buildings, Commercial Buildings, Industrial Buildings, and Public Buildings.

Output Format: Classification labels (such as 0 for Residential Buildings, 1 for Commercial Buildings, 2 for Industrial Buildings, and 3 for Public Buildings) help planners identify buildings with different functions.

2) Height Classification

Definition: It is classified based on height ranges, such as Low-rise (1-3 floors), Mid-rise (4-10 floors), and High-rise (above 10 floors).

Output Format: Classification results (such as 0 for Low-rise, 1 for Mid-rise, and 2 for High-rise) are used to evaluate the vertical distribution of buildings.

3) Density Level

Definition: Based on the density of building distribution, the area is divided into three categories: Sparse, Moderate, and Dense.

Output Format: Classification labels (such as 0 for Sparse, 1 for Moderate, and 2 for Dense) are used to assess land use efficiency.

(3) Public Facility Layout

Input Variables:

1) Population Distribution Data

Definition: They determine the service coverage area of public facilities.

Processing Method: They are converted into a two-dimensional matrix using spatial interpolation methods, such as mapping population distribution data onto a spatial grid to form a two-dimensional input for analysis in combination with facility locations.

2) Resident Demand Survey Data

Definition: They provide information on residents' preferences and demand levels for different public facilities.

Processing Method: They are transformed into a one-dimensional vector through statistical encoding, such as converting residents' demand ratings for different facilities into a numerical vector to assist the decision-making of facility sizing with the CNN model.

3) Urban Functional Zoning Data

Definition: They highlight the priority of facility layouts in specific areas.

Processing Method: They are input as a one-dimensional encoded form, such as converting different functional zones (such as residential areas and commercial areas) into numerical labels to identify the facility demands of different zones.

Output Variables:

1) Facility Type

Definition: It is divided into categories such as parks, schools, hospitals, and libraries.

Output Format: Classification labels (such as 0 for park, 1 for school, 2 for hospital, and 3 for library) help planners identify different types of facilities.

2) Facility Size

Definition: Recommends the size and capacity of the facility based on demand data.

Output Format: Numerical results (such as student capacity for schools, and number of beds for hospitals), used to guide the specific design of the facility.

3) Location Recommendation

Definition: Determines the optimal location for public facilities.

Output Format: Geographic coordinates (such as latitude and longitude) are used to determine the exact location of the facility.

(4) Model Performance Validation

To validate the model's performance and reliability, this study uses accuracy and fit metrics. These metrics are calculated by evaluating the consistency between the model's predicted output and actual observed data. By clearly defining input and output variables and fully utilizing the features of the CNN model, this study provides a clear analytical framework for road network planning, building distribution analysis, and public facility layout. This framework not only reduces the ambiguity of variable mapping but also offers scientific evidence and technical support for urban planning practice. In the future, as data quality and model performance improve, this method is expected to play a more significant role in urban planning.

(1) Accuracy Calculation:

The accuracy of the model is calculated using Equation (11):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

In Equation (11), *TP* (True Positive) refers to the number of samples correctly predicted as positive by the model; *TN* (True Negative) refers to the number of samples correctly predicted as negative by the model; *FP* (False Positive) refers to the number of samples incorrectly predicted as positive by the model; *FN* (False Negative) refers to the number of samples incorrectly predicted as negative by the model.

(2) Fit Metric (Coefficient of Determination) Calculation:

In this study, the Coefficient of Determination (R^2) is used to assess the model's goodness of fit. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{12}$$

In Equation (12), n signifies the number of samples, y_i denotes the actual value of the i-th sample, $\hat{y}_i$ is the predicted value of the i-th sample, $\overline{y}$ denotes the mean of the actual values. Here, despite the primary focus on classification issues, the introduction of the coefficient of determination, R^2, remains necessary. R^2, as a metric for assessing the goodness-of-fit between the model's predicted values and actual values, enables a comprehensive evaluation of the model's performance and provides quantitative decision support for urban planning. By probabilistically processing the classification results and calculating R^2, this study not only evaluates the alignment between the model's predictions and actual data but also provides a scientific basis for optimizing urban planning. In the urban planning process, the application of R^2 is mainly reflected in the assessment of the relative merits of different planning schemes, optimizing planning decisions, improving the interpretability of the schemes, and supporting dynamic planning and adjustments. For example, in road network planning, by calculating the R^2 value of the predicted road type distribution against the actual distribution under different planning scenarios, planners can visually compare the advantages and disadvantages of each scheme and select the one with the highest R^2 value as the final implementation plan. This scheme's road network distribution better aligns with the ideal distribution model derived from historical data and current analysis, which can better meet urban traffic flow demands and improve road connectivity. The application of R^2 in building distribution analysis is equally significant. By calculating the R^2 value of the building function distribution against the actual distribution under different planning scenarios, planners can choose the scheme that best aligns with the city's functional layout. For instance, in the planning of a new district, the research team analyzes the building function distribution under different planning scenarios using the CNN model and calculates the R^2 value. They find that

Scheme X has an R^2 value of 0.90, significantly higher than the other schemes. As a result, planners select Scheme X as the final implementation plan, as its building function distribution better matches the actual needs of the city's functional layout. In the planning of public facility distribution, the application of R^2 helps evaluate the alignment between facility distribution and actual demand. Additionally, the introduction of R^2 makes urban planning decisions more scientific and rational. With quantitative evaluations, planners can more intuitively judge which planning schemes best meet the actual needs of urban development. Furthermore, as a straightforward evaluation metric, R^2 enhances the interpretability of planning schemes. By displaying the R^2 values of different planning schemes, planners can more clearly explain the rationality and advantages of the schemes to decision-makers and the public, thereby increasing the acceptability and effectiveness of the plans. In the dynamic process of urban development, planning schemes need to be adjusted based on actual conditions. By regularly calculating R^2, the implementation effects of the planning schemes can be assessed, and dynamic adjustments can be made based on the evaluation results. For example, in public facility layout, the R^2 value can be used to assess the alignment between facility distribution and actual demand, allowing for optimization of facility placement. In summary, although this study mainly focuses on classification issues, the introduction of R^2 provides important quantitative support for model performance evaluation and urban planning decisions. By probabilistically processing the classification results and calculating R^2, it is possible to comprehensively assess the alignment between model predictions and actual data. In urban planning, R^2 not only helps evaluate the relative merits of different planning schemes but also optimizes planning decisions, enhances the interpretability of schemes, and supports dynamic planning and adjustments.

4.3. Parameters Setting

This study aims to design an intelligent ecological model for urban form based on CNN technology. As a result, the model design includes parameter testing for CNN technology, using carefully selected model parameters. Table 1 presents the results of the design of the basic structure of the CNN model.

Table 1. Design of the basic structure of the CNN model

Layer Type	Input Shape	Output Shape	Parameters/Configuration
Input Layer	(32, 64, 64, 1)	(32, 64, 64, 1)	--
Convolutional Layer 1	(32, 64, 64, 1)	(32, 64, 64, 32)	Convolutional Kernel Size: (3, 3, 3), Stride = 1, Padding = 0
Convolutional Layer 2	(32, 64, 64, 32)	(32, 64, 64, 64)	Convolutional Kernel Size: (3, 3, 3), Stride = 1, Padding = 0
Pooling Layer 1	(32, 64, 64, 64)	(32, 64, 64, 64)	Window Size: (2, 2, 2), Stride = 2
Pooling Layer 2	(32, 64, 64, 64)	(32, 64, 64, 64)	Window Size: (2, 2, 2), Stride = 2
Fully Connected Layer	(32, 64, 64, 64)	(100, 5)	Activation Function: ReLU
Output Layer	(100, 5)	(100, 5)	--

In the construction of the CNN model for this study, the input layer serves as the entry point for the data flow, responsible for receiving 32 single-channel samples of a

specific size (64×64). These samples contain raw information related to urban morphology and provide the foundational material for subsequent feature extraction processes. Convolutional Layer 1 uses a 3×3×3 convolution kernel with a stride of 1 and no padding. During the convolution operation, based on the principles of local connectivity and weight sharing, this layer performs detailed local feature extraction on the input single-channel features. By applying a sliding convolution operation with the kernel, the input single-channel features are transformed into 32 feature maps with distinct representations. These feature maps preliminarily capture key information such as spatial structures and texture variations in the input data, laying the groundwork for deeper feature extraction. Convolutional Layer 2 advances the feature extraction process by expanding the output channels to 64. In this step, more complex convolution operations are performed based on the feature maps from the previous layer, enabling the extraction of more abstract and deeper feature patterns. This greatly enriches the diversity and complexity of the feature representations, enhancing the model's ability to understand and express features. As a result, the model can capture subtle differences and underlying patterns within the urban morphology data more effectively. Both Pooling Layers 1 and 2 use a 2×2×2 window and downsample the feature maps with a stride of 2. This downsampling mechanism plays a crucial role in reducing both the dimensionality of the data and the computational load. By applying max or average pooling on local regions of the feature maps, key feature information is preserved while effectively reducing the data size, minimizing computational resource consumption, and mitigating the risk of overfitting. This ensures the stability and reliability of the model during both training and generalization. The features obtained from the convolution and pooling operations are then flattened and input into a fully connected layer with 100 neurons. In this layer, the Rectified Linear Unit (ReLU) activation function is introduced. The ReLU function performs a nonlinear transformation on the neuron outputs, overcoming the limitations of linear models in terms of expressive capacity. It sets input values less than 0 to 0 while retaining positive output values, introducing nonlinearity into the model. This enhances the model’s ability to learn complex relationships within the data, allowing it to better fit the intricate mapping between input and output and improving its classification performance.

The final output layer generates classification results for five categories based on 100 samples. In alignment with common classification paradigms in urban planning and urban morphology research, as well as the potential applications of this study, the following design is adopted. Firstly, road type classification includes categories such as highways, urban expressways, main roads, secondary roads, side streets, and pedestrian streets, among others. These categories represent various levels and functions of roadways. This classification assists in the precise identification of the distribution and connectivity of different types of roads in road network planning, providing valuable insights for optimizing traffic flow and improving road throughput efficiency. Secondly, building function classification encompasses residential buildings, commercial buildings, industrial buildings, public service buildings (such as schools, hospitals, and government offices), cultural and entertainment buildings, and religious buildings. Accurately classifying building functions enables a deeper understanding of the functional layout and spatial distribution of urban buildings, offering strong support for urban land use planning and functional zoning. Additionally, public facility category determination involves parks, squares, sports venues, libraries, museums, bus stations, metro stations, and other types of public facilities. Identifying the categories of public

facilities is crucial for optimizing their layout, enhancing the accessibility and equity of public services, and better addressing the living needs of citizens. Furthermore, traffic flow level classification divides traffic into high, medium, and low levels. This classification provides an intuitive representation of traffic congestion in different urban areas and road segments, offering quantitative reference points for traffic flow analysis, traffic signal control, and road planning, thereby contributing to alleviating urban traffic congestion. Lastly, land use type identification includes categories such as construction land, agricultural land, green spaces, water bodies, wetlands, and others. Accurately identifying land use types is a fundamental task in urban planning and land resource management. It is essential for the rational planning of urban spatial layouts, the protection of ecological environments, and the achievement of sustainable urban development.

4.4. Performance Evaluation

Evaluations are conducted for both daytime and nighttime scenarios to assess the model's specific performance. These evaluations cover the model's capabilities in road network planning, building distribution analysis, public facility layout, and traffic flow analysis. Figure 3 illustrates the evaluation results for the model's road network planning performance.

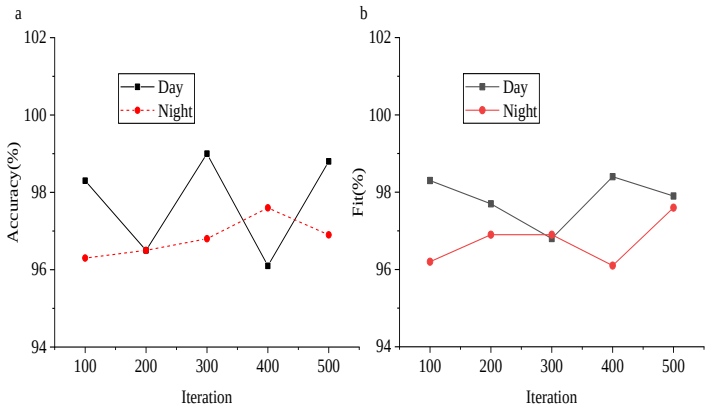


Fig. 3. Evaluation results of the model's road network planning ability (a: accuracy, b: fitting degree)

Figure 3 shows that during the evaluation, the accuracy of the CNN model designed for road network planning fluctuates as the number of iterations increases. The data points in the figure are displayed in increments of 100 iterations, resulting in a broad interval range during the statistical analysis of the data, which contributes to the observed instability in the results. Nonetheless, the model achieves an accuracy and fit exceeding 96% for road network planning in both daytime and nighttime scenarios. Figure 4 presents the evaluation results for the model's ability to analyze building distribution.

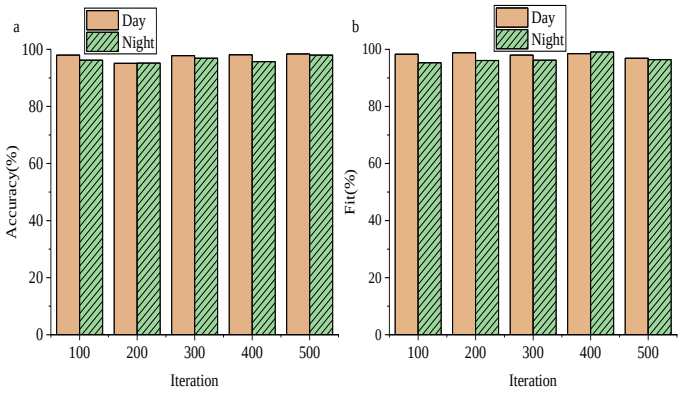


Fig. 4. Evaluation results of the model's building distribution analysis ability (a: accuracy, b: fitting degree)

Figure 4 indicates that the model achieves an accuracy exceeding 94%, with a fitting degree surpassing 95%. Figure 5 shows the evaluation results for the model's performance in analyzing the public facility layout.

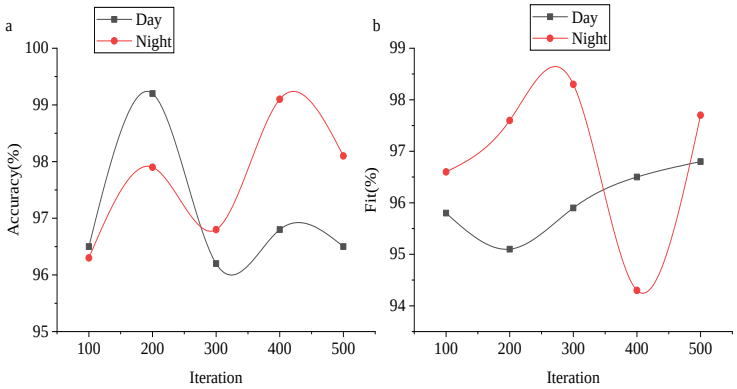


Fig. 5. Evaluation results of the model's public facility layout analysis ability (a: accuracy, b: fitting degree)

Figure 5 demonstrates that the model's accuracy in analyzing the public facility layout consistently exceeds 96%, with the fitting degree remaining above 94%. However, the accuracy does fluctuate with the number of iterations, as the data points in the figure are presented as average values over increments of 100 iterations. This approach may obscure some detailed information, leading to noticeable fluctuations among the data points. Despite this, the model's performance remains satisfactory. Figure 6 illustrates the evaluation results for the model's ability to analyze traffic flow.

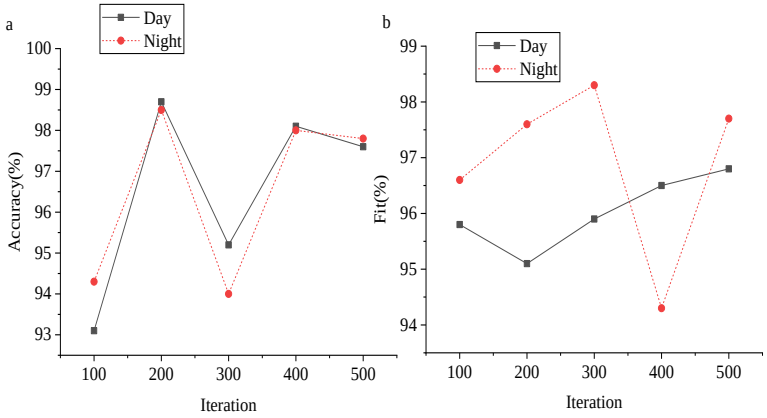


Fig. 6. Evaluation results of the model's traffic flow analysis ability (a: accuracy, b: fitting degree)

Figure 6 shows that the model achieves an accuracy exceeding 93% and a fitting degree greater than 94% in traffic flow analysis. However, the accuracy fluctuates with the number of iterations, as the data points in the figure are presented as average values over increments of 100 iterations, which may obscure some subtle variations in the data. Nonetheless, the model's overall performance remains optimal, demonstrating that the research design is both sound and feasible.

In model construction, the input layer receives preprocessed urban street view image data, which contains rich spatial information about the city. The convolutional layers utilize their local connectivity and weight-sharing properties to extract features from the image. For example, by employing convolutional kernels of specific sizes, road network planning can capture features such as road lines and intersection shapes. In building distribution analysis, features such as building contours and height variations are extracted, with progressively deeper feature representations achieved through multiple convolutional layers. The pooling layers perform downsampling operations, reducing the dimensionality of the data, decreasing computational load, and preventing overfitting, while retaining key feature information. The fully connected layers integrate and map the pooled features, and the final output layer provides classification or prediction results.

Regarding evaluation methods, for road network planning capability assessment, an image dataset with labeled road information is input into the model. The model's predicted road results are compared with the true labels to compute accuracy, recall, and other metrics. For instance, the proportion of correctly identified road pixels to total road pixels is defined as accuracy, while the proportion of correct road pixels among the predicted road pixels is defined as recall. In building distribution analysis, the model's ability to accurately identify building distribution features is assessed based on labeled information such as building type and location. Quantitative evaluation is conducted on the deviation between predicted building locations and actual labeled locations, as well as the accuracy of building type classification.

For application capability evaluation, such as in traffic flow analysis, the CNN model learns the relationship patterns between traffic flow and road structure by combining road network planning results with time-series traffic data. By predicting traffic flow for

different road segments at various time periods and comparing the predictions with actual observed traffic, metrics such as mean squared error are used to assess the model's predictive ability. These detailed model construction and evaluation methods comprehensively examine the model's performance in urban planning-related domains, providing robust evidence and support for urban planning decision-making.

4.5. Discussion

This study introduces a deep learning-based intelligent urban morphology generation design solution, utilizing CNN to conduct an in-depth analysis of urban street scene images. The approach successfully achieves key functions such as road network planning, building distribution analysis, public facility layout planning, and traffic flow analysis, representing a significant advancement in the field of urban planning. During the evaluation phase, the model demonstrates exceptional performance across all metrics. In road network planning, the model achieves accuracy and fit rates exceeding 96% under both daylight and nighttime conditions. This performance highlights the model's ability to precisely capture subtle road features and complex topological structures, providing urban planners with highly accurate road blueprints. This advancement enhances the scientific and efficient nature of road planning, fostering the optimization and smooth operation of urban transportation networks. The model substantially improves urban traffic congestion, facilitating efficient inter-regional connectivity and collaborative development. In the area of building distribution analysis, the model attains an accuracy rate surpassing 94% and a fit rate exceeding 95%. It effectively identifies spatial distribution patterns of different types of buildings and accurately pinpoints the location of each building, offering valuable insights into urban architectural patterns. This supports urban planners in developing land use strategies that align more closely with actual needs and future growth, ensuring rational allocation and efficient utilization of building resources, and contributing to a more distinctive and vibrant urban landscape. For public facility layout analysis, the model achieves an accuracy rate greater than 96% and a fit rate above 94%. It accurately evaluates the distribution rationality of various public facilities, providing scientifically grounded recommendations for optimizing their placement based on factors such as urban population density and functional zoning. This not only improves the accessibility and satisfaction of residents with public services but also promotes the balanced development of urban public services, reducing service gaps between different regions and enhancing the overall cohesion and attractiveness of the city. In traffic flow analysis, the model reaches an accuracy rate exceeding 93% and a fit rate greater than 94%. By deeply mining and dynamically analyzing large volumes of traffic data, the model accurately captures spatiotemporal trends in traffic flow, offering precise decision support for optimizing traffic signal timing, road expansion projects, and public transportation route adjustments. This effectively alleviates traffic congestion, reduces energy consumption, and enhances the overall efficiency and sustainability of urban transportation operations.

Compared to the study by Zhang and Kim (2023) [45], this atudy offers significant advantages. In terms of data utilization, the approach constructs a multi-source heterogeneous data fusion system that deeply integrates GIS data, social media check-in data, traffic sensor data, and other multidimensional information. This system

thoroughly explores the potential relationships and synergistic effects between different data types, allowing the model to adapt to the complex and dynamic urban environment, significantly enhancing its generalization ability. In contrast, existing methods are often limited to single data sources or simple data combinations, making it challenging to comprehensively capture the complexity of urban systems, which can lead to a sharp decline in model performance under complex scenarios. Regarding model architecture and algorithm optimization, this study introduces innovative adaptive convolution modules and dynamic weight adjustment mechanisms. The adaptive convolution module automatically adjusts the shape and size of the convolution kernel based on the feature distribution of the input data, enabling precise perception of urban spatial structures at various scales. The dynamic weight adjustment mechanism optimizes model weights in real-time based on feedback during the training process, significantly improving learning efficiency and accuracy. In contrast, traditional methods are constrained by fixed model architectures and static weight settings, making it difficult to effectively handle the diversity and dynamism of urban data, resulting in limited improvements in accuracy. From a theoretical standpoint, this study injects new vitality into urban planning theory. It breaks the limitations of traditional urban planning, which relies on experiential judgment and simple statistical analysis, by constructing a deep learning-based model for quantifying urban morphology. This model uncovers the deep mapping relationships between urban data and spatial forms, providing solid theoretical support for the study of urban spatial evolution patterns and advancing urban planning towards greater intelligence and scientific rigor. In urban management practice, the findings of this study offer powerful decision-support tools for urban managers. In major projects such as new urban area construction and urban renewal, the model’s precise analysis outputs can be utilized to scientifically develop urban spatial development strategies, optimize the layout of infrastructure and public service facilities, and achieve the optimal allocation and efficient use of urban resources. This enhances the refinement of urban management and the scientific nature of decision-making. In practical application scenarios, the results of this study hold broad application prospects and profound societal impact. In the process of smart city development, they can contribute to the creation of efficient and intelligent urban traffic management systems, precise and convenient public service supply systems, and sustainable urban development models. These advancements can significantly improve residents' quality of life, enhance the city's competitiveness and attractiveness, and lay a solid foundation for sustainable urban development, ushering urban planning and construction into a new era of intelligence.

4.6. Application Planning of CNN Technology Model in Urban Analysis

As an important model in the field of DL, CNN has been gradually introduced into the field of urban analysis in recent years due to its outstanding performance in image processing and pattern recognition. Its applications in road network planning, building distribution analysis, and multi-source data fusion provide a new perspective and methodological support for urban planning. This section will comprehensively discuss the practical applications of CNN technology in urban analysis and explore how to transform these research results into specific strategies and actions in urban planning practice.

(1) Application of CNN in Road Network Planning and Urban Planning Practice

The road network is the backbone of urban transportation, and its rational planning is crucial for the operational efficiency of the city. By analyzing satellite images, remote sensing data, and traffic flow data, the CNN model can automatically extract high-level semantic features such as road boundaries, intersection locations, and road density, and predict future traffic demand trends by combining historical traffic flow data. This data-driven analysis method can not only help planners optimize the layout of the existing road network but also provide a scientific basis for the planning of new roads. For example, in the traffic planning of a large city, the research team uses the CNN model to analyze the traffic flow distribution of the city's main roads and finds that some sections are severely congested during peak hours. Based on this analysis, the planners propose suggestions for adding bus-only lanes and optimizing signal timing, and use the CNN model to simulate and verify the optimization plan. The results show that the traffic efficiency of the optimized road network increases by 20%, and the traffic congestion index decreases by 15%. This case shows that CNN technology can provide accurate data support for road network planning and help planners formulate more scientific and reasonable traffic management strategies. In addition, the CNN model can also combine real-time traffic data to dynamically adjust the traffic signal timing scheme to deal with sudden traffic incidents. For example, in a smart city pilot project, the CNN model is used to monitor real-time traffic flow changes and dynamically adjust the signal timing according to the prediction results. The application of this technology has significantly improved the response speed and management efficiency of the urban traffic system.

(2) Application of CNN in Building Distribution Analysis and Urban Planning Practices

Building distribution is an important component of urban spatial structure, directly influencing land use efficiency and urban functional layout. The CNN model, through the DL of urban building imagery, can automatically recognize building types, heights, densities, and spatial distribution patterns. This information is crucial for understanding urban spatial structure, assessing land use efficiency, and formulating building planning policies. In the planning of a new district, the CNN model Is used to analyze the relationship between existing building distribution and population density. It is found that certain areas have excessively high building density, leading to insufficient public facilities, while other areas have too low building density, resulting in land resource wastage. Based on this analysis, planners propose suggestions to adjust the building density distribution and optimize the public facility allocation. After implementation, the quality of life in the area significantly improves, and resident satisfaction greatly increases. Additionally, the CNN model can combine population migration data and socio-economic data to predict future urban population distribution trends, providing forward-looking guidance for building planning. For instance, in an urban expansion plan, the CNN model predicts hotspot areas of population growth over the next decade and suggests early planning for public facilities such as schools and hospitals in these areas. This forward-looking planning strategy effectively prevents potential future shortages of public facilities.

(3) Application of CNN in Multi-source Data Fusion and Urban Planning Practices

A city is a complex system, and its planning requires comprehensive consideration of various factors. The CNN model can integrate multi-source data to provide more comprehensive support for urban planning. For example, by combining building distribution data with population migration data and environmental monitoring data,

CNN can identify correlations between environmental pollution and building layout in the city. In a study in a coastal city, the CNN model finds that air quality in some high-density residential areas is significantly worse than in other areas, mainly due to poor ventilation caused by unreasonable building layouts. Based on this finding, planners suggest adjusting building orientations and adding green belts, which improves the area's environmental quality. This case demonstrates that CNN technology, through multi-source data fusion, can provide more comprehensive and scientific support for urban planning. Furthermore, the CNN model can also incorporate social media data to analyze residents' satisfaction with the urban environment, providing a basis for public participation in urban planning. For example, in an urban renewal project, the CNN model analyzes residents' reviews of the city environment on social media, discovering that some areas have low greenery levels and poor resident satisfaction. Based on this analysis, planners propose increasing green spaces and parks, which receive high recognition from residents.

(4) CNN and the Intelligent Urban Form Generation

In actual urban planning, the application of CNN models is not limited to data analysis. It can also achieve the intelligent generation of urban forms through technologies such as the generative adversarial network (GAN). For example, by combining CNN and GAN models, researchers can generate urban design schemes that comply with specific planning principles, such as low-carbon cities and smart cities. These generated designs not only provide planners with diverse design options but can also be optimized through simulation and evaluation. Taking a smart city pilot project as an example, the research team uses the CNN-GAN model to generate multiple urban form design schemes and selects the optimal one through simulation and evaluation. This scheme outperforms others in terms of energy consumption, traffic efficiency, and environmental quality, providing valuable references for the construction of smart cities. The application of this technology not only improves the efficiency of urban planning but also offers new possibilities for innovation in urban design.

(5) Practical Applications of Research Results in Urban Planning

Based on the research results of CNN in road network planning, building distribution analysis, and multi-source data fusion, the following specific examples demonstrate how these results can be applied to actual urban planning:

1) Road Network Optimization

In the traffic planning of a medium-sized city, the research team uses the CNN model to analyze the traffic flow distribution of the city's main roads and found that some sections are severely congested during peak hours. Based on this analysis, the planners propose suggestions for adding bus-only lanes and optimizing signal timing, and use the CNN model to simulate and verify the optimization plan. The results show that the traffic efficiency of the optimized road network increases by 20%, and the traffic congestion index decreases by 15%.

2) Building Density Adjustment and Public Facility Optimization

In the planning of a new urban area, the CNN model is used to analyze the matching situation between building distribution and population density. It is found that the building density in some areas is too high, resulting in a shortage of public facilities. According to the analysis results of CNN, the planners propose suggestions to reduce the building density and add public facilities such as schools and hospitals. After implementation, the quality of life in this area has been significantly improved, and residents' satisfaction has greatly increased.

3) Environment - Friendly Urban Planning

In an eco-city planning project, the CNN model combines with environmental monitoring data to analyze the relationship between building layout and air quality. It is found that the unreasonable building layout in some areas leads to poor ventilation and poor air quality. According to the analysis results of the CNN, the planners propose suggestions to adjust the building orientation, add green belts, and set up ventilation corridors. After implementation, the air quality in this area has been significantly improved, making it a model for the construction of an environment - friendly city.

The comprehensive application of the CNN model in urban analysis provides a new perspective and methodological support for urban planning. Through in-depth analysis of road networks, building distributions, and multi-source data, CNN not only helps planners identify existing problems but also provides a scientific basis for future planning. In addition, through intelligent generation technology, CNN further expands the possibilities of urban planning and lays a solid foundation for more scientific, efficient, and sustainable urban development. With the continuous progress of DL technology, the application of CNN in urban planning will be more extensive and in-depth, injecting new vitality into urban governance and sustainable development.

5. Conclusion

This study pioneers the deep integration of deep learning techniques into the generation and design of intelligent urban morphology, effectively overcoming the limitations of traditional methods in processing complex urban spatial data. By leveraging advanced CNN architectures, the study performs detailed processing and in-depth analysis of massive and diverse urban streetscape images. In road network planning, the model accurately identifies the topological structure, hierarchical levels, and connectivity of various road types, providing a solid foundation for the construction of an efficient transportation system. In building distribution analysis, it precisely determines the functional types, spatial layout patterns, and density distribution characteristics of buildings, thereby supporting the rational utilization and development of urban land. For public facility layout, the model scientifically locates the optimal positions, sizes, and service coverage areas for various facilities, significantly enhancing the balance and accessibility of urban public services. In traffic flow analysis, it accurately predicts the dynamic variations in traffic flow at different times and across various road segments, providing critical support for traffic management strategies. Through a rigorously designed evaluation paradigm covering both day and night scenarios, the study comprehensively and objectively assesses the model's performance under varying lighting and environmental conditions. Experimental results clearly demonstrate that the model achieves industry-leading accuracy and fit across the aforementioned key tasks. In road network planning, accuracy exceeds 96%, and the fit exceeds 95%, ensuring that road planning solutions align closely with actual traffic demands. The accuracy of building distribution analysis remains above 94%, with a fit above 95%, providing precise guidance for optimizing urban building layouts. Public facility layout accuracy surpasses 96%, with a fit above 94%, ensuring efficient allocation of public resources. Traffic flow analysis accuracy exceeds 93%, with a fit over 94%, effectively assisting traffic management departments in achieving intelligent traffic control. These findings

inject new vitality into urban planning theory and practice, significantly enriching the technical methods and decision-making foundations for urban spatial analysis. They strongly promote the advancement of urban planning towards intelligence, precision, and scientific rigor. Although there is room for further improvement in the current study, it has already made a critical breakthrough in the field of intelligent urban morphology generation. This lays a solid foundation for subsequent research and holds the potential to spark profound transformations and innovative development within the urban planning field.

References

1. Lv, Z., Li, J., Dong, C., et al.: Deep Learning in the COVID-19 Epidemic: A Deep Model for Urban Traffic Revitalization Index. Data & Knowledge Engineering, 135, 101912. (2021)
2. Yan, X., Ai, T., Yang, M., et al.: A Graph Deep Learning Approach for Urban Building Grouping. Geocarto International, 37(10), 2944-2966. (2022)
3. Menon, V. G., Jacob, S., Joseph, S., et al.: An IoT-enabled Intelligent Automobile System for Smart Cities. Internet of Things, 18, 100213. (2022)
4. Saleem, M., Abbas, S., Ghazal, T. M., et al.: Smart Cities: Fusion-based Intelligent Traffic Congestion Control System for Vehicular Networks Using Machine Learning Techniques. Egyptian Informatics Journal, 23(3), 417-426. (2022)
5. Tekouabou, S. C. K.: Intelligent Management of Bike Sharing in Smart Cities Using Machine Learning and Internet of Things. Sustainable Cities and Society, 67, 102702. (2021)
6. Chen, J., Ramanathan, L., Alazab, M.: Holistic Big Data Integrated Artificial Intelligent Modeling to Improve Privacy and Security in Data Management of Smart Cities. Microprocessors and Microsystems, 81, 103722. (2021)
7. Tong, Z., Ye, F., Yan, M., et al.: A Survey on Algorithms for Intelligent Computing and Smart City Applications. Big Data Mining and Analytics, 4(3), 155-172. (2021)
8. Haseeb, K., Din, I. U., Almogren, A., et al.: Intelligent and Secure Edge-enabled Computing Model for Sustainable Cities Using Green Internet of Things. Sustainable Cities and Society, 68, 102779. (2021)
9. Herath, H., Mittal, M.: Adoption of Artificial Intelligence in Smart Cities: A Comprehensive Review. International Journal of Information Management Data Insights, 2(1), 100076. (2022)
10. Gohar, A., Nencioni, G.: The Role of 5G Technologies in a Smart City: The Case for Intelligent Transportation System. Sustainability, 13(9), 5188. (2021)
11. Fan, Z., Zhang, F., Loo, B. P. Y., et al.: Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences, 120(27), e2220417120 (2023)
12. Zhao, L., Wang, Q., Hwang, B. G.: How to promote urban intelligent transportation: a fuzzy cognitive map study. Frontiers in Neuroscience, 16, 919914 (2022)
13. Neupane, B., Horanont, T., Aryal, J.: Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. Remote Sensing, 13(4), 808 (2021)
14. Yigitcanlar, T., Mehmood, R., Corchado, J. M.: Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. Sustainability, 13(16), 8952 (2021)
15. Neffati, O. S., Sengan, S., Thangavelu, K. D., et al.: Migrating from traditional grid to smart grid in smart cities promoted in development country. Sustainable Energy Technologies and Assessments, 45, 101125 (2021)
16. Bokhari, S. A. A., Myeong, S.: Use of artificial intelligence in smart cities for smart decision-making: A social innovation perspective. Sustainability, 14(2), 620 (2022)

17. Rathee, G., Iqbal, R., Waqar, O., et al.: On the design and implementation of a blockchain enabled e-voting application within IoT-oriented smart cities. IEEE Access, 9, 34165-34176 (2021)
18. Chen, Z., Sivaparthipan, C. B., Muthu, B. A.: IoT based smart and intelligent smart city energy optimization. Sustainable Energy Technologies and Assessments, 49, 101724 (2022)
19. Ageed, Z. S., Zeebaree, S. R. M., Sadeeq, M. A. M., et al.: A state of art survey for intelligent energy monitoring systems. Asian Journal of Research in Computer Science, 8(1), 46-61 (2021)
20. Lv, Z., Chen, D., Lou, R., et al.: Intelligent edge computing based on machine learning for smart city. Future Generation Computer Systems, 115, 90-99 (2021)
21. Javed, A. R., Shahzad, F., ur Rehman, S., et al.: Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. Cities, 129, 103794 (2022)
22. Abbas, S., Khan, M. A., Athar, A., et al.: Enabling smart city with intelligent congestion control using hops with a hybrid computational approach. The Computer Journal, 65(3), 484-494 (2022)
23. Sarker, I. H., Khan, A. I., Abushark, Y. B., et al.: Internet of Things (IoT) security intelligence: a comprehensive overview, machine learning solutions and research directions. Mobile Networks and Applications, 28(1), 296-312 (2023)
24. Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., et al.: IoT for smart cities: Machine learning approaches in smart healthcare — A review. Future Internet, 13(8), 218 (2021)
25. Mabrouki, J., Azrour, M., Fattah, G., et al.: Intelligent monitoring system for biogas detection based on the Internet of Things: Mohammedia, Morocco city landfill case. Big Data Mining and Analytics, 4(1), 10-17 (2021)
26. Zekić-Sušac, M., Mitrović, S., Has, A.: Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. International Journal of Information Management, 58, 102074 (2021)
27. Al-Turjman, F., Zahmatkesh, H., Shahroze, R.: An overview of security and privacy in smart cities' IoT communications. Transactions on Emerging Telecommunications Technologies, 33(3), e3677 (2022)
28. Liu, W., Xu, Y., Fan, D., et al.: Alleviating corporate environmental pollution threats toward public health and safety: the role of smart city and artificial intelligence. Safety Science, 143, 105433 (2021)
29. Hu, Q., Zheng, Y.: Smart city initiatives: A comparative study of American and Chinese cities. Journal of Urban Affairs, 43(4), 504-525 (2021)
30. Deng, T., Zhang, K., Shen, Z. J. M.: A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. Journal of Management Science and Engineering, 6(2), 125-134 (2021)
31. Razmjoo, A., Østergaard, P. A., Denai, M., et al.: Effective policies to overcome barriers in the development of smart cities. Energy Research & Social Science, 79, 102175 (2021)
32. Lv, Z., Qiao, L., Kumar Singh, A., et al.: AI-empowered IoT security for smart cities. ACM Transactions on Internet Technology, 21(4), 1-21 (2021)
33. Mishra, S., Thakkar, H. K., Mallick, P. K., et al.: A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. Sustainable Cities and Society, 72, 103079 (2021)
34. Bhattacharya, S., Somayaji, S. R. K., Gadekallu, T. R., et al.: A review on deep learning for future smart cities. Internet Technology Letters, 5(1), e187 (2022)
35. Fan, C., Zhang, C., Yahja, A., et al.: Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. International Journal of Information Management, 56, 102049 (2021)
36. Chen, C., Jiang, J., Zhou, Y., et al.: An edge intelligence empowered flooding process prediction using Internet of Things in smart city. Journal of Parallel and Distributed Computing, 165, 66-78 (2022)

37. Guo, Z., Shen, Y., Wan, S., et al.: Hybrid intelligence-driven medical image recognition for remote patient diagnosis in Internet of Medical Things. IEEE Journal of Biomedical and Health Informatics, 26(12), 5817-5828 (2021)
38. Majeed, U., Khan, L. U., Yaqoob, I., et al.: Blockchain for IoT-based smart cities: Recent advances, requirements, and future challenges. Journal of Network and Computer Applications, 181, 103007 (2021)
39. Singh, R., Sharma, R., Akram, S. V., et al.: Highway 4.0: Digitalization of highways for vulnerable road safety development with intelligent IoT sensors and machine learning. Safety Science, 143, 105407 (2021)
40. Ahmed, I., Jeon, G., Piccialli, F.: From artificial intelligence to explainable artificial intelligence in Industry 4.0: A survey on what, how, and where. IEEE Transactions on Industrial Informatics, 18(8), 5031-5042 (2022)
41. Ramu, S. P., Boopalan, P., Pham, Q. V., et al.: Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. Sustainable Cities and Society, 79, 103663 (2022)
42. Zhang, Y., Geng, P., Sivaparthipan, C. B., et al.: Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. Sustainable Energy Technologies and Assessments, 45, 100986 (2021)
43. Arulananth, T. S., Kuppusamy, P. G., Ayyasamy, R. K., et al.: Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis. PLOS ONE, 19(4), e0300767 (2024)
44. Fan, Z., Zhang, F., Loo, B. P. Y., et al.: Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences, 120(27), e2220417120 (2023)
45. Zhang, L., Kim, C.: Chromatics in urban landscapes: Integrating interactive genetic algorithms for sustainable color design in marine cities. Applied Sciences, 13(18), 10306 (2023)

Zeke Lian was born in Bengbu, Anhui, P.R. China, in 1996. He received M.S. in Landscape Architect from Beijing Forestry University, China. Now, he works in Landscape ecology School of Ningbo City College of Vocational Technology. His research interests include Urban Planning, Generative Artificial Intelligence (GEI), Ecology, Landscape Architect. Email: lianzeke@nbcc.edu.cn

Hui Zhang was born in Huangshan, Anhui, P.R. China, in 1993. She received PhD in Economics from Pai Chai University, Korea. Now, she works in Business School of Ningbo City College of Vocational Technology. Her research interests include cloud cross-border e-commerce, e-commerce economy, e-commerce supply chain, international trade, logistics management. E-mail: zhanghui11261993@163.com

Ran Chen, a doctoral candidate at Beijing Forestry University's School of Landscape Architecture, focuses on deep learning, digital design, landscape architecture planning, and ecosystem services in his research. He has published 11 academic papers, contributed to three monographs, and holds two patents. Additionally, Chen is a partner, founder, and technical consultant at Beijing Zhijingyuntu Technology Co., Ltd., and leads the AI Application Technology Team at the School of Landscape Architecture, BJFU. He developed an automated landscape architecture design system funded by the National Natural Science Foundation of China, Tongji University, the Ministry of Education of China, and various listed companies. This project has received over 500,000 hits on public social media platforms. Chen has participated in more than 20

academic conferences internationally, gaining recognition in his field. E-mail:13924777217@163.com

Received: December 05, 2024; Accepted: March 03, 2025.

Computer Science and Information Systems 22(3):1299–1330 https://doi.org/10.2298/CSIS241205039F

Impact of Inspirational Film Appreciation Courses on College Students by Voice Interaction System and Artificial Intelligence

Shaohua Fan[1,*] and Yujing Song[2]

[1] School of Business Administration, Chongqing Technology and Business University, Chongqing, 400067, China
fanshaohua@ctbu.edu.cn
[2] Wealth Management School, Chongqing Finance and Economic College, Chongqing, 401320, China
13452366717@163.com

Abstract. To improve the mental health education level of college students and promote the adoption of artificial intelligence (AI) in college education, freshmen from a university are selected as research subjects. Two classes are chosen, with 40 students in each, and they are assigned as the experimental and control groups, respectively. The analysis of the voice interaction system of campus psychological consultation is combined with self-efficacy. The Sixteen Personality Factor Questionnaire (16PF) and the General Self-Efficacy Scale are adopted to analyze the mental health of the two groups of students at the beginning and the end of the semester. The interactive technology in AI is applied to construct a user mental model to achieve voice interaction designing. Knowledge base matching is performed on students' interactive input text, and long short-term memory (LSTM) is adopted to analyze the sentiment type of the input information and then classify it. Moreover, the answer with higher voice confidence is returned as the candidate's answer to the machine. After that, the possible text of each candidate answer is predicted and analyzed according to the interactive condition probability, and the optimal result is fed back to the student interface. Then, it points out the effect of the proposed method in analyzing the influence of inspirational film appreciation courses on the mental health level of college students. The results show a significant difference in the sensitivity factor in the experiment results at the beginning of the semester ($P<0.05$). Meanwhile, the difference in the automaticity factor is very significant ($P<0.05$). Both sensitivity and automaticity factors show different variations than expected in the experimental data at the beginning of the semester, especially the changes in automaticity factors. At the end of the semester, there are no significant differences in the sensitivity and self-discipline of the students. Moreover, it significantly influences students in the experimental group after enjoying the inspirational film, especially in terms of interpersonal communication and emotional management, positively impacting students' mental health. Meanwhile, the students in the experimental group adopt the voice interactive system for mental health consultation, and the machine can give some references to protect students' privacy to a certain extent. Therefore, when adopted to analyze the impact of inspirational film appreciation courses on the mental health of college

* Corresponding author

students, the voice interaction system for campus psychological consultation under AI combined with self-efficacy has a positive effect. It also contributes to recommending college mental health education and the adoption of AI in college education.

Keywords: automatic speech recognition, self-efficacy, inspirational film, mental health, 16PF, Artificial Intelligence, Voice Interaction System.

1. Introduction

In an individual's daily communication, voice is the most commonly used way to exchange information. With the continuous development of voice technology, artificial intelligence (AI) technology has been well-adopted in many domains. In addition, voice interaction technology has more in-depth adoptions in human learning, life, and work [1]. AI and other technologies have allowed voice interaction terminals to gradually enter people's lives, and are no longer limited to devices such as smartphones and computers. Adding voice interaction functions to robots and related equipment allows people to input information through natural language. This is processed by the system and fed back to the user to realize a natural and friendly human-computer interaction (HCI) way. It has become one of the development trends in the HCI field [2].

With the rapid development of the social economy and continuous advancement of technology, when people's lives are satisfied, they seek their emotional and spiritual needs. In other words, they begin to pay attention to their mental health. In recent years, the country has continuously increased its attention to people's mental health, which has caused a wave of psychological consultation. Many people hope to find their value through psychological consultation, thus deeply exploring personal future development and improving mental health levels [3]. As the promoters of social development, college students can't always live under the protection of their parents, nor can they act on their right away, which puts college students in a more subtle position. At this point, they need to work hard to adapt to university life and maintain close contact with the outside society to reduce loneliness [4]. When getting along with peers who are in the stage of psychological development, they may not be able to get effective support from them, and their hearts fluctuate. In addition, they need to face academic, emotional, and communication problems, and the psychological state of college students becomes very important. If college students lack correct self-awareness, as well as a certain degree of stress resistance and self-confidence, it is easy to produce lower self-efficacy. This may affect college students' living standard, psychological state, and their individual development in the future. Therefore, for students, higher self-efficacy has a positive meaning for maintaining a state of mental health [5].

In modern society with such advanced information, college students can broaden their horizons through various channels. They also have richer emotions, more open minds, and active thinking, and pay more attention to the pursuit of freedom and individuality. Then, the external environment may conflict with their psychology. After experiencing the intense study life of middle and high school, they get rid of the supervision of parents and teachers. Some keep their original passion for learning, have clear goals, and work hard to improve themselves, while others think that they can pursue what they call freedom and individuality. As a result, there is no goal for the

future life, and the subsequent study life also appears empty and meaningless. Therefore, timely adjustment of mental state is of great significance for improving the mental health of college students [6].

As an art that integrates thought and culture, film has the unique charm and appeal of information dissemination. Among the various film themes, inspirational films have a potential impact on college students' mental health. In addition, their positive energy, distinct ideas, and strong themes offer valuable insights for student education [7].

In summary, the voice interaction system is combined with college students' self-efficacy from the perspective of educational psychology. It aims to innovate physical and mental health education in the AI era and harness the influence of inspirational films on campus. Students interact with the machine to get the optimal response, and then student information is accurately grasped. This study analyzes the impact of inspirational film appreciation courses on college students' mental health, thereby enhancing their well-being and providing additional support for mental health education. Particularly, the integration of AI-driven voice interaction systems with inspirational film appreciation courses can provide college students with more personalized and privacy-focused mental health support. This method holds significant value, especially in enhancing emotional management and interpersonal communication skills. The innovative method proposed in this study not only offers a novel perspective for traditional mental health education but also opens new avenues for the application of AI technology in the educational domain. By incorporating self-efficacy and emotional regulation mechanisms, this study aims to help college students better cope with academic stress, emotional challenges, and related issues, thereby improving their mental health and overall quality of life. It is anticipated that this study can furnish valuable insights for the future innovation of mental health education models for college students, particularly in the practical integration of AI technology with psychological interventions.

2. Literature Review

In recent years, the application of AI in the field of mental health interventions has gradually become a research hotspot. AI technologies, particularly systems based on Natural Language Processing (NLP), have been widely utilized in emotion analysis, mental health monitoring, and interventions [8]. Pillai (2023) [9] proposed that emotion analysis systems based on NLP could effectively identify human emotional states and recommend corresponding mental health intervention measures based on different emotional conditions. In these studies, AI systems are employed to analyze patients' emotional states, thereby providing personalized mental health support. At the same time, the application of AI in emotional intelligence and mental health diagnostics has also advanced further. Yeke (2023) [10] explored how emotional intelligence systems, particularly those based on machine learning (ML) and deep learning (DL) techniques, could assist mental health professionals in efficiently identifying emotional fluctuations in patients and providing precise psychological interventions. With the continuous advancement of these technologies, AI has been used not only for early screening in mental health but also extensively in the fields of mental health education and

intervention. For example, AI-driven counseling chatbots have been shown to effectively assist individuals in self-regulation and emotional management.

However, despite the initial successes of AI applications in mental health interventions, most existing research focuses on emotion recognition, mental health diagnosis, and static online dialogue systems. Relatively few studies have explored how AI-driven voice interaction systems can be integrated with practical mental health intervention measures, particularly educational interventions targeting college students. Voice interaction systems, by analyzing emotional features in speech, can resonate with users in a more natural manner than text-based systems, demonstrating stronger emotional transmission capabilities [11]. Unlike text or images, voice conveys not only information but also emotions, making it a more intuitive and vivid medium in the mental health domain. Pan (2023) [12] highlighted in their research that voice interaction systems, through the analysis of speech emotions, could identify an individual's emotional state and provide more personalized psychological counseling suggestions. By incorporating emotion recognition and voice emotion analysis technologies, these systems enable mental health professionals to monitor individuals' emotional fluctuations in real time, thus optimizing intervention strategies.

For example, Caulley et al. (2023) [13] investigated AI-based voice psychological counseling systems. They found that voice interaction could enhance students' sense of engagement and strengthen emotional resonance through vocal feedback, thereby improving the effectiveness of mental health interventions. Another significant advantage of voice interaction systems is their ability to provide more privacy-conscious mental health counseling. Compared to traditional face-to-face counseling, voice interaction systems, through their anonymity and non-face-to-face communication, offer students enhanced privacy protection, which is particularly critical in mental health education [14]. Additionally, these systems can dynamically adjust intervention strategies based on students' vocal emotional feedback in real time, making personalized interventions significantly more effective in mental health education.

Inspirational films, as a powerful artistic medium with strong emotional impacts, have been increasingly applied in mental health education in recent years. Film appreciation has been shown to elicit emotional responses in viewers, thereby influencing their emotional regulation abilities and self-efficacy. Yesildag & Bostan (2023) [15] found that inspirational films could improve college students' emotional management skills and social competence, particularly by fostering a more positive mindset when facing stress and challenges. At the same time, film appreciation has been discovered to promote students' sense of social identity and collective belonging, enhancing their psychological resilience. Pan et al. (2023) [16] indicated that after participating in film appreciation courses, college students exhibited improved emotional regulation, better attitudes toward adversity, and a more positive approach to social difficulties. However, despite the demonstrated potential of inspirational films in emotional regulation, most existing research has focused on their indirect effects on emotions, with limited studies exploring the integration of film appreciation courses with voice interaction systems.

Additionally, mindfulness training has gained widespread application in mental health education for college students in recent years. By cultivating self-awareness and emotional regulation skills, it helps students better cope with stress and anxiety. While mindfulness training can alleviate emotional issues to some extent, it often requires

prolonged practice and high levels of individual commitment, which may limit its effectiveness for certain students. In contrast, group counseling promotes emotional support and social skills development through group interactions and shared experiences. However, it faces challenges in privacy protection, individualized interventions, and participation. In comparison to these traditional approaches, the innovative integration of AI-driven voice interaction systems with inspirational film appreciation courses provides a more personalized and immediate intervention. Through vocal emotion analysis, it offers tailored emotional support to each student. This approach addresses privacy concerns inherent in traditional methods and leverages AI's intelligent feedback mechanisms to dynamically adjust intervention strategies, making psychological interventions more flexible and precise. Furthermore, AI-driven voice systems enable non-face-to-face counseling, allowing students to feel more comfortable and secure when discussing sensitive topics.

Despite existing studies on the effectiveness of AI-driven voice interaction systems and inspirational film appreciation courses in mental health interventions, empirical research combining these two approaches remains scarce. Current literature lacks systematic exploration of how to integrate AI-driven voice interaction systems with inspirational film appreciation courses, particularly in the context of mental health education for college students. Most existing studies focus on single intervention methods and lack a multidimensional, cross-technology integrated framework. Regarding privacy protection, existing research has not sufficiently examined how voice interaction systems can enhance student privacy, particularly during mental health counseling. Traditional mental health interventions rely heavily on face-to-face communication, while voice interaction systems, through non-face-to-face methods, can effectively reduce students' privacy concerns. Moreover, voice interaction systems can adjust psychological intervention strategies in real-time based on students' vocal emotional feedback, significantly enhancing the effectiveness of mental health education.

This study innovatively proposes integrating AI-driven voice interaction systems with inspirational film appreciation courses to explore the impact of this comprehensive intervention framework on college students' mental health. The study addresses the gap in existing literature regarding the combination of AI-driven voice interaction systems with educational interventions. Meanwhile, it introduces quantitative tools such as self-efficacy, the Sixteen Personality Factor Questionnaire (16PF), and the General Self-Efficacy Scale (GSES) to comprehensively evaluate the framework's effectiveness. By combining voice interaction systems with film appreciation courses, the method provides more personalized mental health support for college students. Also, it enhances their emotional management and interpersonal skills, ultimately improving their overall mental health. This integrated method offers a novel solution for enhancing emotional regulation, social skills, and emotional management, contributing a new perspective to mental health interventions.

3. Method

3.1. Overall Architecture

In this study, the AI system comprises three primary components: a voice interaction system, a knowledge base matching module, and emotion analysis. These components are integrated through a modular design and work collaboratively to accomplish tasks related to mental health counseling. Specifically, the system architecture is displayed in Figure 1:

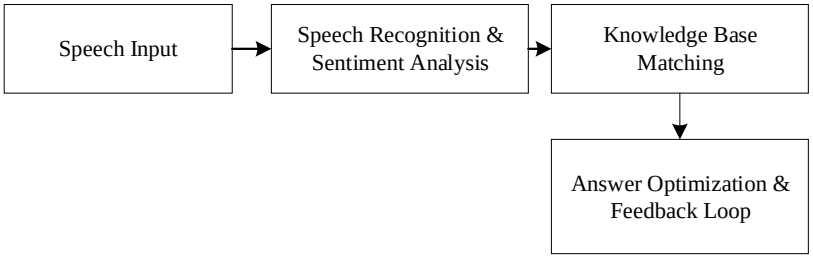


Fig. 1. Overall architecture

In Figure 1, students interact with the system through voice input. The system utilizes speech recognition technology (including preprocessing, feature extraction, and speech signal analysis) to convert the students' voice input into text. A Long Short-Term Memory (LSTM) is employed to perform emotion analysis on the text, determining the students' emotional tendencies. Upon completing this stage, the system identifies the emotional type of the input, which is subsequently used for response generation. The system uses semantic matching algorithms, such as cosine similarity, to match the student's input text with candidate answers stored in a pre-constructed knowledge base. During this stage, the system organizes and stores question-answer pairs using a pre-constructed knowledge graph structure. By querying relevant data in the corpus, it provides the most appropriate candidate answers. Based on the knowledge base matching results, the system further refines and optimizes the responses according to the student's emotional state. The emotion analysis model scores the candidate answers, prioritizing those most consistent with the student's emotional state. The three modules are interconnected via data flow and form a closed-loop feedback mechanism. During each interaction, the system not only provides immediate responses but also continuously refines its knowledge base and models through iterative feedback, enhancing the system's intelligence and adaptability.

The LSTM model employed in the system captures more complex emotional information, such as mixed emotions or subtle emotional changes that students may exhibit during mental health counseling. This innovation enables the system to deliver more precise and emotionally appropriate responses, thereby improving the quality of mental health services. Traditional systems often struggle to provide adequate responses to semantic variations or complex inquiries. However, the proposed system leverages a knowledge graph to link diverse question-answer pairs, maintaining high accuracy and flexibility when addressing open-ended and diverse queries. These technological

advancements enhance the system's intelligence and address computational challenges encountered in real-world applications, ensuring the system's efficiency and accuracy.

3.2. Design of the Speech Recognition Interactive System

The speech signal is not only a communication tool but also contains a lot of information. The process of auto-speech recognition is extracting useful information from speech to complete feature extraction. The speech recognition system greatly affects the effect of speech recognition [17].

The collection and preprocessing of speech signals constitute the first step in the system. Various preprocessing techniques are employed to ensure the quality of the speech signals and the feasibility of subsequent processing. Specifically, the system processes raw audio signals through techniques such as sampling, enhancement, detection, anti-aliasing filtering, and noise reduction. At this stage, noise reduction algorithms are applied to eliminate background noise, while anti-aliasing filters effectively remove unnecessary high-frequency components from the signal, thus improving its clarity and recognizability. Following these steps, the system accurately identifies the start and end points of the speech input, providing high-quality speech signals for feature extraction.

During the feature extraction stage, the system employs multiple methods to extract effective features from the speech signal, which directly influence the efficiency and accuracy of speech recognition. Techniques such as Linear Predictive Coding Cepstrum (LPC), Fast Fourier Transform (FFT), spectral cosine transformation, time-frequency domain analysis, and wavelet analysis are employed to extract both frequency-domain and time-domain features of speech. LPC and FFT are primarily used to extract spectral features of the signal, while time-frequency domain analysis and wavelet analysis capture variations in the speech signal across multiple scales. In addition to these frequency-domain features, other features relevant to speech recognition, such as formants, energy averages, cepstral coefficients, and zero-crossing rates, are also extracted. These features play a crucial role in the subsequent speech recognition process.

Speech recognition is essentially a pattern-matching process, wherein the system compares the input speech signal with a pre-established pattern library. For isolated words and short phrases, the Dynamic Time Warping (DTW) model is adopted, as it can handle temporal misalignment in speech signals, making it particularly suitable for shorter speech segments. For longer sentences, a combination of Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs) is employed [18]. HMMs are primarily utilized to model the temporal relationships within the speech signal, effectively handling its dynamic variations. ANNs utilize a multilayer network structure to learn the nonlinear features of speech, enhancing the system's representational capacity and recognition accuracy. By combining the strengths of HMMs and ANNs, the system remarkably improves its ability to recognize continuous speech with high accuracy.

Speech recognition technology is designed to allow the machine to receive people's voice [19]. After understanding spoken language, the system recognizes and processes the speech, converting it into machine-readable content through specific rules. The speech recognition process is presented in Figure 2.

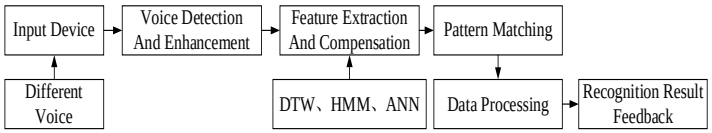


Fig. 2. Speech recognition process

In the above process, the system performs several recognition operations through several links after receiving the speech signal. The first stage is pre-processing, where the system completes tasks such as sampling, enhancement, detection, anti-aliasing filtering, and noise reduction. This enables the system to identify the start and end points of speech vocabulary. The second stage involves extracting the characteristic values of the speech signal, such as formants, energy mean, cepstral coefficients, linear prediction coefficients, zero crossings, etc, which has a greater impact on the efficiency of speech recognition. The third stage is training the signal template. This involves the digital processing of extensive voice signal templates and databases, along with the extraction of voice information, to create a speech signal instruction database for information matching. Finally, it is necessary to match the pattern, analyze and identify the distortion between the voice to be tested and the corresponding template in the pattern library through certain standards, ultimately selecting the most appropriate template for output [20].

In terms of speech confidence calculation, the system evaluates the confidence level of each candidate's response by calculating its predicted probability. To achieve this, the system employs an LSTM model for emotion analysis. LSTM, a specialized recurrent neural network, can handle long-term dependencies in sequential data and utilizes the contextual information of speech inputs to determine emotional tendencies. By analyzing the probability of each candidate's response, the system ultimately selects the response with the highest confidence level and provides it to the user. This process optimizes the selection of candidate responses and ensures that the emotional tone of the feedback aligns with the student's psychological state.

For candidate response selection and knowledge base matching, a dynamically adjustable knowledge base and response templates are designed. The system dynamically adjusts the selection of candidate responses based on the student's input and contextual information to cater to the personalized needs of different users. This innovative approach significantly enhances the accuracy of the system's feedback, making the voice interaction process more intelligent and personalized. Finally, speech synthesis is one of the critical components of the voice interaction system. To enable the system to communicate naturally with students, Text-to-Speech (TTS) technology is utilized. This technology converts the recognized textual information into natural and fluent speech output. Through pre-designed algorithms, TTS technology ensures the naturalness and coherence of the speech output during the conversion process. By leveraging this technology, students can input their queries via speech and receive natural language speech feedback from the system, thereby improving the overall interaction experience [21].

3.3. Self-efficacy

Self-efficacy refers to an individual's judgment, belief, or perception of their ability to complete a task or activity to a certain extent. It represents a person's understanding of their capabilities and their ability to manage external environments. When individuals are full of confidence in their abilities, they can effectively control the external environment and become more active in life and learning [22]. Many scholars regard this self-efficacy as a basic self-evaluation feature that influences behavior and corresponding responses. Meanwhile, it is also an expression of motivational characteristics and a positive attitude [23].

Self-efficacy is often analyzed as an individual's overall confidence in managing various environments and adapting to new situations. It is typically assessed using a general self-report measurement method. The GSES proposed by Schwarzer contains 10 items, which is a 4-point scale. This scale is simple, reliable, and has proved through extensive research and experimentation to be universally applicable across diverse countries and cultural contexts [24].

3.4. User Mental Model

In HCI, the user's mental model helps students understand the intelligent voice system, its information, content, and related skills. However, the system's database is often limited by personal experience. Therefore, the establishment and improvement of the database is crucial for improving the performance of the interactive system. When constructing a component-structured voice interaction user mental model, it must align with the design cognition, user interface, and usage habits, reflecting the user-centered design principles.

When the system receives user input, it must identify the information, understand its purpose, determine the necessary actions, and produce the corresponding output. The role of the user mental model is revealed in Figure 3.

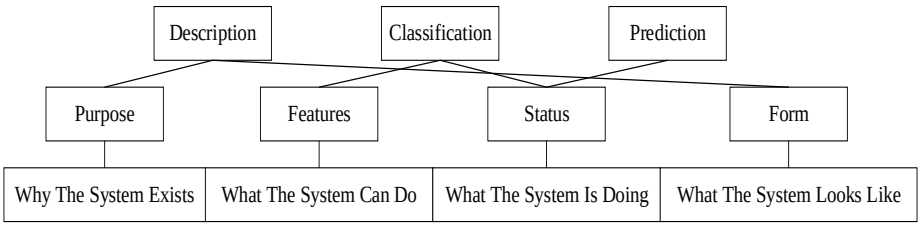


Fig. 3. The relationship diagram of the user mental model

In Figure 3, the system processes the received information to address related issues and ultimately provides the corresponding status based on prior data. During user interaction, they must anticipate the system's model and analyze its functions, interaction logic, information flow, and perception relationships. After the user completes the interactive operation, the mental model can be compared with other models to assess user experience. To further enhance this experience, the interactive model is iteratively refined in line with user cognition, improving its accuracy.

To assess the user's mental model, the first step is to collect user information, primarily through students inputting their queries into the interactive model. The system then matches this information in the background to identify the most appropriate feedback for output. Subsequently, a user mental model is constructed based on the available data, which is continuously optimized through iteration to complete the interactive process.

3.5. LSTM Network

In the field of NLP, DL techniques, particularly LSTM and Recurrent Neural Networks (RNNs), have been extensively applied to text analysis and sentiment classification tasks. LSTM, a specialized type of RNN, addresses the gradient vanishing and explosion problems encountered in traditional RNNs when processing long sequences by introducing memory cells. Compared to traditional RNNs, LSTM can retain and update states over extended time sequences, making it suitable for tasks involving events with long intervals and delays. Consequently, LSTM performs exceptionally well in emotion analysis, speech recognition, and other tasks requiring the capture of long-term dependencies.

The key feature of LSTM is its three gating mechanisms—forget gate, input gate, and update gate. These mechanisms control the flow of information, ensuring that the network effectively retains critical past information while discarding unnecessary details. This design allows LSTM to maintain meaningful context over long sequences.

In this study, an LSTM-based text emotion analysis model is constructed to analyze the textual data input by students through an AI-driven voice interaction system. Specifically, the input text data undergoes tokenization before being processed by the LSTM for lexical classification and semantic matching. The unique structure of LSTM enables it to model sentiment at each time step and ultimately output the overall sentiment category of the text (e.g., positive, negative, or neutral). This model effectively captures students' emotional states, providing precise support for subsequent mental health interventions and feedback. All text data undergoes preprocessing, including lowercasing, punctuation removal, and tokenization. Before being input into the LSTM model, the text is transformed into vectors using Word2Vec word embedding technology, enabling each word to be represented in a high-dimensional space.

The Dropout technique is introduced to prevent overfitting in the LSTM. Dropout randomly discards a portion of neuron connections during training, forcing the network to learn more robust features and improving its generalization ability. This method is particularly effective when the dataset is small or imbalanced. The dropout rate is set to 0.5, meaning 50% of the neuron connections are randomly dropped during each training iteration to mitigate overfitting. During the training process, the cross-entropy loss function is employed, which is widely used in classification tasks to measure the disparity between the model's predicted outputs and the true labels. For optimization, the Adam optimizer is applied, which updates network weights using both gradient momentum and second-order moments, facilitating faster convergence and avoiding local minima.

The training process is as follows. First, students' speech data and mental health assessment data (e.g., self-efficacy, 16PF, and GSES) were used as the training set. The data undergoes preprocessing, where textual data is tokenized and converted into word

vector representations. The annotated sentiment labels (positive, negative, neutral) serve as the target outputs. The dataset is split into training, validation, and test sets in an 80%, 10%, and 10% ratio, respectively.

The model training involves the following parameters. A two-layer LSTM structure is used, with each layer containing 128 hidden units to effectively capture the temporal dependencies in the speech text. The batch size is 32, the number of training epochs is 20, and the initial learning rate is 0.001. An early stopping mechanism is employed to improve model stability and further prevent overfitting, halting training when performance on the validation set ceases to improve. Batch normalization is also utilized during training to reduce internal covariate shift, and Dropout is applied to enhance the model's robustness. With these optimizations and techniques, the LSTM model can effectively classify the sentiment of students' textual input, enabling the voice interaction system to provide accurate feedback on students' emotional states. This supports personalized interventions in mental health education. Evaluation metrics for the model include accuracy, precision, recall, and F1 score.

4. Experiment

4.1. Research Subjects

Two freshman classes from a college are selected randomly as the research subjects, with 40 students in each class. These students are 18-19 years old and are divided into experimental and control groups. The class hours of mental health education courses of the experimental group are increased, that is, the inspirational film appreciation courses and psychological consultation interactive courses are added. The students in the control group receive traditional mental health courses. Evaluations are conducted at the beginning and end of the semester using the Cattell 16PF test to analyze the impact of inspirational film appreciation courses on college students' mental health. The analysis incorporates the voice interaction system combined with self-efficacy. At the same time, students in the experimental group complete questionnaires and self-efficacy scales before courses to evaluate the changes in their self-efficacy after appreciating inspirational films. Then, the uncertainty and individual differences of students after appreciating inspirational films are assessed. Moreover, after class, the voice interaction system of campus psychological consultation is adopted for psychological analysis and evaluation, providing students with comprehensive insights and support.

4.2. Mental Health Level Experiment of College Students

16PF and GSES measure college students' mental health levels, self-efficacy, and psychological changes. The selected scales are of high reliability and validity, which have been confirmed in many studies [25].

The final version of the 16PF contains 187 questions, covering 16 primary personality traits, including warmth, reasoning ability, emotional stability, dominance, liveliness, rule-consciousness, social boldness, sensitivity, vigilance, abstractedness,

privateness, apprehension, openness to change, self-reliance, perfectionism, and tension. These dimensions comprehensively reflect the personality characteristics of the respondents, facilitating the assessment of their mental health status. This study utilizes the 16PF to evaluate respondents' mental health status through self-reporting. To ensure the validity and reliability of the scale in the current study, a reliability analysis is conducted, showing that the internal consistency coefficient (Cronbach's α) of the 16PF exceeds 0.85, verifying its reliability.

Regarding the self-efficacy scale, the GSES proposed by Schwarzer and Jerusalem (1995) was utilized. This scale is a widely used tool for assessing an individual's confidence in their ability to complete tasks across various situations. Here, the self-efficacy scale undergoes a reliability analysis, yielding a Cronbach's α coefficient of 0.88, indicating strong internal consistency. To ensure its applicability to the current research population, prior literature is referenced, and the scale is adjusted during a pilot study to accurately measure college students' mental health and self-efficacy states [26-28]. The GSES is adopted and the scoring adopts the four scores method, which has a relatively high validation effect in many studies.

4.3. Analysis of Voice Interactive System

When students use the interactive system for content consultation, the system needs to identify the content and problems described by the students. Then, the system needs to formalize the operation of the problem, that is, according to the interactive input text of the kth session, the K+1th system pair learns very appropriate and reasonable feedback information in terms of speech and emotion.

In this study, the construction and maintenance of the corpus are central to ensuring the voice interaction system can provide accurate responses. The corpus includes various common questions and answers related to students' mental health counseling and is continuously optimized through manual annotation and semantic matching. The corpus construction begins with data collection from multiple sources, including psychological research literature, online mental health counseling platforms, and expert feedback. All data undergo cleaning and standardization processes to ensure quality and consistency. Furthermore, the corpus is regularly updated to cover the latest mental health-related topics and reflect questions frequently asked by students during actual counseling sessions.

To ensure efficient matching and maintenance of the corpus, a knowledge graph is employed to structurally organize and manage the data. The knowledge graph categorizes and connects the question-answer pairs in the corpus based on themes and semantic relationships, improving the system's response efficiency and accuracy during interactions. Each time new question-answer data is added, a graph-based knowledge-matching algorithm is applied to ensure consistency and coherence of the data.

In the answer prediction process, the proposed system continuously optimizes the recommendation results through an iterative approach. After each student query, the system first converts the question into a semantic vector and performs similarity matching with candidate answers from the corpus. Cosine similarity measures the semantic similarity between the input question and the candidate answers, selecting the most relevant candidates. Subsequently, emotion analysis is conducted using LSTM to filter the candidate answers, prioritizing those that match the student's emotional state.

To maximize semantic confidence, an iterative optimization process is designed. First, candidate answers are scored based on semantic conditional probabilities, incorporating information such as the student's emotional state and voice intonation to calculate the confidence of each answer. In each iteration, the system fine-tunes the semantic matching model based on the previous output and feedback. Meanwhile, it utilizes a gradient descent algorithm to optimize model parameters and ensure that each iteration improves the accuracy and confidence of the recommendation results. This process dynamically adjusts the matching weights based on the individual student's input, providing more personalized suggestions.

After each iteration, the system records user feedback and uses this data to optimize the corpus and model. This closed-loop mechanism ensures that the system can continuously learn and improve, progressively enhancing the intelligence and accuracy of the voice interaction. In conclusion, by integrating knowledge graphs, LSTM emotion analysis, cosine similarity matching, and iterative optimization, the proposed answer prediction and corpus matching process enable efficient handling of students' mental health counseling queries while maximizing the system's semantic confidence. In the interaction process, it is necessary to analyze the semantic similarity, that is, to consider the semantic information of words and the interrelationship between words. The similarity function is shown in Equation (1).

$$S(C_1, C_2) = cos(\theta) = \frac{\sum_{i=1}^{n}(x_i * y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2 * \sum_{i=1}^{n}(y_i)^2}} \quad (1)$$

C_1 and C_2 are the semantic vectors of the interactive text, and the cosine similarity is adopted to calculate the semantic similarity between the interactive texts. Cosine similarity is used to calculate the semantic similarity between the semantic vectors C_1 and C_2, which is of significant importance for analyzing the sentiment of students' speech input. The input text is preprocessed to ensure that the model can accurately capture the semantic information of the text. Specifically, the preprocessing steps include tokenization and the removal of stopwords, effectively reducing interference from irrelevant information. For word embedding, the Word2Vec method converts the text into vector representations, generating high-dimensional vectors for each word. The advantage of cosine similarity lies in its simplicity and intuitiveness, as it effectively measures the angle between two vectors, thus assessing their similarity. In emotion analysis tasks, cosine similarity is employed to compute the semantic similarity between the input text and the candidate answers in the knowledge base. This can help the system filter out the most relevant responses to the user's input. Compared to other similarity measures, such as Euclidean distance or Manhattan distance, cosine similarity has the distinct advantage of being insensitive to the size of the vectors, focusing only on the direction of the vectors. This makes cosine similarity particularly suitable for text data processing, where text lengths may vary, but semantic similarity typically manifests in the distribution of words rather than the length of the text.

In comparison with other methods, Euclidean distance can reflect the overall difference between vectors. However, it involves more complex calculations and is significantly influenced by the size of the vectors, which may introduce biases when dealing with texts of varying lengths. In contrast, cosine similarity normalizes the vectors, ensuring that the calculation is not affected by the length of the text, thereby providing a more accurate reflection of the similarity between the two texts. To sum up, the application of cosine similarity in this study, combining its simplicity, efficiency,

and insensitivity to size, is highly suitable for calculating semantic similarity in emotion analysis. Based on this, knowledge base matching is performed on the kth input information. Through continuous iteration, the answers with the highest semantic confidence for the k+2th time are found to form an answer set. The subsequent interactive text is predicted based on the interaction probability, and the interaction condition probability is as follows.

$$P(T_{HR}^{k+2}/T_{RH}^{k+1}) = \frac{P(T_{HR}^{k+2}/T_{RH}^{*k+1})}{P(T_{RH}^{*k+1})}, T_{RH}^{*k+1} \in L_{RH}^{k+1} \tag{2}$$

L_{RH}^{k+1} represents a set of texts whose semantic similarity between the information in the system corpus and the input text information is greater than the confidence level. The classifier in the interactive system needs LSTM to extract the relevant features of the input sentence information. Then, it performs classification matching through hidden nodes, and further output the classification matching results.

4.4. The Experiments

Before the experiment begins, the research variables and questionnaires need to be determined to evaluate the student's situation.

Experimental group: when designing the course content of the experimental group, the design idea of inspirational film appreciation + collective activities + psychological consultation interaction is adopted. The bi-weekly courses are set, 4 hours each time. Students enjoy inspirational films in the first 2 hours, and group activities are performed in the 3d hour, namely 2-1-1, and course practice is carried out in this mode [29-32]. HCI consultation is conducted in the last hour. In the first and last classes of this course, basic theoretical explanations and comprehensive discussion classes are given. The experimental group has 9 courses, except for the first and last courses, the other courses adopt the 2-1-1 mode. Different themes are designed in each course. The inspirational films arranged for each course are "*The Shawshank Redemption*", "*The Pursuit of Happiness*", "*Examination 1977*", "*Homeless to Harvard*", "*The Liz Murray Story*", "*Forrest Gump*", "*Together with You*", and "*Les Choristes*" [33].

When choosing these films, the main consideration is that the films need to meet the characteristics of college students' viewing. The plot is vivid and rich, the story is more tortuous, and the theme is obvious and bright. At the same time, these films are all inspirational films recommended in official publications and are the comprehensive results obtained after the conversations of many psychology teachers and film appreciation teachers [34-36].

In the experiment, relevant work is conducted from three stages before class, during class, and after class. Before class, teachers give out questionnaires, choose films, and introduce the plot outline to students. Moreover, they observe students' emotional changes and reactions in class, guide students to pay attention to the emotional changes of characters in the play, and solve some emergencies. After class, teachers guide students to fill out questionnaires and organize students to discuss the details of the film, as well as their understanding and psychological reactions. Students communicate with the voice interaction system in the last class, they can conduct voice interactions according to the relevant situations in the film and combine them with their usual

situations to realize the consultation process. After that, the content of the consultation is analyzed by the voice interaction system, and the corresponding solutions are fed back to the interactive interface of the machine to provide students with certain solutions [37-39].

The themes of these films "*The Shawshank Redemption*", "*The Pursuit of Happiness*", "*Examination 1977*", "*Homeless to Harvard*", "*The Liz Murray Story*", "*Forrest Gump*", "*Together with You*", and "*Les Choristes*" can be expressed as follows, hope can be used for spiritual redemption, to pursue dreams and take responsibility, to seize opportunities to change fate, to have firm faith to change a life, to persist in optimistic self-treatment, to rebuild spiritual and humanistic care, and to self-actualize human dignity [40, 41].

The relevant data of the questionnaire and scale are processed by Spss21.0, and then the data are further analyzed by *T-test*.

5. Result and Discussion

5.1. Interactive System Consultation and Comparison of Self-Efficacy in the Experimental Group

After students watch each inspirational film, the GSES is adopted to test them, and the students' self-efficacy enhanced by "Shawshank Redemption" and "Les Choristes" are compared and analyzed, as exhibited in Table 1.

Table 1. Comparison and analysis of general self-efficacy before and after watching inspirational films

	Before watching film	After watching film	T
The Shawshank Redemption	3.32±0.45	2.69±0.45	-6.11
Les Choristes	3.11±0.44	2.71±0.39	-6.08

In Table 1, students exhibit a significant improvement in self-efficacy after watching *The Shawshank Redemption* and *Les Choristes*. First, *The Shawshank Redemption* has a plot with a strong emotional impact and a message of positive energy. The protagonist's resilience and confidence in the face of various challenges resonate with the students. In the film, the main character overcomes seemingly insurmountable obstacles through persistent effort and unwavering belief, prompting students to reassess their self-efficacy and boost their confidence when facing difficulties and challenges. The experimental results indicate a significant increase in students' self-efficacy after watching *The Shawshank Redemption* ($P<0.001$), closely related to the film's emphasis on self-transcendence and the awakening of personal potential.

Moreover, *Les Choristes*, a heartwarming and hopeful film, primarily showcases the kindness and resilience in human nature through the interactions and transformations between the teacher and students. The protagonist in the film changes the fate of a group of troubled students through music and education, allowing them to see more possibilities within themselves and regain confidence. After watching the film, students generally display a calm and positive emotional experience, which enhances their self-efficacy ($P<0.005$). Specifically, many students report that after watching *Les choristes*, they not only let go of past negative emotions but also became more interested in campus life and their future, further improving their mental health.

However, it is noteworthy that although both films have a positive effect on students' self-efficacy, their emotional effects and mechanisms of influence differ. *The Shawshank Redemption* stimulates students' awareness of challenges and their confidence in problem-solving through intense plotlines and the protagonist's unwavering determination. *Les Choristes* evokes inner peace and a positive outlook on life through a warm, caring atmosphere and changes in interpersonal relationships. These two different emotional experiences lead to distinct emotional responses and psychological feedback, resulting in various impacts on students' self-efficacy.

According to the experimental results, the reason for this difference may be directly related to the emotional atmosphere and the intensity of the plot in the two films. *The Shawshank Redemption* is more suitable for students who need strong motivation and belief support when facing personal challenges; *Les Choristes* may be more suitable for students in need of emotional comfort and hope for life. This also suggests that diverse types of inspirational films can evoke different psychological responses in students based on their varying psychological needs, further enhancing their self-efficacy. In conclusion, although both films have a significant positive impact on students' self-efficacy, they achieve this effect through different emotional inspirational mechanisms.

When students use the voice interaction system for psychological consultation in the last class, they obtain feedback results after machine analysis by describing their past or current doubts about certain things. This can provide them with some suggestions for reference. When the students' feelings after consulting are analyzed with the voice interaction system, the feedback from the students can be obtained in Figure 4.

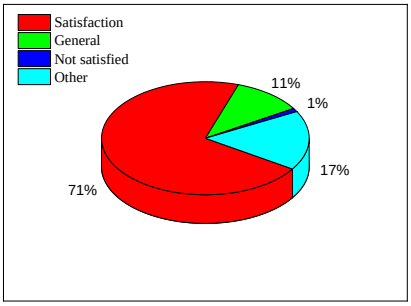


Fig. 4. Satisfaction on voice interaction system consultation

According to the above results, nearly 71% of students believe that using the voice interaction system to consult psychological problems can effectively resolve their

doubts and provide them with corresponding solutions. Moreover, some students argue that leveraging this system to consult with psychological problems can avoid the embarrassment of face-to-face with teachers and the cranky thinking when reading and consulting materials by themselves. This is an effective way to solve related problems and protect students' privacy.

5.2. Semantic Similarity Analysis

8 groups of dialogues are randomly selected from the text database of the system background for system training and testing. There are 6 and 2 sets of dialogues for training and testing, respectively. The theme of each group of dialogues is different, but the content of the dialogues has a strong relevance. Equation (1) is employed to calculate the similarity of the text and get the results in Table 2.

Table 2. Results of semantic similarity of training group text

	X_{1-1}	X_{2-1}	X_{3-1}	X_{4-1}	X_{5-1}	X_{6-1}
X_{1-1}	0.75	0.31	0.28	0.38	0.56	0.61
X_{2-1}	0.69	0.77	0.26	0.33	0.25	0.18
X_{3-1}	0.19	0.21	0.81	0.32	0.21	0.19
X_{4-1}	0.44	0.35	0.11	0.86	0.27	0.22
X_{5-1}	0.42	0.27	0.17	0.26	0.78	0.36
X_{6-1}	0.39	0.21	0.34	0.31	0.33	0.83

It can be concluded that the similarity of each group of dialogues and similar sentences in the text database is higher, all above 0.75, and the similarity of irrelevant sentences is lower than 0.7. Thus, 0.75 is taken as the similarity threshold.

After the semantic text of the two test groups is analyzed, it is found that the similarity of the test group is slightly higher than that of the training group, and the similarity can reach 0.79. Hence, it can be considered that the user mental model interaction system has a good semantic matching effect, and the textual confidence of the constructed semantic set is high, as is the corresponding similarity.

In addition, from the perspective of matching accuracy, the system can provide a highly accurate voice interaction system. Concurrently, it can effectively identify the text information input by students, solve students' doubts, and promote the process of solving mental health problems.

5.3. Comparative Analysis Before and After the Course in the Experimental Group

The students in the experimental and control groups are tested before the beginning of the course. The 16PF questionnaire is used for measurement and evaluation. The sample data are analyzed by T-test, and the results is outlined in Table 3.

Table 3. Comparison of 16PF questionnaire experiment results before the courses

Item	Experimental group	Control group	*T-test*
Warmth X_1	6.17±1.76	6.72 ±1.65	0.91
Reasoning X_2	6.33±1.85	6.56±2.23	0.21
Emotional stability X_3	5.09±1.18	6.44+2.01	0.94
Dominance X_4	7.03±1.34	5.61±1.55	0.53
Liveliness X_5	8.02±1.48	7.33±1.55	0.53
Rule-consciousness X_6	4.59±1.58	8.22±1.66	0.33
Social boldness X_7	7.14±1.57	4.71±1.44	0.45
Sensitivity X_8	5.66±1.86	7.31±1.79	2.41
Vigilance X_9	5.29±1.59	6.68±1.62	-0.33
Abstractedness X_{10}	5.33±1.69	5.22±2.01	-0.47
Privateness X_{11}	6.79±1.78	5.09±1.23	-1.89
Apprehension X_{12}	6.04±1.67	5.96±1.69	-0.77
Openness to change X_{13}	5.41±1.77	5.77±1.92	0.34
Self-reliance X_{14}	3.96±1.25	4.45±1.87	1.55
Perfectionism X_{15}	4.07±1.12	5.11±1.44	3.01
Tension X_{16}	6.72±1.32	6.68±1.47	0.21

Table 3 compares the 16PF questionnaire results of the experimental and control groups. The *T-test* results show that the differences between the two groups are not significant across multiple measurement dimensions. However, certain dimensions, such as Sensitivity (X_8) and Perfectionism (X_{15}), reveal significant differences, which are crucial for understanding the psychological changes of the students in the two groups. Firstly, the significant difference in the Sensitivity (X_8) dimension ($P<0.05$) suggests that, during the experiment, students in the experimental group are more sensitive to emotional stimuli and feedback from others than those in the control group. This result may be attributed to the enhanced emotional and perceptual abilities of the experimental group after undergoing certain interventions (e.g., film viewing, psychological regulation). Since sensitivity involves an individual's response to changes in external situations, this increased responsiveness may stem from the improved emotional regulation and adaptability to external environments of the students in the experimental group following the intervention. Through emotional stimulation from film materials, students in the experimental group experienced strong emotional reactions during the film, which may have contributed to an increased sensitivity to situational and interpersonal responses.

Secondly, the significant difference in Perfectionism (X_{15}) ($P<0.001$) warrants further analysis. Perfectionism reflects an individual's excessively high expectations for themselves and others, often leading to apprehension and stress. In this experiment, the level of perfectionism in the experimental group is significantly higher than that in the

control group. This is possibly due to the emotional resonance triggered by the film content, especially films like *The Shawshank Redemption*, which has a strong emotional impact. The film indirectly heightens the students' focus on self-expectations and societal standards by depicting the protagonist's persistence and effort in the face of adversity. Thus, these can prompt them to have higher expectations for their performance after the experiment. However, this increased tendency towards perfectionism may also lead some students to experience excessive apprehension about the high standards they set for themselves, especially when their real-life performance does not align with these standards.

Moreover, for other dimensions in the 16PF questionnaire, such as Warmth (X_1), Emotional Stability (X_3), Dominance (X_4), and Rule-Consciousness (X_6), the *T-test* results show no significant differences ($P>0.05$). This suggests that, although there are differences in the scores between the experimental and control groups on these dimensions, these differences do not reach statistical significance. A possible explanation is that these dimensions measure relatively stable personality traits, which are less susceptible to short-term psychological interventions. For instance, characteristics such as Warmth and Emotional Stability may require longer-term psychological interventions and self-adjustment to exhibit significant changes. Therefore, no significant differences are observed in this experiment.

In short, no significant differences were found between the experimental and control groups on many psychological traits. However, the significant differences in Sensitivity and Perfectionism indicate that the intervention (such as film viewing) does indeed have an impact on certain aspects of students' mental health. These findings provide directions for future research, specifically on how different psychological interventions can help students better regulate their emotions and mental states, enhance self-efficacy, and avoid excessive tendencies toward perfectionism. After comparing the project factors, the results in Figure 5 are obtained. The secondary factors in Figure 5 encompass individual psychological traits, personality tendencies, and various aspects of interaction with the environment. These factors are derived through data mining and statistical methods based on multidimensional analysis of experimental data, aiming to reveal more detailed psychological characteristics. Among them, Adaptation and Anxiety examines an individual's ability to adapt to changes or new environments, as well as the degree of anxiety experienced. Cowardice and Decisiveness reflect an individual's attitude in the decision-making process. Creative Ability refers to the capacity to generate novel and valuable ideas and solutions, which is crucial for problem-solving and innovation. Environmental Adaptation describes an individual's ability to adjust to different living environments, including both physical and sociocultural environments. High environmental adaptability indicates a better ability to cope with various challenges and changes in life.

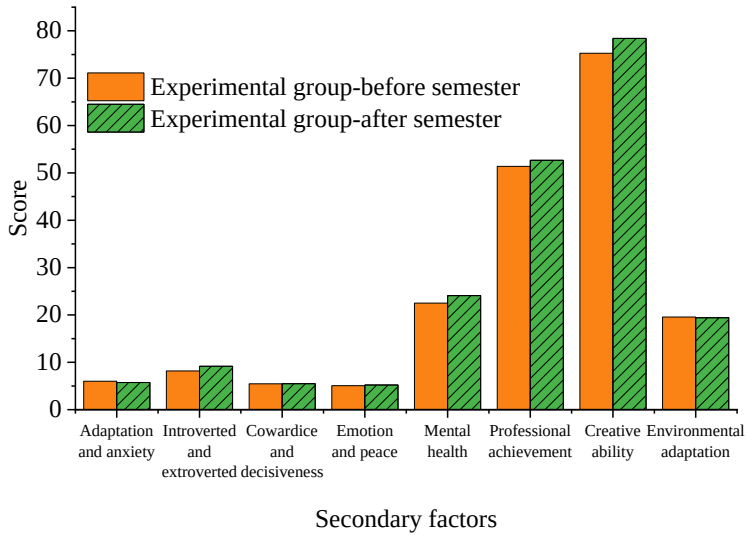


Fig. 5. Comparison of secondary factors of the experimental group at the beginning and end of the semester

The factors of Emotion and Peace and Professional Achievement have a significant difference, $P<0.05$. Apart from that, there is no significant difference in the scores of other secondary factors. To analyze the reasons for such phenomena, first, the significant increase in the scores for Emotion and Peace suggests that the experimental group students experienced noticeable improvements in emotional regulation and psychological balance. Emotion and Peace reflect an individual's emotional fluctuations and psychological state, with lower scores typically indicating emotional instability and greater worry, while higher scores suggest quick action with fewer considerations of consequences. In this experiment, the improvement in Emotion and Peace scores may be closely related to the emotional regulation and psychological stability strategies incorporated during the intervention process. Through activities such as watching films, the experimental group students may have learned more effective methods for regulating their emotions when faced with emotional fluctuations, thereby reducing psychological distress and anxiety. This may also indicate that they can handle academic pressures and personal challenges with greater composure and balance.

Second, the significant difference in Professional Achievement suggests that the experimental group students may have demonstrated a stronger sense of career goal awareness and development potential. Professional Achievement evaluates the impact of personality factors on future career development. Higher scores generally indicate a strong career drive and a clear sense of purpose. In this experiment, the experimental group students' sense of professional achievement significantly increases, possibly due to overcoming psychological and emotional challenges during the intervention, as well as improvements in self-efficacy. Specifically, by watching inspirational films, the students may have been inspired, enhancing their belief in career planning and self-actualization, thus fueling their positive outlook on future career accomplishments.

However, aside from Emotion and Peace and Professional Achievement, other secondary factors do not show significant differences. This illustrates that the experimental group students demonstrate improvements in certain psychological traits. However, the impact of the intervention on other personality factors (such as Sensitivity and Perfectionism) is minimal and does not manifest significant changes in the short term. For example, Sensitivity reflects the degree to which an individual expresses emotions, while Perfectionism relates to the strictness of personal standards and expectations. In this experiment, the scores for these factors do not undergo significant changes. This is possible because these personality traits typically require long-term self-regulation and deeper emotional processing to manifest and may be influenced by more complex external environments and personal experiences. In summary, the significant changes in Emotion and Peace, as well as Professional Achievement, suggest that the experimental group students are positively impacted in terms of emotional regulation and career development awareness. However, the similarities in other personality traits indicate that, while short-term interventions can have positive effects on certain psychological and emotional dimensions, they have a smaller impact on more stable or deeper personality factors.

After the experiment before class, it is also necessary to test the students after all 9 classes, to compare the results of the experimental group before and after the course, and the results are detailed in Table 4.

The results in Table 4 show that students change multiple personality dimensions after the experiment, with significant statistical differences observed in some of these dimensions. Warmth: After the experiment, the warmth score significantly increases ($P<0.05$), rising from 6.72 to 7.11, with a change of -2.11. Individuals with high warmth typically exhibit stronger prosocial behaviors and better communication skills. This change may be related to the enhanced emotional experience students have after watching the inspirational film. The interactions and emotional expressions of the characters in the film might have sparked students' attention and empathy towards others' emotions, encouraging them to be more friendly and outgoing in their social interactions. Apprehension and Perfectionism: These two dimensions also show significant differences ($P<0.005$). Perfectionism increases from 5.11 to 5.21. Although the change is modest, it may reflect an increase in students' self-expectations. Watching the inspirational film may have heightened their pursuit of high standards and encouraged them to focus more on details and outcomes when facing challenges. Apprehension decreases from 5.96 to 5.94, indicating a reduction in anxiety and worry experienced by students after the course. The film's plot may have helped students release negative emotions, reducing their concerns about uncertainty and enhancing their psychological resilience in the face of challenges.

Emotional stability: The score for this dimension slightly decreases (from 6.44 to 5.79), suggesting a decline in emotional stability within the experimental group. While this change is not statistically significant, it may reflect an increased need for emotional regulation during self-reflection. The inspirational film might have prompted students to experience more complex emotional fluctuations. Vigilance: It decreases from 6.68 to 5.21, showing a substantial change, although this difference is not statistically significant. This might indicate that students become less vigilant and sensitive after watching the film, gradually becoming more confident and relaxed. No significant differences are observed in some areas, such as Reasoning, Liveliness, Dominance, and Self-reliance. Despite some fluctuations in the data, no significant changes are observed

in these traits. This suggests that the intervention has a limited impact on these personality dimensions. Meanwhile, additional time or different forms of intervention may be required to induce more profound changes in students' cognitive styles and behavioral tendencies. Rule-consciousness and Sensitivity: The scores decrease from 8.22 and 7.31 to 4.89 and 5.68, respectively, indicating a decline in students' rule-consciousness and sensitivity. This change may reflect a shift toward a more flexible understanding of rules during the experiment, while the reduction in sensitivity could be related to the stabilization of their emotions.

Table 4. Comparison of experiment results of the 16PF questionnaire before and after the courses in the experimental group

Item	After the experiment (experimental group)	Before experiment (experimental group)	*T-test*
Warmth X_1	7.11±1.66	6.72 ±1.65	-2.11
Reasoning X_2	6.74±1.75	6.56±2.23	-0.98
Emotional stability X_3	5.79±1.38	6.44+2.01	-1.44
Dominance X_4	7.41±1.36	5.61±1.55	-0.22
Liveliness X_5	8.12±1.58	7.33±1.55	-0.35
Rule-consciousness X_6	4.89±1.61	8.22±1.66	-0.92
Social boldness X_7	7.22±1.67	4.71±1.44	-0.14
Sensitivity X_8	5.68±1.33	7.31±1.79	-0.27
Vigilance X_9	5.21±1.69	6.68±1.62	0.49
Abstractedness X_{10}	5.55±1.62	5.22±2.01	-0.46
Privateness X_{11}	5.01±1.72	5.09±1.23	1.93
Apprehension X_{12}	5.94±1.51	5.96±1.69	3.33
Openness to change X_{13}	5.88±1.57	5.77±1.92	-1.17
Self-reliance X_{14}	4.96±2.25	4.45±1.87	-1.36
Perfectionism X_{15}	5.21±1.32	5.11±1.44	-3.52
Tension X_{16}	6.52±1.67	6.68±1.47	0.79

Overall, the inspirational film has a positive effect on the personality development of students in the experimental group, particularly in terms of prosocial behavior (Warmth), self-expectations (Perfectionism), and emotional regulation (Apprehension). These changes suggest that watching the film can inspire students to have higher expectations for their future lives and learning while enhancing their social skills. However, not all personality dimensions show remarkable changes, which may be related to factors such as the content and duration of the intervention. Future research could further explore how different types of film content, intervention duration, and

other personality factors interact to improve students' psychological resilience and personality development.

The comparison and analysis of the changes in each secondary factor's score of 16PF in the experimental group are suggested in Figure 6.

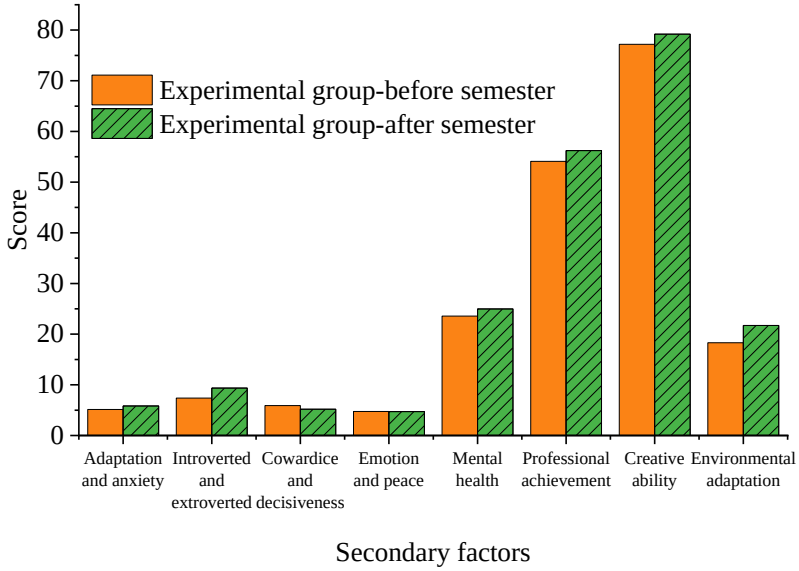


Fig. 6. Comparison of changes in secondary factors of 16PF before and after the experiment

Figure 6 shows that, after the experiment, students' mental health scores significantly improve ($P<0.05$). This change indicates that inspirational films can effectively promote students' emotional health, help alleviate stress, and enhance self-confidence. The positive themes in the films may inspire students' sense of self-worth and hope for the future, thereby improving their overall psychological state and their ability to cope with life's challenges. The Professional Achievement scores also demonstrate a significant increase ($P<0.05$). This improvement may be related to the enhanced career motivation students experienced after watching the inspirational films. The successful cases and themes of perseverance in the films may encourage students to focus more on personal growth and the planning of their future career development.

The increase in scores for Environmental Adaptation also exhibits significant differences ($P<0.05$). This may illustrate that, through inspirational films, students' ability to adapt to their environment has improved. Especially, when facing challenges and changes, they can maintain a positive mindset and adjust their behavior to better cope with new environments. The decrease in Adaptability scores, with significant differences ($P<0.05$). This suggests that, despite the improvement in environmental adaptation, students' reactions to change may have become more cautious or introverted in certain situations. This could be because the challenges and emotional fluctuations in the film's plot prompted students to reflect more on the unknown or change, leading to increased anxiety and discomfort. Anxiety scores distinctly decrease ($P<0.05$), and this

change shows significant statistical significance. This indicates that inspirational films help students alleviate anxiety, enhancing their confidence in facing future uncertainties. The inspirational stories in the films may have provided emotional support, encouraging students to respond more positively to the pressures and challenges of life.

Regarding Action Ability, the experiment does not show substantial changes, indicating that while watching the film has a positive psychological and emotional impact, its effect on actual action ability is relatively limited. This may be because the emotional inspiration and inspirational factors in the films do not effectively translate into specific actions or behavioral changes. Students may require more practical training or activities to convert these positive psychological changes into real actions. Overall, inspirational films positively impact students' mental health, professional achievement, and environmental adaptation, particularly in boosting self-confidence and reducing anxiety. However, the effect on action ability is minimal, suggesting that the psychological motivation provided by the films primarily stays at the emotional and cognitive levels. Furthermore, further practical activities or interventions are needed to help students translate these emotional changes into concrete action.

5.4. Comparative Analysis Before and After the Course in the Control Group

The students in the control group take the 16PF test at the beginning and end of the semester, and the results are listed in Table 5.

The test scores of the control group at the beginning and end of the semester change to some extent, but there is no significant difference. It reveals that the adoption of traditional mental health courses only has a positive impact on students' mental health. However, with the increase in age and grade, students are exposed to more and more external things. The external environment that students face is becoming increasingly complex, and the pressure they need to bear is also increasing. Then, if students are not in a good mental state, they are prone to negative psychological changes such as lack of information, constant worry, and blind conformity. Therefore, if it cannot set up a mental health course suitable for college students every semester to motivate them at any time, it is necessary to consider reforms in teaching content and models. Then, it can provide college students with better psychological consultation, enabling them to deal with some problems calmly and better adapt to social development.

The comparison and analysis of the changes in each secondary factor's score of 16PF in the control group are illustrated in Figure 7.

Figure 7 presents that, in the control group, scores for mental health, professional achievement, and environmental adaptation significantly increased by the end of the semester, with statistically significant differences ($P<0.05$). These changes may be closely related to the students' aging process and the accumulation of life experiences. Throughout the semester, students gradually become better at coping with academic pressure, social challenges, and other life changes. The improvement in mental health scores reflects their maturity in emotional regulation and stress management, enabling them to better handle emotional distress and academic pressure. The increase in professional achievement indicates that students have made more progress in academic accomplishments and social practice. As their plans and expectations for future career development gradually become clearer, they place more emphasis on enhancing their abilities and preparing for their careers. The improvement in environmental adaptation

suggests that, through continuous adaptation to new learning and living environments, students have strengthened their ability to adapt and their self-confidence in both social and campus environments.

Table 5. Comparative analysis of items at the beginning and end of the semester of the control group

Item	At the end of the semester (control group)	At the beginning of the semester (control group)	*T*-test
Warmth X_1	7.41±1.56	6.82 ±1.55	-1.19
Reasoning X_2	6.94±1.35	6.41±2.13	-1.08
Emotional stability X_3	6.18±1.25	5.44+2.01	-1.40
Dominance X_4	6.77±1.33	7.11±1.45	1.52
Liveliness X_5	7.17±1.68	8.33±1.75	1.35
Rule-consciousness X_6	5.62±1.59	4.22±1.46	-1.92
Social boldness X_7	6.89±1.87	7.71±1.64	0.94
Sensitivity X_8	5.86±1.87	6.31±1.69	1.33
Vigilance X_9	4.84±1.79	5.18±2.12	0.79
Abstractedness X_{10}	5.35±1.32	5.12±1.01	-0.49
Privateness X_{11}	6.33±1.02	5.98±1.53	-1.69
Apprehension X_{12}	6.66±1.01	5.98±1.53	-1.69
Openness to change X_{13}	5.79±1.97	5.84±1.22	0.11
Self-reliance X_{14}	5.33±1.25	5.75±1.57	0.62
Perfectionism X_{15}	4.21±1.52	4.31±1.47	0.03
Tension X_{16}	5.52±1.47	5.15±1.66	-1.55

However, the scores for some other factors have decreased, although these changes are not statistically significant. This may be related to the increased pressure and heavier academic tasks throughout the semester, which have led students to gradually show more rational and objective self-evaluation. For example, factors such as creativity and enthusiasm may be affected by negative psychological influences, as reflected in the decrease in scores. This phenomenon suggests that, while students have gradually matured in coping with academic and life challenges, they still have certain shortcomings in stimulating creative thinking and maintaining enthusiasm. Therefore, certain proactive psychological traits, such as creativity and enthusiasm, require further stimulation and cultivation.

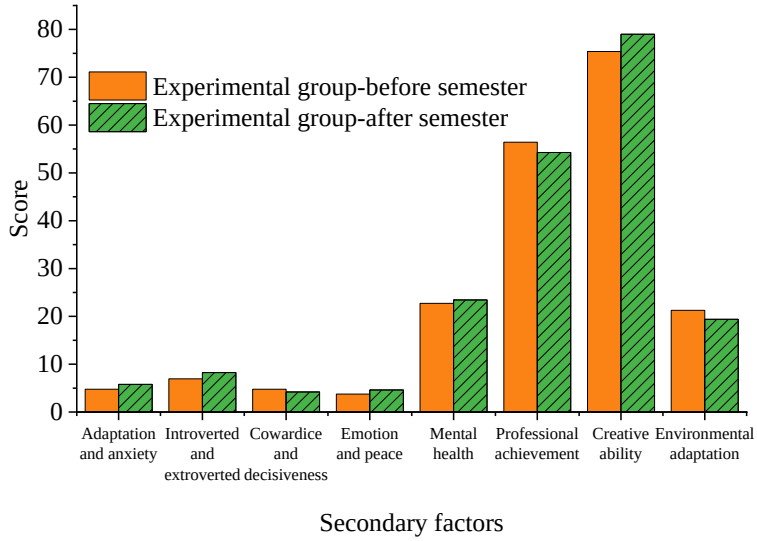


Fig. 7. Comparison of secondary factors of 16PF in the control group

In summary, after the test results of the experimental or control groups at the beginning of the semester are compared and analyzed, the sensitivity and perfectionism differences are significant ($P<0.05$). Moreover, the results show that there is a significant difference in the sensitivity factor in the experiment results at the beginning of the semester ($P<0.05$), and the difference in the automaticity factor is very significant ($P<0.05$). Both sensitivity and automaticity factors exhibit various changes than expected in the experimental data at the beginning of the semester, particularly the variations in automaticity factors. At the end of the semester, sensitivity and perfectionism don't show significant differences. Since the test duration is only one semester, whether for the experimental or control groups, there is a natural maturity in mind, social environmental adaptability, and surrounding environmental conditions. Consequently, it can be excluded that the related items of the control group are affected by the general mental health course during the experiment.

Through comparison, it is verified that inspirational films positively affect the mental health of college students and can help them build effective coordinated thinking to a certain extent. It has a positive impact in terms of interpersonal relationships and emotional management. Meanwhile, it can cultivate the positive thinking of college students and intervene to help them break traditional thinking. However, the impact on students' learning and thinking abilities is relatively small, and another approach is needed to complete the corresponding task. Moreover, when the voice interaction system is used for mental health consultations in the last class, it not only identifies students' concerns and provides helpful suggestions but also ensures privacy, allowing students greater freedom in addressing sensitive issues. By processing the input text, the system matches it with pre-set answers in the database. The most appropriate response is then delivered to the student's interactive interface, offering clearer insights into their concerns and effectively solving college students' mental health problems.

However, it is important to recognize that cultural background can have different effects on individuals' psychological traits and their responses to interventions. For example, Western cultures tend to emphasize individualism and self-expression, while Eastern cultures focus more on collectivism and interpersonal harmony. Therefore, students from diverse cultural backgrounds may experience emotional and behavioral changes brought about by films in different ways. Future research could conduct comparisons in various cultural contexts to explore how these cultural differences influence personality trait changes and the effectiveness of interventions. Additionally, factors such as gender, age, family background, and socioeconomic status may affect individuals' performance in areas such as emotion, social abilities, and self-efficacy. For example, students from different socioeconomic backgrounds may exhibit various coping strategies when facing stress and difficulties, and these differences could influence their reactions to inspirational films. Moreover, gender differences may be more pronounced in certain personality traits, such as warmth and perfectionism. Thus, future research could further validate the conclusions of this study by using a more diverse sample. It includes groups from different age ranges, genders, cultures, and socioeconomic backgrounds, to enhance this study's external validity.

5.5. Contrastive analysis

The AI-based mental health counseling system proposed in this study demonstrates notable advantages in emotion analysis and voice interaction. By using LSTM for emotion analysis and a knowledge graph-based semantic matching algorithm, the system can more accurately recognize and understand students' emotional states and semantic information. Existing AI-based mental health counseling systems primarily rely on traditional emotion classification models. These models generally only identify the polarity of emotions (e.g., positive, negative, neutral) and exhibit certain limitations when dealing with complex and subtle emotional changes. The LSTM model employed in this study captures emotional changes through time-series information, especially in situations involving significant emotional fluctuations, showing superior performance compared to existing methods. Table 6 shows the performance of the proposed system in emotion classification tasks and compares it with traditional methods:

Table 6. Comparison of the accuracy in emotion classification tasks

Emotion classification method	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Traditional emotion classification model	75.4	72.1	73.6	72.8
The proposed model	94.3	92.4	92.7	93.0

In Table 6, after adopting the LSTM model, the accuracy, precision, recall, and F1 score of emotion analysis markedly improved. This illustrates that the proposed method has stronger advantages in capturing complex emotional fluctuations and multidimensional emotional expressions compared to traditional models.

In terms of semantic matching, traditional AI-based mental health counseling systems often rely on keyword matching or rule-based template responses, which cannot effectively understand complex contextual information and emotional needs. The proposed system uses a knowledge graph-based semantic matching algorithm, which understands the semantics of users' queries and dynamically adjusts the weights of candidate answers based on the emotion analysis results. Table 7 compares the accuracy of semantic matching between the proposed system and traditional keyword-matching methods:

Table 7. Comparison of the accuracy of semantic matching

Matching method	Matching accuracy (%)	Average response time (seconds)
Keyword matching	68.5	1.2
The proposed knowledge graph matching	84.7	0.9

It can be found that the proposed semantic matching algorithm notably outperforms the traditional keyword matching method in accuracy while optimizing response time. This indicates that the algorithm improves both semantic understanding accuracy and system response efficiency. Although the proposed system demonstrates innovation in emotion analysis and semantic matching, it still faces some challenges. First, due to the use of DL models and large-scale knowledge graphs, the system's response time may be affected under high-concurrency usage conditions. Second, in emotion analysis, while the LSTM model performs excellently in most scenarios, its accuracy may still decline when dealing with complex or multi-emotional expressions. For instance, in conversations involving rapid emotional intensity changes or irony, the system's emotion analysis model may not fully capture the students' true emotional states. In future research, improvements to the emotion analysis model could be made to enhance the system's accuracy in these complex scenarios.

6. Conclusion

The voice interaction system and self-efficacy analysis are adopted to analyze the effect of inspirational film appreciation courses on the mental health of college students. Two college freshmen classes are set as experimental and control groups. Inspirational film appreciation courses and interactive psychological consultation courses were added for students in the experimental group. By calculating the similarity of the input text and extracting the text features by the LSTM method, the answer with higher confidence is obtained. This answer is then fed back to the interactive interface of the student terminal to provide students with a certain reference. Students in the control group receive

traditional mental health courses. The conditions of the experimental and control groups at the beginning and end of the semester are quantified through the Cartel's 16PF questionnaire test, the voice interaction system satisfaction test, and the self-efficacy test. The influence of inspirational film appreciation courses on the mental health of college students is explored under an analysis of the voice interaction system combined with self-efficacy. However, the research sample selected in the research process is freshman students, and the research results may have certain limitations. At the same time, when the situation of students is studied, it is necessary to analyze the class-attending situation of 80 students to reduce the data deviation. In the era of AI, the voice interaction system and self-efficacy analysis of campus psychological consultation are adopted to investigate the impact of inspirational film appreciation courses on college students' mental health. It is of great significance for the reform of psychological education courses in colleges and universities, as well as the continuous improvement of related evaluation and management mechanisms. Meanwhile, it promotes the adoption of AI in college education.

Acknowledgment. This work was supported by 2023Ministry of Education, Humanities and social science research projects (No.23JDSZ3107) and 2023 Chongqing Education Scientific Planning Projects (No.K23YG2080467) and 2025 General Project of Humanities and Social Sciences Research of Chongqing Municipal Education Commission (No.25SKSZ028).

References

1. Marsh, H. W., Pekrun, R., Parker, P. D., et al.: The Murky Distinction Between Self-Concept and Self-Efficacy: Beware of Lurking Jingle-Jangle Fallacies. Journal of Educational Psychology, Vol. 111, No. 2, 331. (2019)
2. Hatlevik, O. E., Throndsen, I., Loi, M., et al.: Students' ICT Self-Efficacy and Computer and Information Literacy: Determinants and Relationships. Computers & Education, Vol. 118, 107-119. (2018)
3. Lloyd, S. A., Shanks, R. A., Lopatto, D.: Perceived Student Benefits of an Undergraduate Physiological Psychology Laboratory Course. Teaching of Psychology, Vol. 46, No. 3, 215-222. (2019)
4. Grange, C., Miller, A.: Teaching Introduction to Psychology: Promoting Student Learning Using Digital Storytelling and Community Engagement. International Journal of Teaching and Learning in Higher Education, Vol. 30, No. 1, 172-183. (2018)
5. Latikka, R., Turja, T., Oksanen, A.: Self-Efficacy and Acceptance of Robots. Computers in Human Behavior, Vol. 93, 157-163. (2019)
6. Dale, K. R., Raney, A. A., Ji, Q., et al.: Self-Transcendent Emotions and Social Media: Exploring the Content and Consumers of Inspirational Facebook Posts. New Media & Society, Vol. 22, No. 3, 507-527. (2020)
7. Paradis, K.: Types and Tropes: History and Moral Agency in Evangelical Inspirational Fiction. Christianity & Literature, Vol. 69, No. 1, 73-90. (2020)
8. Bharadiya, J.: A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. European Journal of Technology, Vol. 7, No. 1, 58-66. (2023)
9. Pillai, A. S.: Advancements in Natural Language Processing for Automotive Virtual Assistants Enhancing User Experience and Safety. Journal of Computational Intelligence and Robotics, Vol. 3, No. 1, 27-36. (2023)
10. Yeke, S.: Digital Intelligence as a Partner of Emotional Intelligence in Business Administration. Asia Pacific Management Review, Vol. 28, No. 4, 390-400. (2023)

11. Carolus, A., Augustin, Y., Markus, A., et al.: Digital Interaction Literacy Model–Conceptualizing Competencies for Literate Interactions With Voice-Based AI Systems. Computers and Education: Artificial Intelligence, Vol. 4, 100114. (2023)
12. Pan, S.: Design of Intelligent Robot Control System Based on Human–Computer Interaction. International Journal of System Assurance Engineering and Management, Vol. 14, No. 2, 558-567. (2023)
13. Caulley, D., Alemu, Y., Burson, S., et al.: Objectively Quantifying Pediatric Psychiatric Severity Using Artificial Intelligence, Voice Recognition Technology, and Universal Emotions: Pilot Study for Artificial Intelligence-Enabled Innovation to Address Youth Mental Health Crisis. JMIR Research Protocols, Vol. 12, No. 1, e51912. (2023)
14. Bérubé, C., Schachner, T., Keller, R., et al.: Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review. Journal of Medical Internet Research, Vol. 23, No. 3, e25933. (2021)
15. Yesildag, A. Y., Bostan, S.: Movie Analysis as an Active Learning Method: A Study With Health Management Student. The International Journal of Management Education, Vol. 21, No. 1, 100759. (2023)
16. Pan, X., Hu, B., Zhou, Z., et al.: Are Students Happier the More They Learn?–Research on the Influence of Course Progress on Academic Emotion in Online Learning. Interactive Learning Environments, Vol. 31, No. 10, 6869-6889. (2023)
17. Müller, N. M., Seufert, T.: Effects of Self-Regulation Prompts in Hypermedia Learning on Learning Performance and Self-Efficacy. Learning and Instruction, Vol. 58, 1-11. (2018)
18. Perera, H. N., Calkins, C., Part, R.: Teacher Self-Efficacy Profiles: Determinants, Outcomes, and Generalizability Across Teaching Level. Contemporary Educational Psychology, Vol. 58, 186-203. (2019)
19. Herman, K. C., Hickmon-Rosa, J., Reinke, W. M.: Empirically Derived Profiles of Teacher Stress, Burnout, Self-Efficacy, and Coping and Associated Student Outcomes. Journal of Positive Behavior Interventions, Vol. 20, No. 2, 90-100. (2018)
20. Shortway, K., Oganesova, M., Vincent, A.: Sexual Assault on College Campuses: What Sport Psychology Practitioners Need to Know. Journal of Clinical Sport Psychology, Vol. 13, No. 2, 196-215. (2019)
21. Daffin Jr, L. W., Jones, A. A.: Comparing Student Performance on Proctored and Non-Proctored Exams in Online Psychology Courses. Online Learning, Vol. 22, No. 1, 131-145. (2018)
22. Joo, Y. J., Park, S., Lim, E.: Factors Influencing Preservice Teachers' Intention to Use Technology: TPACK, Teacher Self-Efficacy, and Technology Acceptance Model. Journal of Educational Technology & Society, Vol. 21, No. 3, 48-59. (2018)
23. Masitoh, L. F., Fitriyani, H.: Improving Students' Mathematics Self-Efficacy Through Problem-Based Learning. Malikussaleh Journal of Mathematics Learning (MJML), Vol. 1, No. 1, 26-30. (2018)
24. Fuller, B., Liu, Y., Bajaba, S., et al.: Examining How the Personality, Self-Efficacy, and Anticipatory Cognitions of Potential Entrepreneurs Shape Their Entrepreneurial Intentions. Personality and Individual Differences, Vol. 125, 120-125. (2018)
25. De Simone, S., Planta, A., Cicotto, G.: The Role of Job Satisfaction, Work Engagement, Self-Efficacy and Agentic Capacities on Nurses' Turnover Intention and Patient Satisfaction. Applied Nursing Research, Vol. 39, 130-140. (2018)
26. DiGuiseppi, G. T., Meisel, M. K., Balestrieri, S. G., et al.: Resistance to Peer Influence Moderates the Relationship Between Perceived (But Not Actual) Peer Norms and Binge Drinking in a College Student Social Network. Addictive Behaviors, Vol. 80, 47-52. (2018)
27. Hutcheon, T. G., Lian, A., Richard, A.: The Impact of a Technology Ban on Students' Perceptions and Performance in Introduction to Psychology. Teaching of Psychology, Vol. 46, No. 1, 47-54. (2019)

28. Auerbach, R. P., Mortier, P., Bruffaerts, R., et al.: WHO World Mental Health Surveys International College Student Project: Prevalence and Distribution of Mental Disorders. Journal of Abnormal Psychology, Vol. 127, No. 7, 623. (2018)
29. Cuijpers, P., Auerbach, R. P., Benjet, C., et al.: The World Health Organization World Mental Health International College Student Initiative: An Overview. International Journal of Methods in Psychiatric Research, Vol. 28, No. 2, e1761. (2019)
30. Ameral, V., Palm Reed, K. M., Hines, D. A.: An Analysis of Help-Seeking Patterns Among College Student Victims of Sexual Assault, Dating Violence, and Stalking. Journal of Interpersonal Violence, Vol. 35, No. 23-24, 5311-5335. (2020)
31. Richardson, C. M. E., Trusty, W. T., George, K. A.: Trainee Wellness: Self-Critical Perfectionism, Self-Compassion, Depression, and Burnout Among Doctoral Trainees in Psychology. Counselling Psychology Quarterly, Vol. 33, No. 2, 187-198. (2020)
32. Pratt, I. S., Harwood, H. B., Cavazos, J. T., et al.: Should I Stay or Should I Go? Retention in First-Generation College Students. Journal of College Student Retention: Research, Theory & Practice, Vol. 21, No. 1, 105-118. (2019)
33. Jones, P. J., Park, S. Y., Lefevor, G. T.: Contemporary College Student Anxiety: The Role of Academic Distress, Financial Stress, and Support. Journal of College Counseling, Vol. 21, No. 3, 252-264. (2018)
34. Naudé, L., Jordaan, J., Bergh, L.: "My Body Is My Journal, and My Tattoos Are My Story": South African Psychology Students' Reflections on Tattoo Practices. Current Psychology, Vol. 38, No. 1, 177-186. (2019)
35. Schinke, R. J., Stambulova, N. B., Si, G., et al.: International Society of Sport Psychology Position Stand: Athletes' Mental Health, Performance, and Development. International Journal of Sport and Exercise Psychology, Vol. 16, No. 6, 622-639. (2018)
36. Glodowski, K., Thompson, R.: The Effects of Guided Notes on Pre-Lecture Quiz Scores in Introductory Psychology. Journal of Behavioral Education, Vol. 27, No. 1, 101-123. (2018)
37. Pitcher, E. N., Camacho, T. P., Renn, K. A., et al.: Affirming Policies, Programs, and Supportive Services: Using an Organizational Perspective to Understand LGBTQ+ College Student Success. Journal of Diversity in Higher Education, Vol. 11, No. 2, 117. (2018)
38. Bamber, M. D., Morpeth, E.: Effects of Mindfulness Meditation on College Student Anxiety: A Meta-Analysis. Mindfulness, Vol. 10, No. 2, 203-214. (2019)
39. Ellis, J. M., Powell, C. S., Demetriou, C. P., et al.: Examining First-Generation College Student Lived Experiences With Microaggressions and Microaffirmations at a Predominately White Public Research University. Cultural Diversity and Ethnic Minority Psychology, Vol. 25, No. 2, 266. (2019)
40. David, J. L., Powless, M. D., Hyman, J. E., et al.: College Student Athletes and Social Media: The Psychological Impacts of Twitter Use. International Journal of Sport Communication, Vol. 11, No. 2, 163-186. (2018)
41. Savage, M. W., Strom, R. E., Ebesu Hubbard, A. S., et al.: Commitment in College Student Persistence. Journal of College Student Retention: Research, Theory & Practice, Vol. 21, No. 2, 242-264. (2019)

Shaohua Fan was born in Lanzhou, Gansu Province, China in 1980. He obtained a master's degree from Gansu Agricultural University. He currently working at the School of Business Administration, Chongqing Technology and Business University.His research interests include Educational Psychology and College Student Education Management. E-mail: fanshaohua@ctbu.edu.cn

Yujing Song was born in 1981 in Xichang City, Sichuan Province, China. She Obtained a bachelor's degree from Chongqing Normal University. She currently

working at the Wealth Management School, Chongqing Finance and Economics College, her research focus on education. E-mail:13452366717@163.com

Received: December 05, 2024; Accepted: March 05, 2025.

Computer Science and Information Systems 22(3):1331–1346 https://doi.org/10.2298/CSIS241212045S

Leveraging AI and Diffusion Models for Anime Art Creation: A Study on Style Transfer and Image Quality Evaluation

Chao-Chun Shen[1], Shun-Nian Luo[2], Ling Fan[3,*], Chenglin Dai[4]

[1]School of Art Design and Media, Sanda University, China
ghinishen@163.com
[2]School of Information Science and Technology, Sanda University, China
snluo@sandau.edu.cn
[3]Shanghai Technical Institute of Electronics & Information College, China
FL0514@126.com
[4]School of Information Science and Technology, Sanda University, China
ad88105506@163.com

Abstract. The remarkable advancements in artificial intelligence (AI)-driven image generation technologies have brought about a profound transformation across various industries, particularly in new media, video production, and gaming. AI-generated content (AIGC) has emerged as a game-changing, cost-efficient solution for companies seeking high-quality visual assets while operating within constrained budgets and having limited access to traditional human resources. Through the use of sophisticated algorithms, AIGC enables the creation of stunning visuals without relying on conventional, labor-intensive workflows. Among the most prominent techniques, diffusion models have played a pivotal role in the development of AI image generation tools, giving rise to both proprietary platforms like Midjourney and open-source alternatives such as Stable Diffusion. These technologies continue to evolve, benefiting from the collaborative contributions of global programming communities.

This study focuses on advancing the capabilities of Stable Diffusion, an open-source AI image generation model, to address prevalent challenges in style consistency and image quality. By integrating Python and harnessing cutting-edge AI techniques, such as DreamBooth and embedding methods, the research aims to enhance the model's ability to replicate and embed distinct artistic styles. Specifically, the study targets the unique art style of the popular mobile game "Arknights" as a training objective, applying advanced techniques to refine the system's output. The proposed approach demonstrates significant improvements over the baseline model, showcasing enhanced performance in generating style-consistent anime imagery. This research contributes to the evolving landscape of AI-driven art generation, offering novel insights into the application of diffusion-based technologies within creative industries. By utilizing DreamBooth and embedding for style transfer and injection, the study achieves notable efficiency, drastically reducing the time required to train a new model. Ultimately, this work paves the way for more specialized and customizable AI systems in art creation, pushing the boundaries of what AI can achieve in the realm of creative expression.

* Corresponding author

Keywords: AI-Generated Content (AIGC); Diffusion Model; Academic Affairs Management

1. Introduction

The development of AI-generated art can be traced back to advancements in artificial intelligence and computer vision technologies. Early computer graphics (1960s-1980s) focused on technical and scientific drawing, laying the groundwork for later AI art. The rise of neural networks in the 1980s, particularly with the introduction of the backpropagation algorithm, set the stage for deep learning. In the 2010s, deep learning breakthroughs, such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), enabled significant progress in image generation. The advent of style transfer techniques in 2015 further advanced AI art, allowing the fusion of artistic styles with content images, opening new possibilities for AI-driven artistic creation.

Since the dawn of human civilization, art has always accompanied our development. Humans receive external stimuli like sound waves, electromagnetic waves, and pressure via their senses, process them through the central nervous system, and present them in an abstract form in their minds. These abstract concepts are expressed through diverse art forms, such as sculpture and painting. Among them, painting is the most fundamental art expression. Humans transmit the light received by their eyes to the visual center of the cerebral cortex as neural signals. After countless neurons in the visual center process these neural impulses, they become what humans see and imagine, ultimately being presented on a canvas through their hands and feet.

Researchers studying AI painting are also simulating the human brain's operation to build neural network models, enabling AI to learn to draw images that conform to human cognition and aesthetics [1]. The objective of this study is to explore how to achieve image style transfer and artistic style injection using state-of-the-art technologies with a small model, thus improving image quality. This study focuses on combining or enhancing existing style transfer and injection techniques for fast style transfer in small-model-generated images. In existing research, there are no cases of combining Embedding and Dreambooth technologies. Additionally, this paper proposes using FID and MLE simultaneously for cross-analysis to evaluate image quality, which are the key technical innovations of this study.

AI-generated illustrations directly address these issues, helping companies in the social media and gaming industries provide fast and convenient services. They aid startups in cost savings and enable mature art companies to manage art outsourcing [2]. Aiming at commercial companies' need to create anime illustrations and character stand-alone images, an anime image generation system based on the diffusion model solves problems like low output efficiency, inconsistent art styles, high commissioning costs, and unstable delivery times in traditional artist commissioning. It meets commercial demands for efficiency, economy, stability, and convenience in product output, adapting to new business models. This project conducts a basic experiment based on an AI painting system to draw some anime images, specifically avatars, to explore the commercial applicability of AI painting [3].

Taking the open-source model Stable Diffusion as an example, some AI painting models may produce unstable output images due to model architecture or training

process instabilities, leading to significant quality differences and inconsistencies among generated images. Meanwhile, insufficient training data, monotonous samples, or overly strong regularization during training may cause the model to repeatedly generate a large number of extremely similar or detail - less images. This situation is known as "mode collapse" in the industry. The image distribution under mode collapse is clearly far from the real image distribution. To reduce the model's training cost, this paper aims to improve the original Stable Diffusion model, enhancing its stability and the quality of output images.

This paper will leverage the foundational components of the open-source model Stable Diffusion and employ Dreambooth and embedding techniques to fine-tune the model by injecting custom themes for stylistic adjustments. The model will be evaluated using Maximum Likelihood Estimation (MLE) and Fréchet Inception Distance (FID) as metrics. The training objective is to develop a model capable of generating style-consistent images similar to the game's character illustrations, with better performance in maximum likelihood estimation compared to the original model.

2. Literature Review

This section introduces the technical principles and evaluation criteria for model performance of the AI drawing model applied in this design [4]. It mentions the basic principle of the industry's mainstream image generation model, stable diffusion, which utilizes noise as random numbers to generate images.

The Diffusion model is a neural network model that takes descriptive text, random noise, and a sequence of time steps as input and outputs an image. By repeatedly applying denoising operations to the generated random noise image, the Diffusion model produces an output image at each step, which serves as the input for the next denoising operation. This process continues until the number of denoising steps reaches a predefined total, at which point the resulting image is the model's final output. The descriptive text determines the style and content of the image, while the random noise serves as the initial input. Each generated image is assigned a time step number (decreasing sequentially), with the core process being the generation of an image from noise. As illustrated in Figure 1:

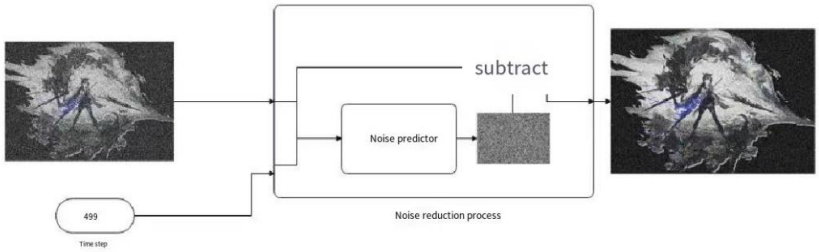


Fig. 1. Forward Propagation Flowchart of the Diffusion Model

The process of continuously subtracting the noise predicted by the noise predictor from the random noise eventually results in a brand-new image being generated from a

completely random noise input [5]. This process of generating an image from noise is referred to as reverse diffusion.

The training of the noise predictor, on the other hand, is known as forward diffusion. Its principle is depicted in Figure 2:

By using both the original image and a version of the image with added noise as input, the noise predictor is trained to output the added noise. Thus, the essence of Diffusion technology lies in teaching neural networks to reverse the process of adding noise to images, thereby enabling image generation. Its principle is depicted in Figure 2:

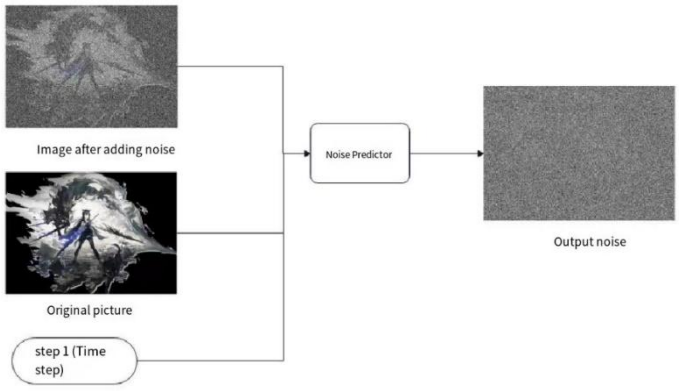


Fig. 2. Reverse Propagation Flowchart of the Diffusion Model

To quantitatively evaluate the quality of images generated by a model, researchers primarily use Maximum Likelihood Estimation (MLE) to assess both the model and the generated images. The idea behind Maximum Likelihood Estimation is that, for a given set of observed data x, we aim to find the parameter θ^* among all possible θ_1, θ_2, ..., θ_n that maximizes the probability of generating the observed data. This leads to the formula 1:

$$\theta^* = \underset{\theta}{argmax} p(x|\theta) \quad (1)$$

In the training process of AI-generated images, we randomly sample mmm data points from the real image distribution $Pdata(x)$, then compute the probability of each data point occurring under the model's generated image distribution $P_\theta(x)$, and multiply these probabilities together. The θ that maximizes this final probability is the parameter θ^* that makes the neural network produce images closest to real images, as shown in formula 2:

$$\theta^* = \underset{\theta}{argmax} \prod_{i=1}^{m} P_\theta(x^i) \quad (2)$$

FID (Fréchet Inception Distance) is a metric used to evaluate the quality of generative models.

FID is calculated based on the Fréchet distance between two probability distributions: one representing the distribution of real images and the other representing the distribution of images generated by the generator model. Specifically, FID quantifies

this gap by computing the feature representations of these two distributions within a pre-trained Inception network and then calculating the Fréchet distance between these representations.

A lower FID value indicates a smaller gap between the generated images and the real image distribution, thus indicating better performance of the generative model. Based on the FID value, the performance of the generative model can be interpreted and evaluated.

The calculation of FID typically involves the following steps: computing the feature means and covariance matrices for both the real image dataset and the generated image dataset, as shown in formulas 4 and 5:

$$\left\|\mu_{real} - \mu_{gen}\right\|_2^2 \quad (4)$$

$$d^2\left(\Sigma_{real}, \Sigma_{gen}\right) = \left\|\Sigma_{real} + \Sigma_{gen} - 2\left(\Sigma_{real}\Sigma_{gen}\right)^{\frac{1}{2}}\right\| F \quad (5)$$

Then, the Fréchet distance between them is calculated, reflecting the similarity between the two distributions.

$$\mathrm{FID} = \left\|\mu_{real} - \mu_{gen}\right\|_2^2 + Tr\left(\Sigma_{real} + \Sigma_{gen} - 2\left(\Sigma_{real}\Sigma_{gen}\right)^{\frac{1}{2}}\right) \quad (6)$$

Table 1. Functional Modules Table

Neural Network Layers	Function
Preprocessing Layer	Uniformly process the size, pixel, and other attributes of the images being trained
Noise Predictor (Forward Diffusion Layer)	By inputting both the original image and the image with added noise, it trains its ability to predict the noise.
Reverse Diffusion Layer	By combining with the noise output by the noise predictor, it achieves the effect of generating an image from complete noise.
Dreambooth Model	Inject custom art style into the existing image drawing model.
FID (Fréchet Inception Distance)	FID is used to calculate the closeness between the distribution of generated images and that of real images.
Maximum Likelihood Estimation Layer	Among the existing parameters, Maximum Likelihood Estimation (MLE) is utilized to identify a set of neural network parameters that make the distribution of images generated by the model closest to the distribution of real images.

3. Methodology

This section primarily introduces the primary modules and hierarchical structure of the image generation system. The system comprises the Rendering Module, Style Injection Module, and Data Assistance Module.

3.1. System Implementation

The system adopts the stable diffusion model as its fundamental architecture, utilizing the open-source model darkSushiMixMix as the initial training model. These two open-source components have already implemented the functions of the Forward Diffusion Layer and Reverse Diffusion Layer, enabling the system to randomly generate images with various styles and details, fine-tuned according to different prompt words. However, the current primitive system lacks optimization for specific art styles and suffers from severe overfitting and extreme lack of diversity when inputting numerous prompt words. The purpose of this system is to build upon these two modules, enabling the neural network to learn the prompt word "Arknights" and, upon inputting this prompt, output images resembling the art style of the "Arknights" series of illustrations. Additionally, the system aims to inject the essence of the Arknights art style into the model.

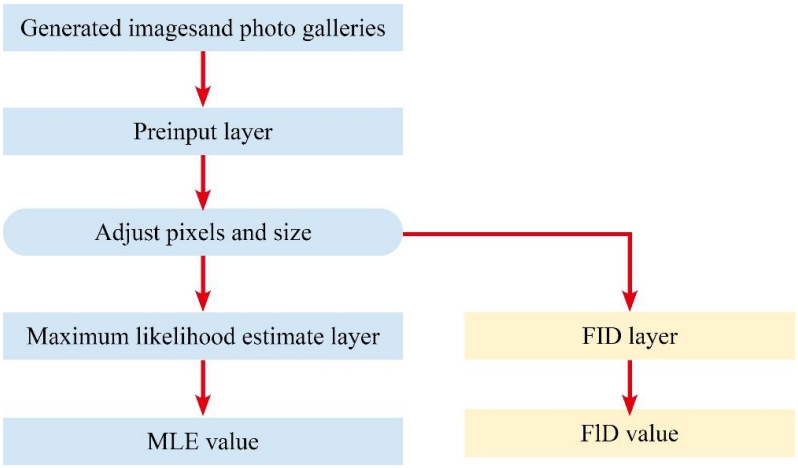


Fig. 3. Flowchart of the Data Assistance Module

As mentioned above, Figure 3 illustrates the operation process of the model evaluation system, which inputs the image, initializes, and calculates MLE and FID.

3.2. Data Assistance Module

The Data Assistance Module comprises the Preprocessing Layer, MLE Calculation Layer, and FID Layer. The input of this module is the generated images and a real image set, and the output is the MLE and FID values for the generated images compared to the real image set.

The Preprocessing Layer aims to standardize the dimensions and pixels of input images, simplifying MLE and FID value calculations. First, it uses the Python imread() command from cv2 to load generated and real images in color format. Then, the input images are resized to 512*512 pixels, enabling the MLE and FID layers to extract more representative features and enhance the reliability of the final MLE and FID outputs.

The MLE Layer takes the images processed by the Preprocessing Layer as input. Its output is the KL divergence between these images and the image set. After extracting the feature vectors, the MLE value is calculated. Since the MLE value equals the KL divergence value [6], the program directly computes the KL divergence between the generated and real image sets.

The FID Layer receives images from the Preprocessing Layer, extracts their feature vectors, and calculates the FID value. In FID calculations, the feature vector distributions of real and generated images are regarded as two high - dimensional Gaussian distributions, and their Fréchet Distance is computed. This approach allows FID to offer a more comprehensive evaluation of image quality.

3.3. Image Rendering Module

This module is composed of the open-source generative model framework, Stable Diffusion, and the generative model DarkSushiMixMix_225D, which is a Diffusion Model within the Stable Diffusion framework. In the Stable Diffusion framework, various Diffusion Models can be employed to achieve diverse artistic styles for image generation, such as realism, ink painting, abstraction, and more [7]. These models are typically implemented based on deep neural networks, necessitating the use of corresponding pre-trained models for parameter initialization and inference [8].

The Style Injection Module leverages Dreambooth and Embedding techniques to inject desired keywords and artistic styles into DarkSushiMixMix_225D without retraining the model from scratch. Current large AI models, also known as foundation models, have been trained on billions of data points, making them highly generalized, versatile, and practical for various drawing scenarios. However, these models often struggle to meet specific requirements for detail control or particular drawing styles. To address this issue efficiently in terms of time and cost, researchers have proposed fine-tuning techniques for large models, including Dreambooth and Embedding.

Dreambooth, introduced by Google in August 2022, is a novel deep learning technique for fine-tuning existing text-to-image models [9]. Its goal is to generate more detailed and personalized output images by fine-tuning pre-trained text-to-image models. It enables users to "feed" custom image information to the model and generate diverse images through simple names and prompts, while preserving the critical visual

features desired by the user. By fine-tuning the model with just 3-5 images and text prompts, Dreambooth can effectively generate new images that accurately replicate the appearance of the input images. This study utilizes the Dreambooth website provided by Google. After providing the training set, the corresponding Dreambooth file is automatically generated after a period of time, which can then be imported for use [10].

In a trained model, the Text Coder functions as a dictionary, combining input text and word vectors to guide the UNET in initializing noisy images [11].

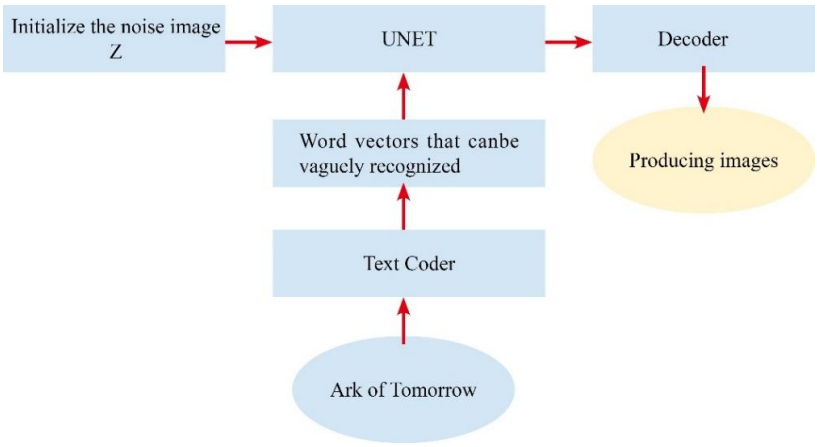


Fig. **4**. Flowchart of Embedding Algorithm

However, when encountering novel keywords that are not part of the Text Coder's vocabulary, traditional approaches would require retraining the entire model, significantly reducing its flexibility. To address this, the Embedding algorithm is introduced. It trains the Text Coder to find word vectors that share similar characteristics and styles with the new keywords, enabling fine-tuning of the model without altering its fundamental structure or Text Coder.

4. Model Evaluation

This section generated 100 images using the original generative model, and calculated 100 MLE and FID values respectively. Through the scatter plot, an intuitive distribution of image quality was obtained [12]. The corresponding indicators for evaluating model performance were then obtained by calculating the average values of MLE and FID. After that, the Embedding and Dreambooth technologies, which can cleverly avoid the difficulties of large model training, such as high resource consumption and poor effect, were introduced [13].

4.1. Introduction to the Real Image Dataset

The system employs 200 images of Arknights characters collected from the authoritative source station, Prts. Arknights, as the training targets for the model to generate images, serving as the real image dataset. To ensure the stability of model training, all selected images for the real image dataset exhibit the following characteristics: 1) They belong to the mature and consistent art style of Arknights. 2) There are no apparent drawing inconsistencies or errors. 3) The backgrounds are clean, devoid of redundant interfering elements. All images are in JPG format.

As depicted in Figures 5 and 6, the anime images produced by the original DarkMix model exhibit issues such as unclear hierarchical structures, indistinct art styles, and severe lack of details. In Figure 5, the female character's facial details are severely lacking, with important features like the nose and mouth missing [14]. Figure 6 shows a girl with a bizarre hairstyle that seemingly blends a ponytail with loose hair, and an unnatural connection between the head and body. It is evident that the initial model-generated images have rough craftsmanship, distorted characters, and dull expressions. Additionally, many details in the characters' clothing and hair are lost, making the images aesthetically unpleasing to the average viewer.

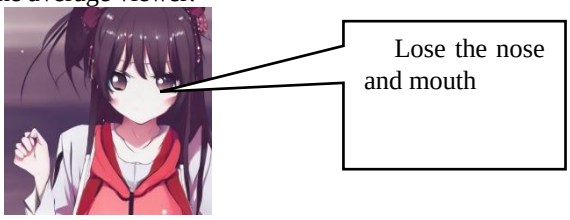


Fig. 5. Generated image of the original model 1

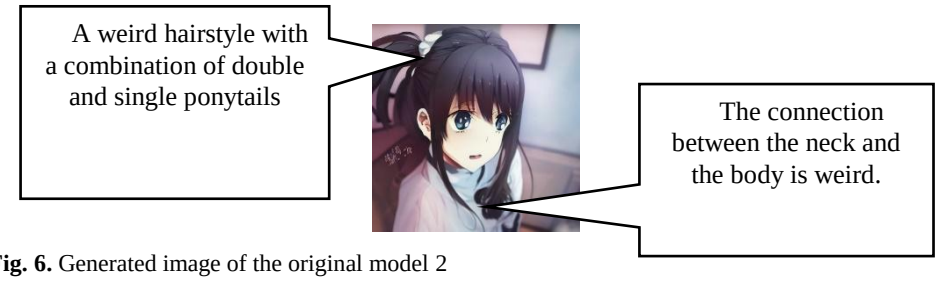


Fig. 6. Generated image of the original model 2

We randomly generated 100 images with the original model and input them into a data auxiliary module to calculate MLE and FID. The results are presented in the scatter plot in Table 4-4, where the horizontal axis represents the maximum likelihood estimate (MLE) values obtained by comparing the 100 generated images with the real image dataset, and the vertical axis corresponds to FID values. Since the maximum MLE is equivalent to the minimum KL divergence, we use KL divergence as a substitute for MLE for convenience. It is evident that points closer to the lower left corner indicate higher similarity between generated and real images. For the original model, MLE values mostly ranged from 0.5 to 1.5, with an average of 1.36882, while FID values

were generally distributed between 95 and 115, averaging 105.6795. Using the MLE and FID scatter plot provides a more intuitive evaluation of model performance.

Fig. 7. Scatter plot of image quality

4.2. Output Image Set and Performance of the Current Model

After injecting the art style using the aforementioned methods, as depicted in Figures 8 and 9, we observe improved image detail richness, clearer and smoother lines, and more vivid character expressions [15]. Figure 8 shows a female character with normal noses and mouths, distinct lines, and high recognizability. Figure 9 reveals a girl with a three-dimensional bangs and ponytail, and naturally flowing hair, eliminating the chaotic appearance. It is evident that the images generated by the existing model exhibit refined craftsmanship, well-defined body curves, and clear expressions. Additionally, more details are present in the characters' clothing and hair, making the images aesthetically pleasing to the average viewer.

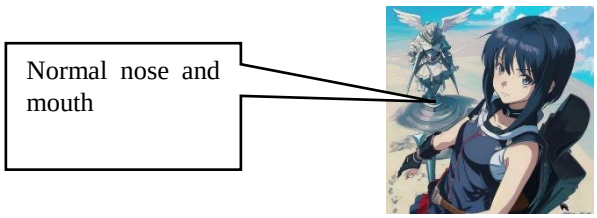


Fig. 8. Output Image 1 from the Current Model

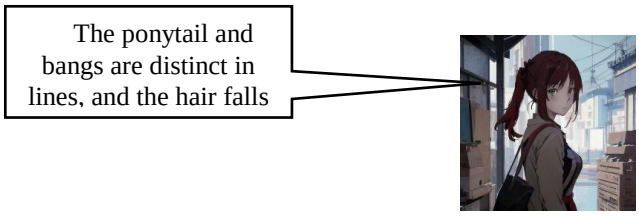


Fig. 9. Output Image 2 from the Current Model

As shown in Figure 10, it can be observed that the anime images generated by the current model have the highest distribution of MLE values within the range of 0.2 to 1, with an average of 1.15472. The FID values are generally distributed within the range of 80 to 120, with an average of 100.9565. It can be noticed that both the average values of MLE and FID have decreased, and most of the scatter points are closer to the lower left corner compared to before. Therefore, it can be preliminarily judged that the current model is superior in performance to the original model.

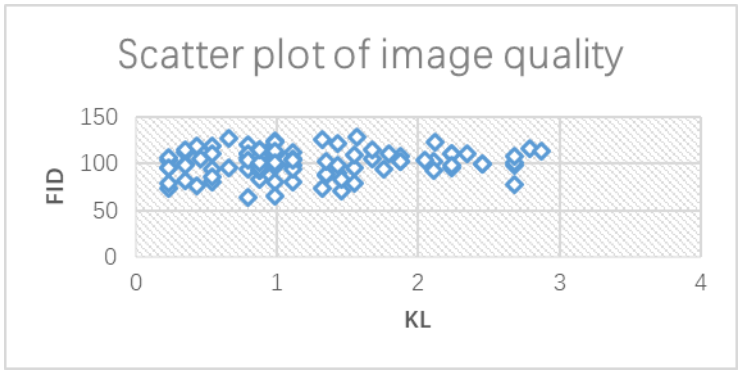


Fig. 10. Scatter plot of image quality

4.3. Implementation of Data Assistance Module

The data assistance module loads and preprocesses images from a folder, unifies their sizes, and passes the image list to the FID and MLE calculation layers. Finally, it generates a corresponding scatter plot based on the data provided by the FID and MLE layers. The functional flow of this module is illustrated in Figure 11.

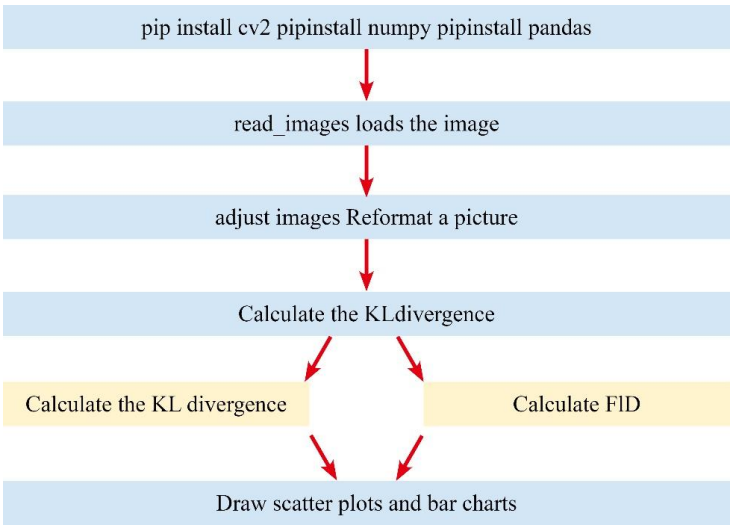


Fig. 11. Data Assistance Module Flowchart

The primary function of the preprocessing layer is to uniformly process the images from the input data assistance module by reading them into an image list from a specified folder using the cv.imread() function and resizing each image to 512*512 pixels [16]. After reading the images, you can call the custom testing function show_image(image_paths) to display the images, verifying whether they have been successfully loaded into the system. The function show_image() creates a window to display the image and waits for any user input before closing the window. Once the read_images() function is verified to be working correctly, the image list is fed into the adjust_images() function for preprocessing. This function iterates through each image in the list and resizes it to 512*512 pixels using the cv2.resize() function, finally placing all the images into a new list.

The FID (Fréchet Inception Distance) and MLE (Maximum Likelihood Estimation, approximated by KL divergence) calculation layers are primarily designed to convert the list of preprocessed images into one-dimensional arrays, extract their feature vectors, and utilize dedicated FID and KL divergence computation functions to obtain the FID distance and KL divergence between images. These metrics are then plotted on a scatter plot for visualization.

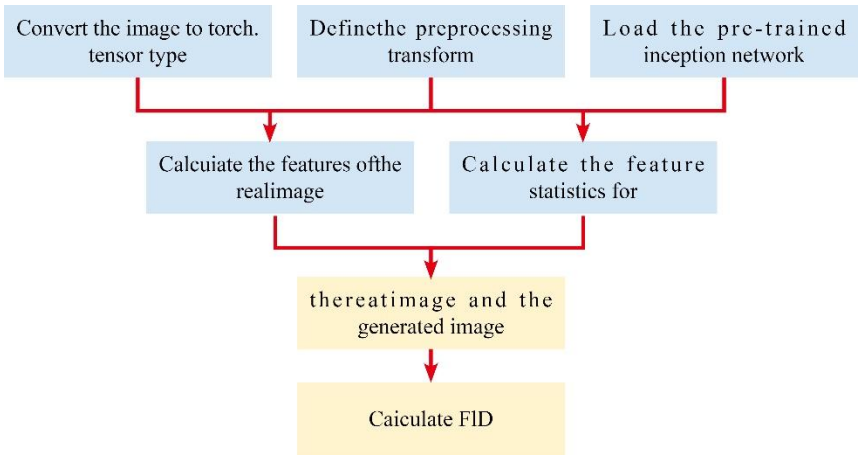


Fig. 12. FID Layer Function Flow Chart

Figure 12 illustrates the operation process of the FID calculation system, which initializes the model for computation, calculates the features of the two images, and computes their FID.

5. Conclusions

The purpose of this system is to optimize the existing image generation model technology, aiming to obtain a model with better image generation performance [17]. Additionally, the system proposes to adopt both FID and MLE as indicators for evaluating the quality of the generated model and uses these two metrics to optimize the model. Based on the original diffusion model framework, Dreambooth and Embedding technologies are employed as optimizations, addressing issues such as long training time, difficulty in training, high energy consumption, and poor results successfully lightening the problem.

Specifically, Dreambooth endows the model with more powerful detail depiction and style transfer capabilities [18]. The Embedding technology successfully enables the system to create content in a specific style based on the keyword "Arknights."

Ultimately, the images generated by this model successfully surpass the original model's images in terms of FID and MLE indicators, demonstrating that this system has better image generation performance.

The optimization of the original image generation system using Embedding and Dreambooth technologies has been achieved, enabling the model to surpass the original model's image quality in specific art styles. A method for evaluating the model's performance using a combination of MLE and FID with scatter plots has been proposed. After optimization, the MLE of the model decreased by approximately 2.4, and the FID decreased by approximately 5, indicating that the distribution of images generated by the optimized model is closer to the real distribution of images than before.

Overlapping the two scatter plots reveals that most of the points from the optimized model are significantly closer to the bottom left and more dispersed than those of the original model [19]. This suggests that the distribution of images generated by the optimized model is closer to the real distribution of images but there are instances where

the quality of generated images declines after adjustment, indicating reduced stability in the new model. It is speculated that this is due to overfitting of the model to certain specific images. Thus, while lightweight injection techniques like Dreambooth and Embedding significantly improve training efficiency and avoid the high costs of retraining large models, they may also somewhat reduce model stability [20].

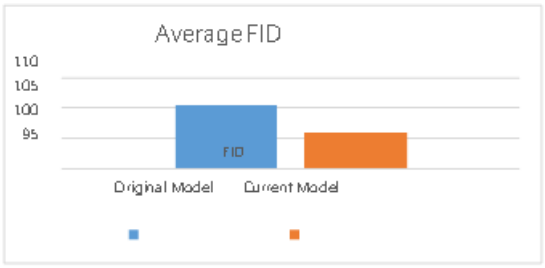


Fig. 13. Average FID

As shown in Figure 13, after optimization with Dreambooth and Embedding, the average FID decreased by approximately 5. FID, as an important indicator of the distance between two images (i.e., image similarity), indicates that the lower the FID, the more similar the two images are, meaning the generated image is closer to a real image. Table 5-1 proves that from the FID evaluation metric, the image generation quality of the current model is slightly higher than that of the original model, demonstrating better specialized generation capabilities.

As shown in Figure 14, after optimization with Dreambooth and Embedding, the average KL decreased by approximately 0.2. KL, as an indicator of the similarity between an image and the distribution of an image set, is also a crucial metric for image similarity. A lower KL indicates a higher likelihood that the image is real, meaning the generated image is closer to a real image. Table 5-1 proves that from the KL evaluation metric, the image generation quality of the current model is slightly higher than that of the original model, excellently fulfilling the task of generating images in specific art styles.

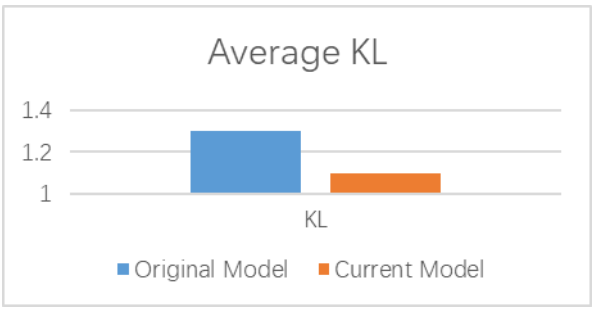


Fig. 14. Average KL

Our research still has many limitations, such as overfitting issues and room for improvement in image quality. To address these problems, we have considered the following improvement methods, such as increasing the model size and enhancing the diversity of the training set images.

Acknowledgment. The research is conducted under the Humanities and Social Sciences Planning Fund Project of the Ministry under the title of "Research on the Social Media Communication of Mainland China's Image in Taiwan", Project No. 24YJAZH122, with special thanks.

References

1. Alex Krizhevsky, Ilya Sutskever, Geoffrey E.Hinton. ImageNet Classification with Deep Convolutional Neural Networks. (2012).
2. Dennis Elbrachter, Dmytro Perekrestenko, Philipp Grohs, and Helmut Boelcskei. 2021. Deep Neural Network Approximation Theory. IEEE Transactions on Information Theory (2021).
3. Daneshfar, Fatemeh, Ako Bartani, and Pardis Lotfi. "Image captioning by diffusion models: a survey." *Engineering Applications of Artificial Intelligence* 138 (2024): 109288.
4. Paiva, José Carlos, José Paulo Leal, and Álvaro Figueira. "Comparing semantic graph representations of source code: The case of automatic feedback on programming assignments." Computer Science and Information Systems 00 (2024): 4-4.
5. Brokman, Jonathan, et al. "MONTRAGE: Monitoring Training for Attribution of Generative Diffusion Models." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024.
6. Alimisis, Panagiotis, et al. "Advances in diffusion models for image data augmentation: A review of methods, models, evaluation metrics and future research directions." *Artificial Intelligence Review* 58.4 (2025): 1-55.
7. Osorio, Pedro, et al. "Latent diffusion models with image-derived annotations for enhanced ai-assisted cancer diagnosis in histopathology." *Diagnostics* 14.13 (2024): 1442.
8. Blake Bullwinkel, Kristen Grabarz, Lily Ke, Scarlett Gong, Chris Tanner, and Joshua Al-len. 2022. Evaluating the fairness impact of differentially private synthetic data. arXiv preprint arXiv:2205.04321 (2022).
9. Turner, D. Bruce. "A Diffusion Model for an Urban Area." Journal of Applied Meteorolo-gy and Climatology, vol. 3, no. 1, 1964, pp. 83–91.
10. Rahman, Abidur, et al. "Implementation of diffusion model in realistic face generation." *2024 9th International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 2024.
11. Rogers, Everett M. "A Prospective and Retrospective Look at the Diffusion Model." Jour-nal of Health Communication, vol. 9, no. S1, 2010, pp. 13–19.
12. Juan Miguel Lopez Alcaraz and Nils Strodthoff. 2022. Diffusion-based time series imputation and forecasting with structured state space models. arXiv preprint arXiv:2208.09399 (2022).
13. Kidder, Benjamin L. "Advanced image generation for cancer using diffusion models." *Biology Methods and Protocols* 9.1 (2024): bpae062.
14. Samy Bengio and Yoshua Bengio. 2000. Taking on the curse of dimensionality in joint distributions using neural networks. IEEE Transactions on Neural Networks (2000).
15. Croitoru, Florinel-Alin, et al. "Diffusion Models in Vision: A Survey." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 9, 2023, pp. 10850–10869. DOI: 10.1109/TPAMI.2023.3261988.
16. Moser, Brian B., et al. "Diffusion models, image super-resolution, and everything: A survey." *IEEE Transactions on Neural Networks and Learning Systems* (2024).

17. Tomer Amit, Eliya Nachmani, Tal Shaharbany, and Lior Wolf. 2021. SegDiff: Image segmentation with diffusion probabilistic models. arXiv preprint arXiv:2112.00390 (2021).
18. Peter, Ojonugwa Oluwafemi Ejiga, Md Mahmudur Rahman, and Fahmi Khalifa. "Advancing AI-Powered Medical Image Synthesis: Insights from MedVQA-GI Challenge Using CLIP, Fine-Tuned Stable Diffusion, and Dream-Booth+ LoRA." *Conference and Labs of the Evaluation Forum*. 2024.
19. Turker, Anil, and Ender M. Eksioglu. "3D convolutional long short-term encoder-decoder network for moving object segmentation." *Computer Science and Information Systems* 21.1 (2024): 363-378.
20. Andrew Campbell, Joe Benton, Valentin De Bortoli, Tom Rainforth, George Deligiannidis, and Arnaud Doucet. 2022. A continuous time framework for discrete de-noising models. arXiv preprint arXiv:2205.14987 (2022).

ChaoChun Shen is currently Professor in the School of Art Design and Media, Department of Network and New Media, Sanda University. His research interests include AIGC in new media, Social media big data, content algorithm, News communication analysis. He has served as chief editor and general manager of new media company, there are more than 40 related papers, including SCI, EI, CSSCI. He is member of IEEE.

Shun Nian Luo graduated with a Doctor of Philosophy in Information Management from Dayeh University, Taiwan. He has previously worked at institutions in Taiwan such as Chinfon Securities, Mitsukoshi-MetLife Insurance, and Taojiang University of Science and Technology. Since 2020, he has been working at the School of Information Science and Technology, Shanghai Sanda University. He has won awards including the Third Prize in the College Middle-aged Group of the 7th Shanghai Teachers' Calligraphy Competition and the Excellent Instructor Award in the Practice Competition of Huawei ICT Competition - Shanghai Division. His research areas include information security, system analysis and design, Unity game engine, 3D animation, robot ROS (Robot Operating System), and embedded systems.

Ling Fan is an associate professor at the School of Shanghai Technical Institute of Electronics & Information College of designt and art. Her areas of expertise include gamified teaching, AI-based art creation, new media communication, artificial intelligence, and art education.

Chenlin Dai is a student who has been admitted to the master's program at the University of Southern California (USC) this year. He majored in software engineering during his undergraduate studies.

Received: December 12, 2024; Accepted: March 07, 2025.

Computer Science and Information Systems 22(3):1347–1360 https://doi.org/10.2298/CSIS241218048L

Usage Intention of the Reservation System of Taipei Sports Center from the Perspective of Technology Readiness Index

Kuan-Yu Lin[1], Chun-Yu Chao[2], Xiang-Ting Zhou[3], Jui-Liang Hsu[4], and Che-Jen Chuang[5,⋆]

[1] Beijing Sport University, Sports Coaching College,
Beijing 100084, People's Republic Of China
kuanyu0706@qq.com
[2] General Education Center, National Kaohsiung University of Hospitality and Tourism, Taiwan
nike@staff.nkuht.edu.tw
[3] Department of Healthcare Industry Technology Development and Management,
National Chin-Yi University of Technology, Taiwan
shehe1016328@gmail.com
[4] Department of Leisure and Sports Management, Cheng Shiu University, Taiwan
joe50503388@yahoo.com.tw
[5] Department of Tourism & Leisure Management, Vanung University, Taiwan
chuangchejen@gmail.com

Abstract. This study explored the usage intention of the reservation system of Taipei Sports Center from the perspective of Technology Readiness Index. The research subjects were the consumers of the reservation system of Taipei City Sports Center. In the questionnaire distribution part of the study, a total of 350 questionnaires were sent out, and 350 questionnaires were returned. After the invalid questionnaires were deducted, there were 322 valid questionnaires, and the valid questionnaire recovery rate reached 92%. Statistical software SPSS 25.0 and AMOS 25.0 were then used for analysis. The research results are as follows: Hypothesis 1 Optimism had a significant positive influence on usage intention; Hypothesis 2 Innovation had a significant positive influence on usage intention; Hypothesis 3 Inadaptability had a significant negative influence on usage intention; Hypothesis 4 Insecurity had a significant negative influence on usage intention.

Keywords: Reservation System, Technology Readiness Index, Usage Intention.

1. Introduction

According to the Department of Accounting and Statistics of the Executive Yuan in 2023 [1], the annual total salary of the people of Taiwan has been growing year by year, from NTD 470,000 in 2007 to NTD 518,000 in 2022.The Sports Administration, MOE, in order to enhance the national sports culture and improve the lifestyle of the people, has formulated the Reference Guidelines for National Sports Center Planning since 2011 to provide a legal basis for county and municipal governments across Taiwan to effectively subsidize the county and municipal governments to build national sports centers with a

⋆ Corresponding author

wide range of sports. However, the traditional way of making reservations at sports centers often has such problems as low efficiency and lack of information transparency, so it cannot meet the needs of modern consumers for convenience and efficiency. Although the reservation system of Taipei Sports Center has been built with an online reservation function, the operation interface of the system is still complicated, which makes it difficult for some consumers, especially the elderly, to familiarize themselves with the use of the system. In addition, the lack of security of the system makes consumers worry about the protection of their personal information. All these problems seriously affect the willingness of consumers to use the reservation system. Therefore, this study focused on how to improve the reservation system to increase consumers' usage intention of the system. so that recreational sports can become a part of people's lives [6].

Similarly, it can be inferred that there has been a drastic change in people's lifestyles, with people increasingly relying on the convenience in life brought by smartphones and apps [34]. During the COVID-19 pandemic, the Taiwanese government severely restricted social distances and places of access, resulting in a dramatic change in people's lifestyles in terms of daily diet, leisure and personal health care, which accelerated the creation of various digitalized products, and an increase in the number of people who pay attention to their own physical and mental well-being [23]. [30] found that the a priori basis for individual cognitive judgments is the use of technology readiness to analyze the propensity of individual traits to use new technologies. Since different consumer characteristics may have an impact on the perception of new technologies, scholars have argued that whether a consumer has adequate technology readiness (TR) will influence the consumer's acceptance of using new technologies. [26] defined technology readiness as "people's habit of accepting and using new technologies to accomplish goals in daily life or work". The Technology Readiness (TR) theory proposed by Parasuraman refers to a psychological state that describes a user's attempts to use a new technology to achieve goals in daily life, learning or work performance. Technology readiness measures the degree to which an individual is receptive to technological information rather than the ability to accept it [8]. While some people have positive feelings of convenience, efficiency and joy in using new technology products or services, others have negative feelings in terms of safety, elimination risk and control, and they even fear technology. [22] pointed out that the use of new technology products or services can result in a psychological conflict between positive and negative feelings of control/chaos, freedom/slavery, novelty/cliché, ability/inability, efficacy/inefficacy, fulfillment of a need/creation of a need, assimilation/segregation, and fascination/unfascination, and their corresponding enablers and inhibitors. Positive enablers refer to consumers' beliefs that the use of new technology products or services can enhance the control, flexibility and efficiency of their daily lives, while negative inhibitors are the negative perceptions of consumers' inability to manage and control the use of new technology products or services [15]. [28] investigated the correlation between consumers' technology readiness in using smartphones and their preference for using business newsletters, which generally refers to the sending of caring, advertising or promotional messages to existing customers to attract new customers. [7] argued in his study that the advantages of efficiency, convenience and usefulness of using a new technology will positively affect consumers' perceived usefulness, which will make consumers have a more positive attitude towards the new technology. [38] investigated the correlation between technology readiness and job satisfaction in a study with the research

subjects of insurance salespersons and customer service workers, and the results showed that optimism and innovativeness in technology readiness were positively related to job satisfaction; on the other hand, inadaptability and security concerns were negatively related to job satisfaction. [37] also concluded in their study that technology readiness had a positive effect on online service quality and online shopping behavior. [27] pointed out that, for consumers who are unable to adapt to new technologies, the innovative technological environment is not very helpful to life.

Technology interacts with and develops along with users' life patterns, and adjusts to changes in the social environment and in response to users' needs, dramatically affecting people's habits of use. Web applications on the Internet have become the indicators of popular applications [13]. In the norm of mutual constraints between providers and consumers coordinated through the reservation-based service system, not only the provider can provide the reservation holders with the protection of their rights and benefits, but also the consumers must follow the agreed time to be present to receive the service [10]. [19] pointed out that the main source of business profits is reflected in the consumer services, consumers can make reservation for services through the Internet system, and they can know the status of the reservation, the reservation time and the number of reservations in real time. In other words, people's daily life has been inseparable from the Internet, and it has become a mainstream that companies must serve consumers through the Internet. With the establishment and help of a web-based reservation system, people can go online to make reservations at any time, regardless of time and place, and know the status quo of the reservation time and the number of people who have made reservations at the same time [14]. The user-friendly interface design can save the user's operation time and improve the user-friendliness, which is a way of great development potential to make reservations in the current environment of Internet penetration [17]. In the past, many scholars discussed the theories related to the usage intention of technology products. [16] found in their research on reservation systems that consumers' use of reservation systems not only reduced on-site waiting time, but also facilitated the exchange of information with operators [11]. [3] argued that the design of the parking reservation system consisted of an Android mobile phone reservation subsystem, a parking management subsystem and a parking ticket recognition subsystem, which not only supported real-time searching of neighboring parking spaces, but also provided parking space reservation services. [31] indicated that the parking space search system can provide drivers with information about the usage status of nearby parking spaces, which greatly reduces the time and money costs compared to the traditional random parking space search method, and is beneficial to the maintenance of traffic order around the parking spaces. [12] designed a school bus reservation system based on the study of the existing school bus schedule and conducted a field study. The results proved that it could save students' transportation costs and improve the safety and convenience of students' rides. [33] showed that an online reservation system can provide better service by reducing operating costs and providing more diversified channels for consumers. However, different users have different usage intentions of the reservation system over time, and the reservation system of Taipei City Sports Center in this study will also change according to consumers' usage patterns and acceptance of technological information. Existing studies have focused on the analysis of a single dimension and neglected the interactions between different dimensions [26]; [12]. Moreover, existing studies rarely focus on the specific area of the reservation systems of sports

centers and lack an in-depth discussion of practical application scenarios [31]. Therefore, this study further explored the effects of optimism, inadaptability, innovation and insecurity on usage intention from the perspective of technology readiness, and made specific suggestions to improve the innovation, usability and security of the system to increase users' usage intention and to improve the overall quality of service. These findings will not only enrich the application of technology readiness theory in specific domains, but also provide new perspectives for the optimization of the reservation system of sports centers.

2. Research Methods

2.1. Research Structure and Hypotheses

1. Research Structure
   Based on the technology readiness theory [26], this paper developed the research structure including the four variables of optimism, innovation, inadaptability and insecurity, as well as their impact on usage intention. Each variable was measured through multiple observation indicators, and the path arrows indicated the causal relationship between the variables, as shown in Figure 1.

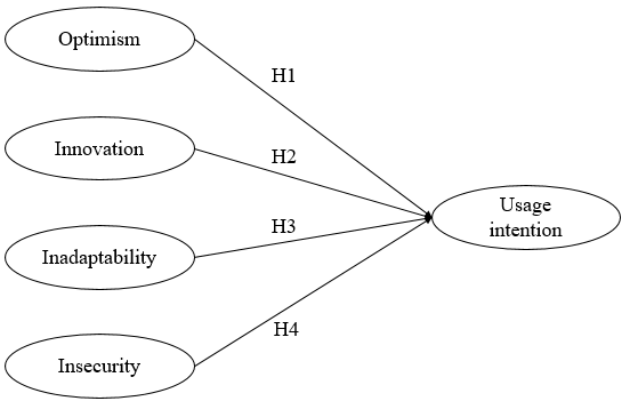


Fig. 1. Research Structure

2. Research Hypotheses
   Based on the previous literature review as well as the purpose and research structure of this study, the following hypotheses were proposed:
   H1: "Optimism" has a significant positive influence on "usage intention".
   H2: "Innovation" has a significant positive influence on "usage intention".
   H3: "Inadaptability" has a significant negative influence on "usage intention".
   H4: "Insecurity" has a significant negative influence on "usage intention".

2.2. Research Subjects and Research Tools

1. Research Subjects
   This study took the consumers using the reservation system of Taipei City Sports Center as the research subjects of this study. The intentional sampling method was adopted to sample a total of 350 consumers of the reservation system of Taipei City Sports Center from December 05, 2023 to December 30, 2023 outside Taipei City Sports Center. During the questionnaire collection process, a total of 350 questionnaires were collected and 322 valid questionnaires were obtained after the invalid questionnaire were deducted, with a valid questionnaire recovery rate of 92
2. Research Tools
   The content of the questionnaire was compiled based on the literature review, including the four scales for optimism, innovation, inadaptability and insecurity. The questionnaire content included basic data, 4 questions in optimism scale, 5 questions in innovation scale, 5 questions in inadaptability scale, 5 questions in insecurity scale, and 3 questions in usage intention scale. The questionnaire was modified from the studies of [26], [25], [4], [36], [35], and [29], respectively, and was measured on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The data were analyzed using SPSS 25.0 for descriptive statistics to understand the basic characteristics of the samples. AMOS 25.0 was used for structural equation modeling to validate the hypotheses and assess the model fit. The AMOS software is effective in building models to validate hypotheses, confirming relationships between observations and latent variables, and it also has the function of handling complex relationships between multiple variables.

3. Research Results

3.1. Descriptive Statistical Analysis

After the samples of the study on the usage intention of the reservation system of Taipei City Sports Center from the perspective of technology readiness were collected, the invalid questionnaires were deleted. Then descriptive statistics were used to analyze and understand the background information of the valid samples, and the analysis of the data of this study was summarized as shown in Table 1.

3.2. Confirmatory Factor Analysis

1. Validation of Convergent Validity
   This study examined the convergent validity of the research dimensions of optimism, innovation, inadaptability, insecurity and usage intention. The results of the analysis showed that the factor loading values of all the dimensions ranged from 0.53 to 0.90, the composite reliability values ranged from 0.82 to 0.92, and the values of average variance extracted ranged from 0.51 to 0.69, indicating that this study had convergent validity [21]. The details are shown in Tables 2 to 6.

Table 1. Summary of Descriptive Statistics of Background Data of Research Samples

Background variables	Classification criteria	Sample size	Percentage %	Cumulative percentage %
Gender	Male	156	48.4%	48.4%
	Female	166	51.6%	100%
Age	20 years old or below	27	8.4%	8.4%
	21-30 years old	110	34.2%	42.5%
	31-40 years old	95	29.5%	72.0%
	41-50 years old	47	14.6%	86.6%
	51-60 years old	31	9.6%	96.3%
	61 years old or above	12	3.7%	100%
Educational level	High (vocational) school or below	114	35.4%	35.4%
	University (junior college)	166	51.6%	87%
	Graduate school (or above)	42	13%	100%
Marital status	unmarried	148	46%	46%
	married	145	45%	91%
	others	29	9.0%	100%

Table 2. Summary of Confirmatory Factory Analysis of Optimism (Source: Compiled by this study)

Estimates of model parameters							Convergent validity		
Latent variable	Observational variable	Unstandardized factor loading	Standard deviation	C.R	p	Factor loading	SMC	Composite reliability	Average variance extracted
Optimism	Optimism 1	1.00				0.68	0.46	0.83	0.51
	Optimism 2	1.14	0.10	12.08	***	0.79	0.62		
	Optimism 3	1.19	0.10	11.50	***	0.79	0.62		
	Optimism 4	1.05	0.10	11.00	***	0.74	0.54		
	Optimism 5	0.76	0.09	8.60	***	0.53	0.28		

Table 3. Summary of Confirmatory Factory Analysis of Innovation (Source: Compiled by this study)

Estimates of model parameters							Convergent validity		
Latent variable	Observational variable	Unstandardized factor loading	Standard deviation	C.R	p	Factor loading	SMC	Composite reliability	Average variance extracted
Innovation	Innovation 1	1.00				0.73	0.53	0.89	0.62
	Innovation 2	1.02	0.08	13.42	***	0.76	0.57		
	Innovation 3	1.14	0.08	14.12	***	0.85	0.72		
	Innovation 4	1.18	0.08	14.08	***	0.83	0.68		
	Innovation 5	0.99	0.08	12.76	***	0.76	0.57		

Table 4. Summary of Confirmatory Factory Analysis of Inadaptability (Source: Compiled by this study

Estimates of model parameters							Convergent validity		
Latent variable	Observational variable	Unstandardized factor loading	Standard deviation	C.R	p	Factor loading	SMC	Composite reliability	Average variance extracted
Inadaptability	Inadaptability 1	1.00				0.50	0.44	0.83	0.51
	Inadaptability 2	0.71	0.07	9.14	***	0.50	0.34		
	Inadaptability 3	1.01	0.08	11.69	***	0.64	0.65		
	Inadaptability 4	0.98	0.08	11.40	***	0.78	0.59		
	Inadaptability 5	0.88	0.08	10.67	***	0.58	0.50		

Table 5. Summary of Confirmatory Factory Analysis of Insecurity (Source: Compiled by this study)

Estimates of model parameters							Convergent validity		
Latent variable	Observational variable	Unstandardized factor loading	Standard deviation	C.R	p	Factor loading	SMC	Composite reliability	Average variance extracted
Insecurity	Insecurity 1	1.00				0.82	0.67	0.92	0.69
	Insecurity 2	1.05	0.05	18.05	***	0.84	0.70		
	Insecurity 3	1.09	0.05	19.41	***	0.90	0.81		
	Insecurity 4	1.05	0.05	19.05	***	0.88	0.77		
	Insecurity 5	0.95	0.06	14.44	***	0.73	0.53		

3.3. Fit Analysis

In the structural model analysis of this study, the indices of chi-square value, chi-square value/degree of freedom (X^2/df), GFI, AGFI, RMSEA, CFI, etc., were used as the fit

Table 6. Summary of Confirmatory Factory Analysis of Usage Intention (Source: Compiled by this study)

Estimates of model parameters							Convergent validity		
Latent variable	Observational variable	Unstandardized factor loading	Standard deviation	C.R	p	Factor loading	SMC	Composite reliability	Average variance extracted
Usage intention	Usage intention 1	1.00				0.84	0.70	0.82	0.61
	Usage intention 2	0.91	0.07	12.52	***	0.77	0.59		
	Usage intention 3	0.82	0.07	12.18	***	0.73	0.53		

indices for the overall model analysis of this study with reference to the research opinions of [2] and [32]. In this study, the corrected ratio of X^2/df was 2.17, which was consistent with the statement of [32] that the smaller the ratio of X^2 to the degree of freedom, the better; [21] stated that the closer the values of GFI and AGFI are to 1, the better, and the corrected ratios of GFI and AGFI in this study were 0.89 and 0.86, respectively. [24] indicated that the best RMSEA value is less than 0.08, and the corrected RMSEA value of this study was 0.06; the best standardized value of CFI is more than 0.90, and the corrected CFI of this study was 0.93. All these showed that the overall model fit of this study was good and the indices met the standard, as shown in Table 7.

Table 7. Overall Model Fit Analysis (Source: Compiled by this study)

Fit indices	Acceptable range	This study model	Model fit judgement
X^2 (Chi-square)	The smaller, the better	477.301	
Ratio of X^2 to degree of freedom	<3	2.17	Good fit
GFI	$>.80$	0.89	Good fit
AGFI	$>.80$	0.86	Good fit
RMSEA	$<.10$	0.06	Good fit
CFI	$>.90$	0.93	Good fit

The empirical results of this study are shown in Table 8. For H1, the path value of optimism on usage intention was 0.69 (p=0.00), H1 was established, that is, the optimism of the users of the reservation system of Taipei City National Sports Center had a significant positive influence on the usage intention of the reservation system. This result of this study is in line with the results of the studies by [25] and [18]. The possible reason is that the consumers of Taipei City Sports Center can make reservations anytime and anywhere through the APP, so that the situation that the facilities and equipment are fully occupied when they arrive at the site can be avoided, thus increasing the usage intention the reserva-

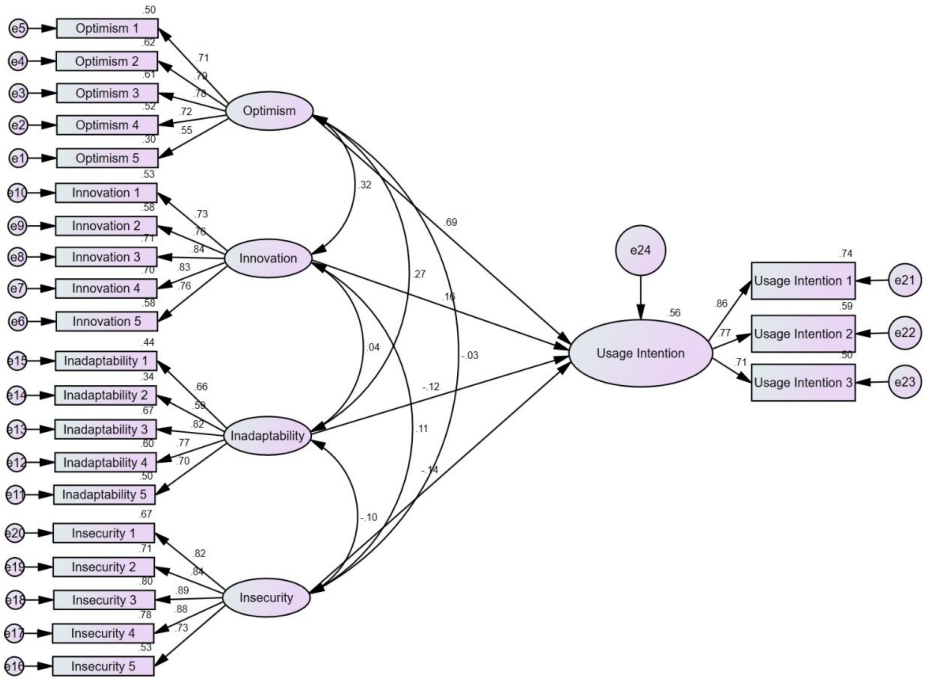


Fig. 2. Study Model Verification Results

Table 8. Empirical Results of Research Hypotheses (* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$)

Hypotheses	Path relationship	Path value	Hypothesis supported or not
1	"Optimism" has a significant positive influence on "usage intention".	0.69***	Supported
2	"Innovation" has a significant positive influence on "usage intention".	0.16***	Supported
3	"Inadaptability" has a significant negative influence on "usage intention".	-0.12**	Supported
4	"Insecurity" has a significant negative influence on "usage intention".	-0.14***	Supported

tion system of Taipei City Sports Center. In Hypothesis 2, the path value of innovation on usage intention was 0.16 (p=0.00), so H2 was established. That is, innovation of the users of the reservation system of Taipei City Sports Center had a significant positive influence on the usage intention of the reservation system. This finding is consistent with the findings of [18] and [20], which may be attributed to the fact that the consumers of Taipei City Sports Center recognized the innovative technology services and conveniences provided

by the reservation system, thus increasing the usage intention of the reservation system of Taipei City Sports Center.

In H3, the path value of inadaptability on usage intention was -0.12 (p=0.02), H3 was established, that is, the inadaptability of the users of the reservation system of Taipei City Sports Center had a significant negative influence on the usage intention of the reservation system. The result is in line with the results of [18] and [20]. There is a possible reason that the users of the reservation system of Taipei City Sports Center learned to adapt to the operation of the reservation system of Taipei City Sports Center and the correctness of the information, which affected the usage intention of the reservation system of Taipei City Sports Center. In H4, the path value of insecurity on the usage intention of the reservation system was - 0.14 (p=0.00), H4 was established, i.e., insecurity of the users of the reservation system of Taipei City Sports Center had a significant negative influence on the usage intention of the reservation system, which is in line with the results of [5] and [9]. It is possible that consumers of Taipei City Sports Center may be concerned about their personal security and privacy when registering for the reservation system, which may affect the usage intention of the reservation system of Taipei City Sports Center.

4. Conclusion and Suggestions

4.1. Conclusion

This study examined the usage intention of the reservation system of Taipei City Sports Center from the perspective of technology readiness, and the results showed that: (1) The path value of optimism on usage intention was 0.69 (p=0.00), so H1 was established. The optimism of the consumers in using the reservation platform of Taipei City Sports Center as examined in this study had a significant positive influence on the usage intention of the reservation system of Taipei City Sports Center. (2) The path value of innovation on usage intention was 0.16(p=0.00), so H2 was established. The innovation of the consumers in using the reservation platform of Taipei City Sports Center as examined in this study had a significant positive influence on the usage intention of the reservation system of Taipei City Sports Center. (3) The path value of inadaptability on usage intention was -0.12(p=0.02), so H3 was established. The inadaptability of the consumers in using the reservation system of Taipei City Sports Center as examined in this study had a significant negative influence on the usage intention. (4) The path value of insecurity on usage intention was -0.14(p=0.00), so H4 was established. The insecurity of the consumers in using the reservation system of Taipei City Sports Center as examined in this study had a significant negative influence on the usage intention. From the theoretical point of view, this study enriched the application of technology readiness theory in a specific domain, and provided a new perspective for understanding the acceptance of technology services by users. From a practical point of view, this study provided specific recommendations for sports center managers to optimize the design of the reservation system, such as improving the system's ease of use, and enhancing data security and privacy protection.

4.2. Suggestions

In recent years, as people pay more attention to health and sports, the demand for the use of sports centers is increasing. However, there are now some problems with the reservation system of Taipei City Sports Center, such as complicated operation, insufficient data

security, and lack of innovation. These problems seriously affect the users' usage intention. In order to address these challenges, this study analyzed the multidimensional user psychological factors such as optimism, innovation, inadaptability and insecurity from the perspective of technology readiness, and proposed novel and specific suggestions for improvement to enhance the service quality and user satisfaction of the reservation system of Taipei City Sports Center.

The results of this study showed that optimism had a significant positive effect on the usage intention of the reservation system of Taipei City Sports Center. Therefore, in order to strengthen the driving force of the positive feelings of the users of the sports center in terms of technology readiness, the sports center industry can use the new technological marketing techniques to market such intelligent technology services as the online venue flow inquiry system and the venue reservation system to the consumers of the sports center, and explain what new benefits and conveniences they can bring to increase the level of optimism of the sports center users towards the intelligent technology services of the sports center. In addition, innovation had a significant positive impact on the usage intention of the reservation system of Taipei City Sports Center. The system manager of the sports center should be able to quickly understand and respond to the market demand, so as to facilitate the introduction of safer, more complete and easy-to-understand technology services and products for the sports center, such as the "AI Intelligent Badminton Court" and "Swimming Pool Drowning Detection System" for consumers to use. This can reduce the powerlessness and fear of sports center consumers to the new technology, make they believe that technology sports venues provide online, intelligent, interactive, immersive and other technology services can bring flexibility, efficiency, safety and convenience to meet their needs, so as to improve the quality of technology services in sports centers.

The results of this study showed that insecurity had a significant negative effect on the usage intention of the reservation system of Taipei City Sports Center. Consumers may have slightly different views on data security and privacy protection of technology services in technological sports centers due to the age difference. Younger consumers may be more skillful in the use of technology, but their concern and alertness to security issues may be relatively low, and they may easily neglect security because of convenience; while older consumers may pay more attention to information security and user interface because they may have suffered from security problems or been cheated in their past life and work experiences, which makes them have a deeper understanding of the security risks of technology services. Therefore, it is suggested that sports center operators should effectively enhance the capacity of their technology service systems to meet security and privacy protection challenges and strengthen the user interface. In this way, they will be able to protect the security and usage of the reservation system and other technology services of the sports center, and maintain the trust and reputation of consumers and the sports center at the same time. In addition, inadaptability had a significant negative impact on the usage intention of the reservation system of Taipei City Sports Center. It is suggested that sports centers should strengthen the instructions for the use of the reservation system in the system. For example, more detailed operating instructions, digital teaching materials or Frequently Asked Questions (FAQ) can be included in the system to help consumers quickly familiarize themselves with the system functions. And the service staff also need to improve the communication and problem-solving explanation and operation

skills, which will enable them to assist consumers more effectively in solving problems encountered in using the new system, so that the consumers of sports centers can quickly adapt to the new technological service system, and the overall quality of service can be improved.

In this study, the confirmatory factor analysis was conducted to examine the convergent validity of the research dimensions of optimism, innovation, inadaptability, insecurity and usage intention. Th correctness and practicability of the analytical model was validated by the structural modeling analysis with such indices as chi-square values, chi-square values/degrees of freedom (X^2/df), GFI, AGFI, RMSEA and CFI in the fit analysis.

References

1. Accounting, Statistics, E.Y.: Statistical results for median annual total salary and distribution of employees in industry and services, 2022. (2003), https://reurl.cc/A286xY
2. Bagozzi, R.P., Yi, Y.: On the evaluation of structural equation models. Journal of the academy of marketing science 16, 74–94 (1988)
3. Chang, C.K.: Design and development of a parking reservation system (2014)
4. Chang, T.H.: The relationship among use intention of credit card and brand image (2021)
5. Chen, S.Y.: An investigation of technology readiness, attitude and behavior intention toward application in community residents (2021)
6. Chou, Y. H., K.C.N.K.C.H.: A comparative study of the effectiveness of government recreation centers and privately owned sports facilities. Journal of Taiwan Society for Sport Management 12(4), 255–284 (Dec 2012)
7. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly pp. 319–340 (1989)
8. Erdoğmuş, N., Esen, M.: An investigation of the effects of technology readiness on technology acceptance in e-hrm. Procedia - Social and Behavioral Sciences 24, 487–495 (2011)
9. Fan, S.T.: An empirical study of technology readiness, attitude and behavior intention toward instant messenger in travel agency (2010)
10. Hsu, J.S.: Discussing the influences of factory service appointment system on customer satisfaction: Taking central automobile south taichung plant as an example (2002)
11. Hu, M., Xu, X., Li, X., Che, T.: Managing patients' no-show behaviour to improve the sustainability of hospital appointment systems: Exploring the conscious and unconscious determinants of no-show behaviour. Journal of Cleaner Production 269, 122318 (2020)
12. Huang, H.C.: Research and implementation of a school bus reservation system (2023)
13. Huang, Y.W., Tsai, C.H., Lin, T.P., Huang, S.K., Lee, D., Kuo, S.Y.: A testing framework for web application security assessment. Computer Networks 48(5), 739–761 (2005)
14. Joslyn, J.S.: Healthcare e-commerce: connecting with patients. Journal of Healthcare Information Management 15(1), 73–84 (2001)
15. Ke, C.Y.: A case study of mobile device services for discussing the acceptance of consumers behavior using technical readiness. (2012)
16. Küçük, A., Demirci, M., Kerman, G., Özsoy, V.S.: Evaluating of hospital appointment systems in turkey: Challenges and opportunities. Health Policy and Technology 10(1), 69–74 (2021)
17. Kwak, J.K.: Analysis of the waiting time in clinic registration of patients with appointments and random walk-ins. International Journal of Environmental Research and Public Health 20(3), 2635 (2023)
18. Liu, P.T.: Research on the technology readiness and continuous use intention of elementary school teachers: Taking "classes have internet, students use tablets" as an example (2024)

19. Liu, Y.T.: Design and implementation of adaptive streaming media player based on http live streaming protocol (2012)
20. Lu, Z.F.: Investigating consumers' attitudes and usage intentions towards ar by integrating technological readiness and theory of planned behavior (2024)
21. Mardia, K.V., Kent, J.T., Taylor, C.C.: Multivariate analysis. John Wiley & Sons (2024)
22. Mick, D.G., Fournier, S.: Paradoxes of technology: Consumer cognizance, emotions, and coping strategies. Journal of Consumer research 25(2), 123–143 (1998)
23. Modgil, S., Dwivedi, Y.K., Rana, N.P., Gupta, S., Kamble, S.: Has covid-19 accelerated opportunities for digital entrepreneurship? an indian perspective. Technological Forecasting and Social Change 175, 121415 (2022)
24. Mw, B.: Alternative ways of assessing model fit. Testing structural equation models (1993)
25. Pan, Y.S.: A study on attitude and willingness to use health passbook app based on technology readiness-taking military, public and educational personnel as an example (2024)
26. Parasuraman, A.: Technology readiness index (tri): A multiple-item scale to measure readiness to embrace new technologies. Journal of Service Research 2(4), 307–320 (2000)
27. Park, S.Y., Zhu, K.: Advances in sno2 for efficient and stable n–i–p perovskite solar cells. Advanced materials 34(27), 2110438 (2022)
28. Sophonthummapharn, K., Tesar, G.: Technology readiness and propensity of cell phone users to subscribe to commercial messaging services. Marketing Management Journal 17(2), 81–95 (2007)
29. Sun, S.C.: Integrating technology readiness and post-acceptance model: A case study of mobile service (2010)
30. Walczuch, R., Lemmink, J., Streukens, S.: The effect of service employees' technology readiness on technology acceptance. Information & management 44(2), 206–215 (2007)
31. Wang, C.K.: Smart parking space monitoring and reservation system. department of electrical engineering (2020)
32. Whittaker, T.A., Schumacker, R.E.: A beginner's guide to structural equation modeling. Routledge (2022)
33. Wu, H., Lu, N.: Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. International Journal of Medical Informatics 107, 107–119 (2017)
34. Wu, H.C., Huang, W.Y.: A study on using the technology acceptance model to explore the behavioral intentions of consumers using app to watch sports events. Physical Education Journal 53(1), 75–94 (Mar 2020)
35. Wu, T.Y.: Exploring the influential factors of user satisfaction with retailer app from the perspective of information system success model and technology readiness index - a study of convenience store app (2023)
36. Yang, W.H.: The influence of technology readiness and green knowledge on purchase intention in esg environment-a case study of electric vehicle consumers in southern taiwan (2023)
37. Zeithaml, V.A., Parasuraman, A., Malhotra, A.: Service quality delivery through web sites: a critical review of extant knowledge. Journal of the academy of marketing science 30(4), 362–375 (2002)
38. Zhan, W.S.: A study on the impact of boundary spanners' technology readiness on their job satisfactions in the service industry (2002)

Kuan-Yu Lin is a PhD student of Sports Coaching College at Beijing Sport University. His research field focuses on Sports training and management, Analysis of Taekwondo techniques and tactics, and Outdoor education exploration. His academic papers have been published in journals like Journal of Taekwondo Sports, Journal of Physical Education Fu Jen Catholic University, etc.

Chun-Yu Chao is a Assistant Professor of the General Education Center, National Kaohsiung University of Hospitality and Tourism. Research expertise: Sports, Photography, Coffee, Calligraphy.

Xiang-Ting Zhou is a master at the School of Department of Healthcare Industry Technology Development and Management, National Chin-Yi University of Technology Research expertise: Healthcare Industry Technology Development and Management.

Jui-Liang Hsu is a Associate Professor of the Department of Leisure and Sports Managemen at Cheng shiu University. He obtained his Ph.D. from the Department of Bio-Industry Technology, Da Yeh University. His research field focuses on Leisure and Sports Managemen. His academic papers have been published in journals like International Conference on Economics, and Management of Business, Innovation and Technology, IOP Conference Series: Earth and Environmental Science, the International Journal of Water, Life Science Journal. To date, he has obtained five utility model patents from the Republic of China.

Che-Jen Chuang is an Associate Professor in the Department of Tourism and Leisure Management at Vanung University. He has previously served as the Dean of International & Cross-Strait Affairs and the Dean of the Chinese Language Center at Vanung University. He obtained his Ph.D. from Maejo University, School of Tourism Development. His research expertise includes tourism planning strategic management, leisure behavior research, and research methodology in tourism.

Received: December 18, 2024; Accepted: March 03, 2025.

Computer Science and Information Systems 22(3):1361–1378 https://doi.org/10.2298/CSIS241218049L

Applying MSEM to Analyze People's Cognitive Behavior towards Virtual Reality Sport Experience

Yan-Hui Li[1], Cheng-Sheng Lin[2], Che-Jen Chuang[3], Jui-Liang Hsu[4], and Yu-Jui Li[5,*]

[1] Doctoral Program for Transnational Sport Management and Innovation, College of Management, National Taiwan Sport University, Taiwan
1111704@ntsu.edu.tw
[2] Department of Agricultural Technology, National Formosa University, Taiwan.
sheng8876@nfu.edu.tw
[3] Department of Tourism & Leisure Management, Vanung University, Taiwan
Chuangchejen@gmail.com
[4] Department of Leisure and Sports Management, Cheng Shiu University, Taiwan
joe50503388@yahoo.com.tw
[5] Department of Recreation and Sports Management, University of Taipei, Taiwan
li542058@utaipei.edu.tw

Abstract. Since the traditional Structural Equation Modeling is unable to correctly grasp the preferences and behaviors of consumers at multiple levels, this study combined the virtual reality experience model to analyze the motivations and cognitive behaviors of the public towards sports and fitness. The factor analysis was adopted to measure the latent variables of the virtual reality sports experience function, and then Multilevel Structural Equation Modeling was conducted based on the latent variables to find out the optimal model combination and the best path relationship between people's experience situations and latent variables, to truly grasp the motivations and goals of people's sports experience. The empirical results showed that the main factors of the latent variables of people's satisfaction and loyalty of the group level were different from those of the individual level, which indicated that there was a difference between the group and individual sports preferences in virtual reality sports experience, and it indirectly affected the consumption intention of the group and the individual for virtual reality sports experience.

Keywords: Multilevel Structural Equation Modeling (MSEM), Virtual Reality (VR), consumption intention.

1. Introduction

With the advancement of information technology, emerging industries have the momentum for growth and potential huge production value, and the sports industry is also growing rapidly with the widespread application of digital technology. According to the data of [14], the Global Sports Technology Market Size was valued at USD 20.13 billion in 2022, and the Worldwide Sports Technology Market is expected to reach USD 76.39 billion by 2032, demonstrating the development potential of the sports industry. On the other hand, the outbreak of the COVID-19 pandemic in 2020 has changed people's consumption and exercise habits, indirectly giving opportunities to combine technology and sports fields to create more diverse sports products, services and business models. [10] pointed out

that the pandemic has promoted the rise of virtual events around the world, and cross-border and cross-domain sports exchanges are held online. With pandemic prevention considerations, people have reduced their going out and socializing, creating a global trend of home gym, and giving rise to the vigorous development of sports technology. In the post-pandemic era, the sports industry has integrated technologies such as 5G, sensing devices, Artificial Intelligence (AI), and Virtual reality (VR) to provide consumers with a more diverse experience. In other words, since the COVID-19 pandemic, whether it is for leisure sports, consumption, education, learning, or emotional communication, people's huge demand for digital remote interaction has been triggered. The breakthrough development of 5G communications, AI, imaging and other technologies also provides a new interactive experience for sports and leisure, and pushed mankind into the stage of digital transformation, allowing the concepts of virtual and reality to be realized in human life. The development of VR technology and equipment is an important key to promote the mixing of reality and VR. Through the aid of innovative equipment, we can overcome the limitations of human mobility, enhance the sense of realism of long-distance interaction and the benefits of sports and leisure, reduce the cost and risk of outdoor sports, increase people's willingness to play sports, which is an important way to alleviate stress by promoting the physical and mental health of the people. [15] indicated that VR technology can not only simulate the natural environment and promote the sense of participation and enjoyment of sports, but also meet the needs of different groups through customized virtual environments, thus increasing the degree of participation and time spent in sports. [30] hold a similar view, arguing that different applications of VR can create new value in the sports industry, including new marketing methods and business models, and sports consumers can obtain more innovative and diversified interactive experiences through VR [30]. The current applications of VR in the sports industry include sports event broadcasting, sponsor relations and fan management, sports retail, sports museums, sports training and analysis, sports participation, and sports facilities [31].

Since each person has different athletic abilities, the goals they aim to achieve through exercise also vary. VR can also allow users to customize their exercise settings and scenarios according to their preferences and goals. For instance, some VR apps can let users choose different levels of difficulty. [2] indicated thatVR technology not only simulates natural environments, enhancing the sense of engagement and enjoyment in exercises, but also meets the needs of different groups through customized virtual environments, thereby increasing exercise participation and duration. VR not only brings exercise benefits to regular users but also makes it easier for individuals undergoing rehabilitation to use and achieve the desired rehabilitation effects.In the virtual world, real-life scenarios can be created for physical therapy exercises such as stretching, strength training, gait training, coordination, balance, and functional practice through task-based games. These tasks can be customized based on questionnaires or personal preferences and include activities like moving a cup to a designated location, picking fruit from trees, and kicking a ball [16]; [40].

The related products of the combination of VR with sports equipment have also been launched. The relevant research on consumer behavior of VR sports technology also provides reference information for the continuous expansion of the sports industry. As can be seen from the relevant literature, the issue of consumer behavior towards VR sports technology is mostly explored from the perspective of technology acceptance. For exam-

ple, [26] utilized the Technology Acceptance Model (TAM) to explain user attitudes and behavioral intentions in a VR surfing experience. [8] explored VR acceptance through the TAM. [43] proposed an extended version of the TAM that addresses some aspects of VR. Nonetheless, there are still other aspects of consumer behaviors that can be explored in exploring consumer behavior towards VR. For example, consumers' use of VR sports technology not only includes individual participation, but also group participation. This also means that there are different motivations and purposes for individual and group participations, so it is not suitable to analyze the motivations and purposes of people's multilevel leisure sports by traditional SEM. [30] also stated that empirical research that focused on consumer behavior at the intersection of sport content and media technology acceptance is limited. On the other hand, a general framework incorporating various consumers' contributions as main antecedents of customer satisfaction and loyalty, is lacking. Therefore, this paper examined the influence of individual and group consumers on consumption intention in terms of loyalty and satisfaction of VR consumers.

VR sports experience can change people's movement patterns. Elevating fitness to a gaming role will bring a whole new definition to exercise. The effort you put into VR sports will reap immediate in-game rewards and long-term health benefits.

2. Literature Review

Communication with consumers has been the focus of many sports service-related industries [1]; [9]. Generally speaking, consumers' evaluation of the individual's physical and mental pleasure and actual experience during the process of using facilities and enjoying services can be regarded as consumer satisfaction. [17] pointed out that consumer satisfaction is a perception of pleasure or discomfort that an individual obtains after comparing the perceived benefits of a product or service with his or her own expectations. This concept can be found in [41], which also believes that consumers' evaluation of their satisfaction with a product or service is based on a comparison of their prior expectations and their subsequent performance obtained after the actual experience. When consumers believe that the subsequent performance evaluation of the product or service is higher than the original prior expectations, there will be a result of satisfaction, and otherwise, dissatisfaction. It can be seen from the above definition that consumers or participants' comparison of their previous expectations or past experiences with their feelings after actual experience is influenced by various personal factors and environmental factors. Therefore, consumers' actual feelings and personal experiences of participating in activities are particularly important. VR devices for different sports provide consumers with various personal experiences. The quality of these experiences is worthy of further exploration on the satisfaction with the VR devices or services for the sport.

How to promote loyalty has always been the goal of marketers for business organizations, as consumers' higher consumption intention on a product or service means a higher level of loyalty. [34] believe that customer loyalty consists of the commitment of customers to a brand or company to maintain a long-lasting relationship, which will be manifested through attitudes and behaviors. Attitudes include the intention to repurchase or purchase other products of the company, the willingness to recommend the company to others, and to reduce the loss of competition; while behaviors include repeat purchases, purchasing other products of the company, and recommending the company

to others. [21] suggested that there are two types of factors that contribute to consumer loyalty, including customer loyalty generated when the services and products offered by a company are superior to those of other brands, and the likelihood that a customer will repurchase the brand in the future or favor another brand. Thus, it is worthwhile to explore whether the hands-on experience and use experience of VR sports devices are good enough for consumers to choose the products and services once again. In addition, [45] proposed the "Loyalty Triangle" model as a framework for building customer loyalty, which consists of three dimensions: (1) service process: the process of service operation, which includes all activities that require the involvement of the customer and the service provider; (2) value creation: including value addition and acquisition; (3) database management/communication: creating a database of customer preferences, providing special services to customers, and utilizing customer contacts such as advertising letters and event notifications. Through these three dimensions, a company or organization can better focus on its consumers.

When weighing the likelihood of purchasing a product or service, consumers often make decisions on purchasing based on their own perceptions of the product or service. According to [44], consumer intention is a subjective tendency to a particular product formed on consumers' own experiences and the collection of external information about the product during the process of purchasing the product. [11] considered that the consumer's main reason before his or her making the decision on purchasing a product is that he or she has the need for the product, and the consumer decides to purchase the product to satisfy that need, while the evaluation criteria are based on personal experience and the external environment. Thus, consumer intention is a subjective decision. This is consistent with [3] Theory of Planned Behavior, which states that behavior intention occurs before the actual behavior, and behavior intention is an important basis for predicting the occurrence of the behavior. In other words, VR devices of different sports provide consumers with different hands-on experiences, and it is worthwhile to further analyze whether they generate positive consumption intentions, and what factors influence consumption intentions.

In terms of the relationship between variables, [13] explored the evolution of Marketing 4.0 and empirically examined its impact on customer satisfaction and purchase intention. The finding showed that the impact of customer satisfaction on purchase intention is highly significant. [33] on E-commerce also demonstrated that customer satisfaction positively affects purchase intention. On the other hand, related studies also indicated the impact of loyalty on consumption intention [5]; [12]; [18].

3. Research Methods

This study focused on analyzing people's experience and cognitive behaviors of sports and fitness in the context of VR sports experience. The online people's sports experience with VR was sampled and analyzed, the factor analysis was adopted to measure the latent variables of the VR sports experience function, and Multilevel Structural Equation Modeling (MSEM) based on the latent variables was used to find out the best model combination and the best path relationship between people's experience situation and the latent variables. The purpose was to grasp the motivations and goals of people's sports

behaviors, and to establish the optimal sports experience model as well as physical and mental health strategies based on the people's motivations and demand for leisure sports.

In this study, we used factor analysis to analyze people's cognitive behavior towards VR sports and fitness experience, divided the participants of VR sports experience into individuals and groups, and established three latent variables, namely satisfaction, loyalty, and consumption intention. At the same time, the optimal path relationship between the latent variables was established with MSEM. Multilevel data are often characterized by clustered data, which makes the measured sample data have special dependencies, resulting in violation of the assumption of sample independence and invalidating the statistical test. Normal regression analysis and variance analysis cannot deal with this problem, so multilevel analysis techniques must be adopted. Otherwise, the analysis data wouldl be confounded by the hierarchical relationships and lead to erroneous statistical conclusions.

Recent studies have utilized the SEM architecture to deal with multilevel data for Multilevel SEM (MSEM), such as [35]; [27]; [42]; [19]; [20]; [22]; [28]; [32]. Traditional MLM analysis and SEM are independent and unrelated statistical techniques, each with its own merits. If a database has both multilevel structure and latent variable estimation needs, it is necessary to combine these two analytical methods. Therefore, MSEM has been developed to solve the problems of multilevel data structure and latent variable estimation at the same time [6]; [7]; [37]; [38]; [23]; [25]; [29]. In this study, factor analysis was used to measure the latent variables of the VR sports experience function, and based on the latent variables, MSEM was conducted to find out the optimal combination of the models and the optimal path relationship between people's experience situation and the latent variables. The purpose was to understand the motivations of people's sports behaviors and consumption intentions, and to optimize the benefits of sports and VR technology.

The basic hypotheses of the MSEM of this study are as follows.

H1: People's satisfaction with VR sports and fitness experience has a positive effect on consumption intention.

H2: People's loyalty to VR sports and fitness experience has a positive effect on consumption intention.

H3: People's satisfaction with VR sport and fitness experience has a positive effect on loyalty.

Since people of VR sports experience includes both individuals and groups of people with different motivations and purposes, it is not suitable to analyze the motivations and purposes of people's multilevel leisure sports with traditional SEM. This study mainly applied MSEM to explore different groups of people's demand for sports functions in VR sports experience, and conducted the estimation and analysis of the potential contextual variables to grasp the demand of individuals and groups of people in VR leisure sports experience, and to improve the functional benefits of VR sports experience. The aim was to propose the best VR sports experience marketing strategy in an accurate and customized manner to achieve the goals of optimizing people's needs of physical and mental health, physical fitness, and recreation.

According to [37]; [39], in a multilevel data structure, observations at the group level are represented by averages at the individual level, so that the estimates of the intergroup variables contain information about the intragroup variables. In this case, the S PW matrix is the maximum likelihood estimator of the parent matrix ($\sum W$) But the S' B matrix is not the maximum likelihood estimator of the intergroup matrix ($\sum B$). Instead, it is the

linear integration weighted maximum likelihood estimator of the intergroup matrix ($\sum B$) and the intragroup parent matrix ($\sum W$), as shown in the following two equations.

$$S_{PW} = \sum W \tag{1}$$

$$S'_B = \sum W + C_g \sum B \tag{2}$$

In multilevel data, the observation levels of individual and group levels are different, and the number of people C_g in each group in Eq. (2), is used for weighting. When groups are equal in size, C_g is a fixed constant, and the model is called a balanced model; in an unbalanced model with unequal group sizes, C g is a variable. [37]; [39] suggested to ignore the effect of group size differences and replace the C g weight with a post hoc estimate of group size (c^*), which is very close to the average group size, to derive an intergroup observation matrix. The purpose is to simplify the model and achieve the convergence of the MSEM model. c^* is defined as follows.

$$C^* = \frac{N^2 - \sum_{g=1}^{G} C_g^2}{N\,(G-1)} \tag{3}$$

where N is the total number of samples and G is the number of groups. Eq. (3) is the Limited Information Maximum Likelihood (LIML) solution, which is called MUML solution (MUthén's ML) by [37]; [39] and pseudobalanced solution (PBL) by [36]. S'_B is very close to the estimate of $\sum B$ when the sample sizes for intergroup and intra groups are both large. But, when the sample size is small, the bias becomes more severe and the parameter estimates and standard deviations become incorrect. Therefore, the analysis of the MSEM model in this study followed the two-stage procedure suggested by [4], where the most appropriate measurement model is identified and then followed by structural model analysis. The intergroup and intragroup structural models with latent variables (η) are defined as follows.

$$\eta_B = \alpha_B + B_B \eta_B + \varepsilon_B \tag{4}$$

$$\eta_W = B_W \eta_W + \varepsilon_W \tag{5}$$

Eq. (4) and Eq. (5) define the intergroup and intragroup matrix relationships with latent variables of VR sports experience in this study, i.e., the basic form of MSEM.

Since the parent intragroup variance ρ_w^2 and the parent intergroup variance $\rho_b^2 p$ of the vector of variables derived from the sample observation data of the people of VR sports experience is an unbiased statistic, if the ratio between the intergroup variance and the total variance of the variables, i.e., ICC (Intra-Class Coefficient), is measured, the intergroup heterogeneity or intragroup homoscedasticity of the observed variables can be analyzed, as shown in Eq. (6).

$$ICC_M = \rho_M = \frac{\rho_b^2}{\rho_b^2 + \rho_w^2} \tag{6}$$

ICC_M represents the proportion of intergroup variation in the variance of the sample observations. If ICC_M is less than 0.3, it means that the intergroup variation is not

significant and can be handled by the traditional method. On the contrary, if ICC_M is very large, it means that the intergroup variation is significant and must be handled by the multilevel analysis technique [18], or it will result in biased statistical estimation. According to the observed values of the sample of people of VR sports experience in this study, ICC_M was 0.68, which indicated that there were group activities and multilevel data characteristics in the games and sports of VR sports experience, so this study adopted MSEM for empirical analysis.

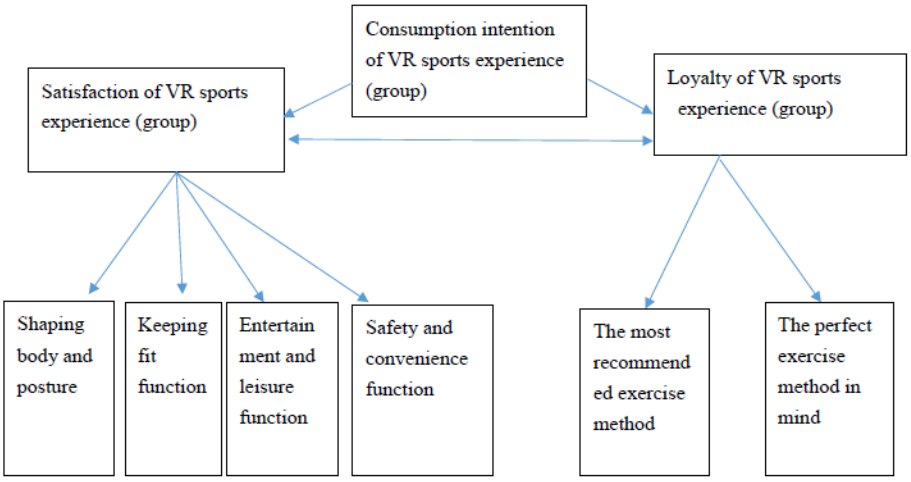
According to [24], the estimated statistic of the intergroup matrix is affected by the sample characteristics, and when the ICC between groups is low, the estimation of factor loadings is prone to be biased, resulting in the underestimation of the error variance and standard deviation, which leads to the rejection of the null hypothesis and committing the statistical Type I error. Therefore, [36] suggested that the sample characteristics should be homogeneous within groups, and the differences should be significant between groups. Also, when the number of samples in each group is not equal, the parameterization should be performed with a more stringent Type I error rate (e.g., a = 0.01). The larger the number of samples and groups, the more complicated the statistical estimation model. The multilevel structural model analysis must be based on a solid measurement model to have appropriate estimation solutions for the parameters measured. Therefore, this study was based on the two-stage MSEM analysis suggested by [29] to find out the most appropriate measurement model, and then carry out the structural model analysis to explore the path relationship between the satisfaction, loyalty and consumption intention of individuals and groups of VR sports experience. The MSEM of the motivations and purposes of the people who had VR sports experience in this study is shown in Figure 1 below.

Group-Level Structural Equation Modeling Path Relationships

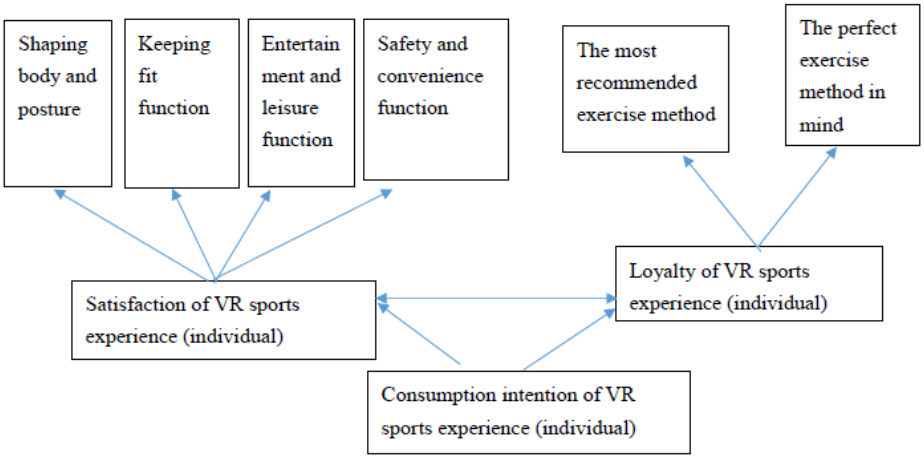
Figure 1 shows that the upper level is the group-level path relationship. The explanatory variable of people of VR sports experience on the left side extracted a corresponding group-level latent contextual explanatory variable (VR sports experience group's satisfaction), and another group-level latent contextual explanatory variable was on the right side (VR sports experience group's loyalty), and the dependent variable in the middle extracted a corresponding group-level latent contextual explanatory variable (VR sports experience group's consumption intention), which is called the latent contextual dependent variable. The three latent variables can be used as important variables in the structural model to conduct regression analysis among the latent variables. Similarly, the lower level is the individual-level path relationship, and the regression analysis between latent variables was also conducted.

4. Empirical Analysis and Results

The sample data of this study were sorted from the data on the people of VR sports experience collected from January to June 2024 through a digital questionnaire. The purpose of this study was to explore the motivational behaviors of the people of VR sports experience, analyze the impact of the satisfaction and loyalty of the motivational behaviors of the VR sports experience groups and individuals on their consumption intention of VR sports experience, in order to understand the relationship between the latent variables of the participants in VR sports experiences such as interactive games, dance, yoga and



(a) Group-Level Structural Equation Modeling Path Relationships



(b) Individual-Level Structural Equation Modeling Path Relationships

Fig. 1. Multilevel Network Structural Equation Modeling (MSEM) for People of VR Sports Experience

qigong, to propose customized marketing service strategies to enhance the consumption intention of people of VR sports experience, and to improve people's goal of sports as well as their physical and mental health.

The sample data of this study consisted of 62 different VR sports experience groups and 1650 VR sports experience individuals. Among the 62 VR sports experience groups, there were 10 female groups, 32 male groups, and 20 mixed groups of women and men. Most of the people of VR sports experience were middle-aged people between 40 and 50 years old (accounting for 41.6%). A total of 1,650 questionnaires were collected as the samples of people of VR sports experience, of which 1,425 were valid, including 786 (55.2%) were from women and 639 (44.8%) from men, with the majority of the participants being women. In the groups of VR sports experience, the minimum number of participants was 6, the maximum number was 25, and the average number of participants in the groups of VR sports experience was 16.5 persons.

Based on the sample data, two latent independent variables "satisfaction with VR sports experience" (with four items, Cronbach's a =.92), "loyalty to VR sports experience" (with two items, Cronbach's a = .91), and the latent dependent variable "consumption intention" (Cronbach's a = .95) were chosen via factor analysis. The reliability of these three latent variables was high, indicating good data consistency.

In this study, Mplus8 was used for the empirical analysis of the MSEM, which has the advantage of directly calculating the covariance matrix of the sample without the need to measure the intergroup and intragroup variance matrix relationships, so that the data have consistency and completeness.

Table 1. Definition of input and output variables

Items of participants' motivational behaviors	Narrative statistics			Sa: Satisfaction with VR sports experience				Lo: Loyalty to VR sports experience		Cu: Consumption intention	
	M	SD1	SD2	Sa1	Sa2	Sa3	Sa4	Lo1	Lo2	Cu1	Cu2
Sa1: Shaping body and posture	3.61	0.25	0.21	1.00	0.82	0.78	0.75	0.72	0.79	0.82	0.81
Sa2: Keeping fit function	3.36	0.12	0.35	0.66	1.00	0.76	0.88	0.67	0.76	0.88	0.80
Sa3: Entertainment and leisure function	3.82	0.31	0.28	0.71	0.68	1.00	0.88	0.82	0.78	0.78	0.71
Sa4: Safety and convenience function	3.38	0.14	0.34	0.68	0.73	0.72	1.00	0.65	0.65	0.86	0.82
Lo1: The most recommended exercise method	3.58	0.22	0.18	0.82	0.68	0.83	0.78	1.00	0.68	0.88	0.86
Lo2: The perfect exercise method in mind	3.47	0.34	0.26	0.67	0.72	0.73	0.72	0.85	1.00	0.81	0.89
Cu1: Great value for money	3.39	0.28	0.32	0.83	0.75	0.69	0.65	0.76	0.62	1.00	0.91
Cu2: Planning to purchase	3.35	0.24	0.27	0.64	0.63	0.81	0.68	0.59	0.82	0.82	1.00

Note: The lower triangles are the individual-level correlation coefficients of VR sports experience and the upper triangles are the group-level (group) weighted correlation coefficients. SD1 is the standard deviation of 1425 VR sports experience individuals at the individual level; SD2 is the weighted standard deviation of 62 VR sports experience groups.

As shown in Table 1, for the statistics and correlation coefficients of the 62 VR sport experience groups with the 1425 samples, the mean score of satisfaction, loyalty and consumption intention of VR sport experience participants ranged from 3.35 to 3.82 out of 5. The variance of the scores of motivational behaviors of VR sport experience participants was consistent, with the standard deviation ranging from 0.14 to 0.34. The correlations between items were all significant at the 0.01 level, with coefficients ranging from 0.59 to 0.91, and the correlations between the two consumption intention items (Cu1 and Cu2) were even higher than 0.80, which indicated that the consistency of the questionnaire items was very high.

CFA was conducted on eight observational variables to verify the validity of three latent variables, namely "satisfaction with VR sports experience", "loyalty to VR sports experience" and "consumption intention". The ICC_M was measured to be 0.68 according to the sample data of this study, so the data structure had multilevel characteristics. In the CFA test, the single-level CFA was firstly conducted on the 1425 individual-level sample data to confirm the factor structure of the VR sports experience at the individual level, and then multilevel CFA was carried out to test the factor validity of the three latent variables at the group level. There were 8 observed variables for the single-level CFA model validation. The independent model (SEM1) had no factor constraints and no correlation between the observed variables. The other single-level CFA models were divided into three models based on whether the latent variables were correlated or not: the orthogonal three-factor model (SEM2), which assumed that the correlation among the three latent variables was zero; the diagonal three-factor model (SEM3), which also meant that there was a correlation among the three latent variables and correlation analyses were carried out; and the independent variable single-factor model (SEM4), which assumed that the "satisfaction with the VR sports experience" and "loyalty to VR sports experience" were integrated. The results of the four model fit test indicators are shown in Table 2.

able 2 shows that, among the three hypothesized models, the best data fit was obtained from the diagonal three-factor model (SEM3), with $X^2(48)$ = 378.66, p¡.001, X^2/df=9.62, RMSEA=0.086, CFI=0.971, TLI=0.952, and SRMR=0.062. Comparatively, the fit of three-variable orthogonal SEM2 and independent variable single-factor model SEM4 was not good, indicating that the three latent variables were not independent and uncorrelated, nor were they combined into a single latent variable. Therefore, in this study, the three-variable diagonal model was the best fit for the individual-level factor structure path, and this model was used as the basis for the multilevel CFA analysis.

According to Figure 2, the factor loading values of the measured variables of the latent variables were all greater than 0.80, indicating that the explanatory power and consistency within the latent variables were very high. The correlation between the two latent variables was 0.86 ($p < .001$), indicating high satisfaction with the VR sports experience and high loyalty to the VR sports experience. The correlation between the two independent variables and the dependent variable was also very high. The correlation coefficients between "satisfaction with VR sports experience " and "consumption intention" as well as "loyalty to VR sports experience and "consumption intention" were both 0.85 ($p < .001$).

According to the MSEM, the structure of latent variables at both the individual level and the group level was considered for analysis. This study retained the above-mentioned best model at the individual level (the diagonal model of three latent variables). The latent variables at the group level were sequentially divided into four different structural models.

The first multilevel model (MSEM1) was an independent model, where the eight latent variables were assumed to be irrelevant and independent of each other. This model did not make any assumptions or restrictions on all latent variables, and was regarded as the most elastic baseline model among the group-level path. The structure of latent variables at the group level was set according to the individual level model, which was divided into a three-variable orthogonal model (MSEM2), a three-variable diagonal model (MSEM3), and an independent variable single-factor model (MSEM4). The fit test results of these four groups of multilevel CFA models are shown in Table 2.

As seen in Table 2, MSEM1 did not have any constraints on the latent variable structure, and the results of this model's fit test were the least satisfactory, with $X^2(68)$ =693.78, $p < .001$. The SRMRB indicator (0.768) at the group level showed a very large standardized difference between groups at the group level, and a relatively small residual difference at the individual level (SRMRW=0.068), indicating that the MSEM1 model fit was most unsatisfactory. In the group level, the fit test results were even more unsatisfactory if the three latent variables were assumed to be under the triple orthogonal model (MSEM2), with $X^2(63)$ =641.62, $p < .001$, RMSEA=0.092, indicating that the three latent variables were partially correlated at the group level, and the modeling at the group level was poor as indicated by SRMRB=0.612. As to the two models MSEM3 and MSEM4, the model fit test results were very close. However, the three-factor diagonal model (MSEM3) was found to be better, as shown by the test indicators of $X^2(63)$ =252.82, $p< .001$, RMSEA=0.028, SRMRB=0.112, and SRMRW=0.032.

According to the results of CFA test in this study, the three-factor diagonal model was the best model in the individual level of people of VR sports experience, and the latent independent variables were two-factor diagonal in the group level. Therefore, this study used the MSEM3 structural model to analyze the behavioral motivation and satisfaction of people of VR sports experience as well as the influence of loyalty on consumption intention.

The empirical results of the MSEM in this study are as shown in Figure 2. The R^2 at the group level and the R^2 at the individual level of the model were 0.86 and 0.82, respectively, which indicated that the two latent independent variables in the model can effectively explain the latent dependent variable.

According to the results of the empirical analysis, the path relationship of the group level MSEM was consistent with the basic hypotheses of this study. The satisfaction with VR sports experience had a positive impact on loyalty, the influence coefficient was 0.13, and t-value was significant, which meant that there was not much correlation between the two latent independent variables, and this would not lead to the problem of collinearity statistical fallacy between the latent independent variables. Among the factors of the latent independent variable, satisfaction with VR sports experience (Sa), the most influential factor was the entertainment and leisure function of VR sports experience (Sa3), with an influence coefficient value of 0.38 and a significant t-value, which indicated that the VR sports experience (group) paid the most attention to the entertainment and leisure function of the VR sports experience, and this factor also indirectly influenced the group's consumption intention. The second factor was that VR sports experience had the function of shaping body and posture, with an influence coefficient value of 0.35 and a significant t-value, which indicated that VR sports experience groups believed that the function of shaping body and posture had the second largest influence on their satisfaction. Among

Table 2. Definition of input and output variables

Intragroup	Intergroup	X^{2*}	df	X^2/df	RMSEA	CFI	TLI	SRMR	
								Intragroup	Intergroup
Single-level CFA									
SEM1 independent model		4033.12	52	98.62	0.365	0.000	0.000	0.486	
SEM2 three-variable orthogonal		1282.62	50	24.36	0.279	0.732	0.726	0.425	
SEM3 three-variable diagonal		378.66	48	9.62	0.086	0.971	0.952	0.062	
SEM4 independent variable single-factor		821.25	50	16.71	0.261	0.866	0.887	0.162	
Multi-level CFA									
MSEM1 three-variable diagonal	Independent model	693.78	68	10.26	0.065	0.896	0.936	0.068	0.768
MSEM2 three-variable diagonal	Three-variable orthogonal	641.62	63	10.69	0.092	0.923	0.942	0.055	0.612
MSEM3 three-variable diagonal	Three-variable diagonal	252.82	63	4.36	0.028	0.972	0.972	0.032	0.112
MSEM4 three-variable diagonal	Independent variable single-factor	371.60	61	6.25	0.046	0.943	0.951	0.039	0.267

The X^2 values of all single-level and multi-level models reached the significant level of 0.001.

the important factors of the group's loyalty (Lo) to VR sports experience, the most recommended exercise method (Lo1) had the largest influence coefficient, with a coefficient value of 0.42 and a significant t-value. This indicated that the VR sports experience group believed that they would recommend their teammates and friends and relatives to participate in the VR sports experience activities together, and they had a rather enthusiastic loyalty, and loyalty also indirectly affected the consumption intention of the VR sports experience group. Of the two latent independent variables, the loyalty of VR sports experience activities had a greater effect on the latent dependent variable (consumption intention), with a coefficient value of 0.45 and a significant t-value, indicating that the consumption intention on VR sports experience (group) was strongly influenced by the loyalty of the public.

As shown in Figure 2, the paths of latent variables in the individual level MSEM were all consistent with the basic hypotheses of this study. People's satisfaction in VR sports experience activities had a positive impact on customer loyalty, with an influence coefficient of 0.16, a significant t-value, and the influence coefficient was within a reasonable range that did not lead to covariance statistical problems among the latent independent variables. In the individual-level model, the main factor affecting the satisfaction of individuals in VR sports experience activities (Sa) was that VR sports experience has the function of shaping body and posture (Sa1), with an influence coefficient of 0.38 and a significant t-value, which indicated that the VR sports experience people paid more attention

to the body shape and appearance than the entertainment, leisure, keeping fit functions, and it indirectly influenced people's consumption intention. The main factor of loyalty (Lo) of VR sports experience people (individuals) was the belief that VR sports experience was a perfect exercise method (Lo2), with an influence coefficient value of 0.40, and a significant t-value. This indicated that the psychology of perfect exercise method (Lo2) of VR sports experience was the biggest factor to establish people's loyalty to VR sports experience and it indirectly affected people's consumption intention a great deal. Of the two latent independent variables, the one with greater influence was the loyalty to VR sports experience (Lo), with an influence coefficient of 0.47 and a significant t-value.

The empirical analysis found that the behavioral motivations and preferences of the group level and individual level of VR sports experience were different, leading to differences in satisfaction and loyalty, which indirectly affected the consumption intention of VR sports experience. The behavior of VR sports experience at the group level had some of the attributes of individual people, but the VR sports experience group, considering the common recreational and leisure behaviors of the group and the spirit of mutual assistance, formed another independent group. Therefore, in order to promote the physical and mental health of busy modern people and facilitate leisure sports anytime and anywhere, the needs for sports and leisure by groups and individuals must be comprehensively grasped in the function of VR sports experience and digital related virtual effects, so as to accurately and efficiently provide customized sports and leisure services, and to maximize the benefits of sports, entertainment and health.

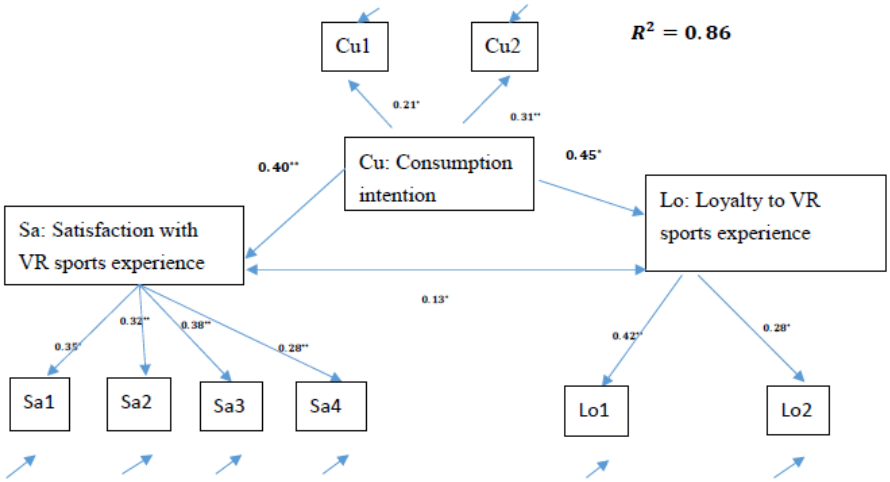
5. Conclusion and Suggestions

5.1. Conclusion

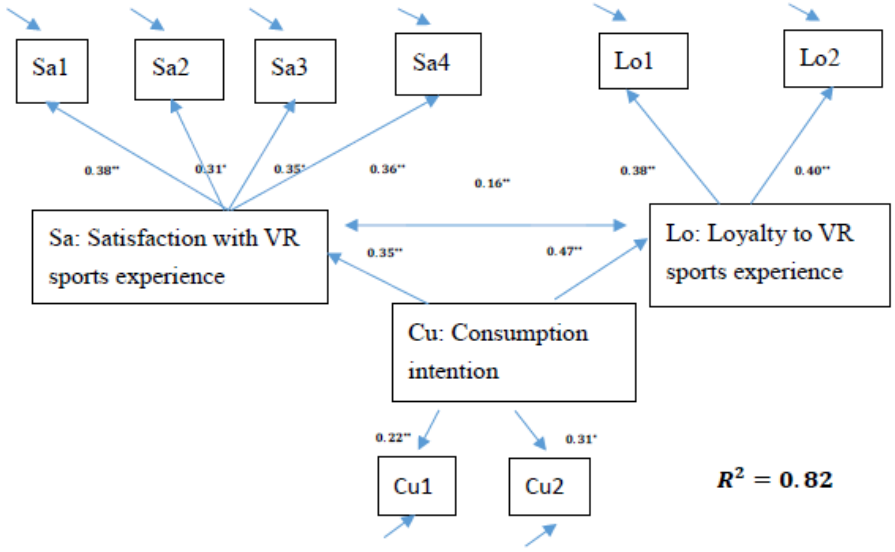
Among the users of VR sports experience, there were significant differences in the demand for sports experience and leisure attributes between groups and individuals, and the traditional SEM model cannot correctly capture the preferences and behaviors of tourists at multiple levels. The empirical results showed that the main factors of the latent variables of satisfaction and loyalty of the group level were different from the main factors of the individual level, which indicates that the sports preferences of VR sports experience groups were different from the sports preferences of individuals, and this indirectly affected the sports consumption intention of groups and individuals.

5.2. Suggestions

1. For VR sports experience operators
   According to the results of this paper, the motivation of VR sports experience groups and individuals was different, leading to differences in their satisfaction and loyalty, and indirectly affecting their intention of exercise and consumption. In terms of the latent variables of satisfaction, VR sports experience groups paid most attention to the recreational and leisure function of VR sports experience (Sa3), while VR sports experience individuals paid attention to the body and posture shaping function of VR sports experience (Sa1). Therefore, it is recommended that related operators of VR sports experience should make timely adjustments according to the functions



(a) Group-Level Structural Equation Modeling Path Relationship



(b) Individual-Level Structural Equation Modeling Path Relationship

Fig. 2. Multilevel Structural Equation Modeling Path Relationships for VR Sports Experience People Note: The above path coefficients are standardized coefficients; **p < 0.05

and services emphasized by VR sports experience in the future planning of VR sports experience functions and services to meet the demand, such as adding the types of VR sports experience games and sports, and increasing the functions of group cooperation and the interactivity of sports. In addition, the main purpose of VR sports experience for individual participants is shaping body and posture, so the operators must provide a variety of on-demand courses and exercise methods, not only to make sports as fun as playing, but also to enable individual participants to get more involved in the VR sports through the provision of accurate service content.

2. For people of VR sports experience
   The results of this study showed that in terms of loyalty, VR sports experience groups thought that VR sports experience was the most recommended exercise method (Lo1), while VR sports experience individuals considered VR sports experience as the most perfect way of exercise and leisure (Lo2). Thus, both group and individual VR sports experience participants, after experiencing special and VR leisure sports, had positive behavioral motivations for the functions and benefits of VR sports experience, and achieved the sports benefits of physical and mental health. It is thus recommended that people who want to have VR sports experience may choose the suitable devices to achieve better sports benefits. For example, some people suffer from gym phobia, because they are too demanding but afraid of failing and causing embarrassment. At this point, VR sports experience devices are suitable for individuals to exercise and workout at home. At the same time, with the high-tech 3D dynamic capture technology or real-time sports posture correction and trajectory guidance, it allows individuals to have the opportunity to choose the right VR device to maximize the benefits of exercise, and to engage in exercise in a very safe environment, which will increase the consumption intention.

References

1. Abeza, G., O'Reilly, N., Reid, I.: Relationship marketing and social media in sport. International Journal of Sport Communication 6(2), 120–142 (2013)
2. Ahmad, M.A., Singh, D.K.A., Mohd Nordin, N.A., Hooi Nee, K., Ibrahim, N.: Virtual reality games as an adjunct in improving upper limb function and general health among stroke survivors. International journal of environmental research and public health 16(24), 5144 (2019)
3. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes 50(2), 179–211 (1991)
4. Anderson, J.C., Gerbing, D.W.: Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin 103(3), 411 (1988)
5. Anderson, K.C., Knight, D.K., Pookulangara, S., Josiam, B.: Influence of hedonic and utilitarian motivations on retailer loyalty and purchase intention: a facebook perspective. Journal of Retailing and Consumer Services 21(5), 773–779 (2014)
6. Bentler, P.M., Liang, J.: Two-level mean and covariance structures: Maximum likelihood via an em algorithm. Multilevel modeling: Methodological advances, issues, and applications pp. 53–70 (2003)
7. Bliese, P.D.: An introduction to multilevel modeling techniques. Personnel Psychology 53(4), 1062 (2000)
8. Capasa, L., Zulauf, K., Wagner, R.: Virtual reality experience of mega sports events: A technology acceptance study. Journal of Theoretical and Applied Electronic Commerce Research 17(2), 686–703 (2022)

9. Chanavat, N., Bodet, G.: Experiential marketing in sport spectatorship services: A customer perspective. European Sport Management Quarterly 14(4), 323–344 (2014)
10. Cho, C.H.: Developing a digital ecosystem for sports technology in the post-pandemic era to create business opportunities. National Sports Quarterly 51(3), 45–50 (2022)
11. Cohen, D.: Consumer behavior. New York, Toronto (1981)
12. Das, G.: Linkages of retailer awareness, retailer association, retailer perceived quality and retailer loyalty with purchase intention: A study of indian food retail brands. Journal of Retailing and Consumer services 21(3), 284–292 (2014)
13. Dash, G., Kiefer, K., Paul, J.: Marketing-to-millennials: Marketing 4.0, customer satisfaction and purchase intention. Journal of business research 122, 608–620 (2021)
14. Electronics, S..: Spherical Insights.: Global Sports Technology Market. Spherical Insights (2023)
15. Flowers, E.P., Freeman, P., Gladwell, V.F.: A cross-sectional study examining predictors of visit frequency to local green space and the impact this has on physical activity levels. BMC Public Health 16, 1–8 (2016)
16. Flowers, E.P., Freeman, P., Gladwell, V.F.: A cross-sectional study examining predictors of visit frequency to local green space and the impact this has on physical activity levels. BMC Public Health 16, 1–8 (2016)
17. Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., Bryant, B.E.: The american customer satisfaction index: nature, purpose, and findings. Journal of marketing 60(4), 7–18 (1996)
18. Foroudi, P., Jin, Z., Gupta, S., Foroudi, M.M., Kitchen, P.J.: Perceptional components of brand equity: Configuring the symmetrical and asymmetrical paths to brand loyalty and brand purchase intention. Journal of business research 89, 462–474 (2018)
19. Goldstein, H., Browne, W.: Multilevel factor analysis modelling using markov chain monte carlo estimation. In: Latent variable and latent structure models, pp. 225–243. Psychology Press (2014)
20. Goldstein, H., McDonald, R.P.: A general model for the analysis of multilevel data. psychometrika 53(4), 455–467 (1988)
21. Griffin, J.: The internet's expanding role in building customer loyalty. direct marketing, 59 (7), 50–53 (1996)
22. Hox, J., Moerbeek, M., Van de Schoot, R.: Multilevel analysis: Techniques and applications. Routledge (2017)
23. Hox, J.J.: Applied multilevel analysis. TT-publikaties (1995)
24. Hox, J.J., Maas, C.J.: The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. Structural equation modeling 8(2), 157–174 (2001)
25. Hoyle, R.H.: Handbook of structural equation modeling. Guilford Publications (2014)
26. Huang, Y.C., Li, L.N., Lee, H.Y., Browning, M.H., Yu, C.P.: Surfing in virtual reality: An application of extended technology acceptance model with flow theory. Computers in Human Behavior Reports 9, 100252 (2023)
27. Jedidi, K., Ansari, A.: Bayesian structural equation models for multilevel data. In: New developments and techniques in structural equation modeling, pp. 149–178. Psychology Press (2001)
28. Jöreskog, K.G., Sörbom, D.: Lisrel 8.80. Lincolnwood, IL: Scientific Software International Inc (2006)
29. Kaplan, D., Elliott, P.R.: A didactic example of multilevel structural equation modeling applicable to the study of organizations. Structural Equation Modeling: A Multidisciplinary Journal 4(1), 1–24 (1997)
30. Kunz, R.E., Santomier, J.P.: Sport content and virtual reality technology acceptance. Sport, Business and Management: An International Journal 10(1), 83–103 (2020)
31. Lee, S.Y., Lin, Y.H.: Digital transformation and application of virtual reality in sports industry. Sport and Health 11(2) (Sep 2022)

32. Lee, S.Y.: Handbook of latent variable and related models, vol. 1. Elsevier (2011)
33. Lee, V., Park, S., Lee, D.: The effect of e-commerce service quality factors on customer satisfaction, purchase intention, and actual purchase in uzbekistan. Global Business & Finance Review (GBFR) 27(3), 56–74 (2022)
34. Love Lock Jones, T.O., Sasser Jr, W.E.: Why satisfied customer defect. Jurnal Harvard Business Review 73(6) (2011)
35. Marcoulides, G.A., Schumacker, R.E.: New developments and techniques in structural equation modeling. Psychology Press (2001)
36. McDonald, R.P.: The bilevel reticular action model for path analysis with latent variables. Sociological Methods & Research 22(3), 399–413 (1994)
37. Muthén, B., Satorra, A.: Multilevel aspects of varying parameters in structural models. In: Multilevel analysis of educational data, pp. 87–99. Elsevier (1989)
38. Muthén, B.O.: Multilevel covariance structure analysis. Sociological methods & research 22(3), 376–398 (1994)
39. Muthén, B.O.: Mean and covariance structure analysis of hierarchical data (2011)
40. Norouzi-Gheidari, N., Hernandez, A., Archambault, P.S., Higgins, J., Poissant, L., Kairy, D.: Feasibility, safety and efficacy of a virtual reality exergame system to supplement upper extremity rehabilitation post-stroke: a pilot randomized clinical trial and proof of principle. International journal of environmental research and public health 17(1), 113 (2020)
41. Oliver, R.L.: What is customer satisfaction? Wharton Magazine 5, 36–41 (1981)
42. Ravi, S.: Multilevel statistical models, 3rd edn (2005)
43. Sagnier, C., Loup-Escande, E., Lourdeaux, D., Thouvenin, I., Valléry, G.: User acceptance of virtual reality: an extended technology acceptance model. International Journal of Human–Computer Interaction 36(11), 993–1007 (2020)
44. Schiffman, L., Kanuk, L.: Consumer Behavior. Prentice Hall International Editions Series, Prentice Hall (2000)
45. Shoemaker, S., Lewis, R.C.: Customer loyalty: the future of hospitality marketing. International journal of hospitality management 18(4), 345–370 (1999)

Yan-Hui Li is a PhD student of Sports Coaching College at Beijing Sport University. His research field focuses on Sports training and management, Analysis of Taekwondo techniques and tactics, and Outdoor education exploration. His academic papers have been published in journals like Journal of Taekwondo Sports, Journal of Physical Education Fu Jen Catholic University, etc.

Cheng-Sheng Lin received his PhD from National Chung Hsing University in Taiwan (2006). He is currently an Assistant Professor at Formasa University. Research expertise: agricultural digital technology, economic benefit analysis and agricultural digital marketing.

Che-Jen Chuang is an Associate Professor in the Department of Tourism and Leisure Management at Vanung University. He has previously served as the Dean of International & Cross-Strait Affairs and the Dean of the Chinese Language Center at Vanung University. He obtained his Ph.D. from Maejo University, School of Tourism Development. His research expertise includes tourism planning strategic management, leisure behavior research, and research methodology in tourism.

Jui-Liang Hsu is a Associate Professor of the Department of Leisure and Sports Managemen at Cheng shiu University. He obtained his Ph.D. from the Department of Bio-Industry

Technology, Da Yeh University. His research field focuses on Leisure and Sports Management. His academic papers have been published in journals like International Conference on Economics, and Management of Business, Innovation and Technology, IOP Conference Series: Earth and Environmental Science, the International Journal of Water, Life Science Journal. To date, he has obtained five utility model patents from the Republic of China.

Yu-Jui Li is an Associate Professor of the Department of Recreation and Sports Management at University of Taipei, and also concurrently serves as the Dean of General Affairs at University of Taipei. He obtained his Ph.D. from the School of Sport and Exercise Science, University of Northern Colorado. His research field focuses on Sports Management, Sport Administration, Sports Marketing and Marine Sport.

Received: December 18, 2024; Accepted: March 10, 2025.