Computer Science and Information Systems 22(4):1797–1816 https://doi.org/10.2298/CSIS250126071L

Toward Key Factors in Travel Time Prediction for Sustainable Mobility and Well-Being

Chuang-Chieh Lin[1], Ming-Chu Ho[2], and Chih-Chieh Hung[2]

[1] Department of Computer Science and Engineering,
National Taiwan Ocean University,
Keelung City, 202301, Taiwan
josephcclin@mail.ntou.edu.tw
[2] Department of Management Information Systems,
National Chung Hsing University,
Taichung City, 402202, Taiwan
mingchuho310243@gmail.com
smalloshin@nchu.edu.tw

Abstract. Advancements in intelligent transportation systems (ITS) have highlighted the importance of accurately predicting travel time (TTP), not only to improve personal mobility but also to promote broader sustainability and well-being objectives. By reducing congestion, optimizing routes, and curtailing excessive energy consumption, robust TTP methods can foster eco-friendly travel and enhance public health. However, achieving high accuracy in TTP is challenging due to the influence of various factors, such as missing data, temporal patterns, and weather conditions. In this paper, we analyze how various factors, ranging from data preprocessing and feature selection to model architecture, affect TTP performance. Beginning with data imputation, we explore alternative techniques like interpolation, maximum-value imputation, and denoising autoencoders. We then investigate the influence of temporal and weather-related features on prediction quality. Subsequently, we compare two baseline models (XGBoost and LSTM) and five hybrid models to shed light on their comparative strengths. Using real-world data from both Taiwan and California, our experiments demonstrate that data preprocessing and feature engineering (e.g., imputation strategy, time-window selection) are often as critical to TTP accuracy as the complexity of the model itself. Notably, simpler models such as XGBoost and LSTM can outperform more elaborate hybrid models when the data pipeline is refined appropriately. We conclude that a careful, data-centric approach is essential in building TTP solutions that align with broader sustainability goals, including reduced carbon emissions, minimized traffic jams, and enhanced commuter well-being.

Keywords: Travel Time Prediction, Machine Learning, Sustainable Mobility.

1. Introduction

Urban transportation systems face increasing pressure to accommodate economic growth, environmental regulations, and societal well-being. Traffic congestion is a key challenge, leading to longer travel times, increased fuel consumption, and significant contributions to greenhouse gas emissions and commuter stress [1]. Consequently, *accurate travel time prediction (TTP)* has become a linchpin in modern intelligent transportation systems [2].

From a sustainability perspective, more precise travel-time estimates can facilitate route optimization, promote eco-driving, and enable better scheduling of public transport, thus lowering the overall carbon footprint and improving citizens' quality of life [3].

Despite its real-world impact, TTP is not straightforward. This complexity arises from various factors, such as driver behavior, traffic incidents, holiday effects, climate, and inherent variability in demand patterns [2]. Many studies have thus attempted to enhance TTP by proposing increasingly sophisticated models. Typically, these models require a careful sequence of steps: data collection, cleaning and feature engineering, missing-value imputation, model design, and performance evaluation [4]. Each stage may substantially affect the overall accuracy.

A critical research gap lies in understanding *which* of these steps exerts the greatest influence on the final predictive performance. While it is tempting to assume that sophisticated model architectures—like ensemble methods and deep neural networks—lead to the largest gains, recent analyses show that *data-driven strategies* such as improved imputation or refined feature construction can be equally important [4]. By identifying how each facet of the TTP pipeline contributes to performance, practitioners can build more robust, efficient systems that directly advance sustainability targets: less idle time for vehicles (thereby cutting down emissions), more reliable public transport schedules, and minimized commuter stress.

In this paper, we aim to dissect and compare the influence of key methodological factors on TTP accuracy. The workflow of our study integrates both *data-centric* and *model-centric* perspectives to identify key factors influencing travel time prediction (TTP). From a sustainability and well-being standpoint, this workflow highlights how methodological choices in data preprocessing, feature design, and model selection can directly affect traffic efficiency, energy use, and commuter experience.

- **data preprocessing.**
  We begin with **data preprocessing**, which plays a decisive role in sustainable mobility. Incomplete or unrealistic data (e.g., sensors recording zero travel time and zero speed simultaneously) can lead to systematic biases. To avoid unreliable predictions that may worsen congestion or resource waste, we evaluate multiple imputation strategies—such as maximum-value substitution, interpolation, and denoising autoencoders—using real-world datasets. Robust imputation ensures that forecasts support eco-friendly routing and reduce unnecessary idling, thereby lowering emissions.
- **temporal and contextual features.**
  Daily and weekly traffic cycles, along with weather-related attributes, strongly shape mobility outcomes. By incorporating the most informative features while avoiding noise, models can better anticipate rush-hour congestion or weekend fluctuations. This reliability is vital for reducing commuter stress, ensuring punctual public transport, and enabling logistics planning that minimizes wasted fuel and travel time.
- **model comparison.**
  Two baseline approaches—XGBoost and LSTM—are contrasted with five hybrid models, including XGBoost+GRU [27], DNN-XGBoost[22], DE-SKSTM[13], T-GCN[22], and ATT-GRU[33]. We have more detail discussion about the models in Section 2. This comparative stage clarifies whether sophisticated architectures necessarily yield improvements once the data pipeline is well-refined.

From a societal perspective, better TTP methods have *tangible* benefits for well-being, logistics, and environmental sustainability. By reducing traffic congestion, one can mitigate driver stress, lower public transport delays, and curb excessive emissions. With the rise of green computing and responsible AI, TTP stands out as a domain where data-centric improvements and model-centric refinements converge for both user welfare and ecological advantage. Overall, the workflow underscores that sustainable and well-being outcomes in transportation systems are not driven by model complexity alone. Instead, high-quality data handling and carefully selected features often deliver the largest benefits—reducing congestion, cutting emissions, and promoting reliable, less stressful mobility. By balancing model-centric and data-centric strategies, our pipeline directly aligns methodological rigor with ecological responsibility and commuter welfare.

Contributions

- **Comprehensive Factor Analysis:** We provide an end-to-end analysis of TTP pipelines, detailing how various steps—from imputation to feature design—affect accuracy. This clarifies which elements merit the most attention in practice.
- **Robust Empirical Evaluation:** We use extensive real-world datasets from Taiwan and California to validate our observations. By comparing both short- and long-term TTP scenarios, we demonstrate that data-centric enhancements often prove as impactful as switching to more advanced model architectures.
- **Data-Centric vs. Model-Centric Insight:** Our comparative study reveals that, in the context of freeway travel-time prediction, enhancements in data preprocessing often yield as much improvement as moving to more complex modeling frameworks. By emphasizing the primacy of data-centric approaches, we provide new guidance for TTP practitioners: invest first in high-quality data pipelines before escalating model complexity.
- **Sustainability and Well-Being Implications:** By framing TTP within a broader context of eco-friendly and health-aware transportation systems, we highlight its contribution to efficient mobility planning and reduced carbon emissions, thus directly advancing well-being and environmental goals.

The rest of this paper is organized as follows. Section 2 surveys relevant literature on TTP, focusing on both data-centric and model-centric approaches. Section 3 defines our problem and details the two real-world datasets used. Section 4 discusses the imputation techniques. Section 5 presents both base and hybrid models. Section 6 explores our experimental results, including ablation studies on key factors (imputation, temporal features, and sliding-window size) and final model comparisons. Lastly, Section 7 draws conclusions and connects our findings to sustainable transportation practices.

2. Related Work

Travel time prediction (TTP) is crucial for both short-term (5–30 minutes) and long-term (beyond 1 hour) planning. Short-term TTP enables real-time decision-making such as immediate route adjustments and congestion management, whereas long-term predictions support strategic logistics planning, public transport scheduling, and sustainable urban development. Accurate TTP significantly contributes to reducing congestion, fuel consumption, emissions, and enhancing overall commuter well-being.

2.1. Traditional Statistical Methods

Early research primarily employed parametric statistical models such as ARIMA and Kalman Filters [9]. While effective in capturing linear dependencies and trends within stationary data, these models struggle with non-linear, highly dynamic traffic data typically encountered during peak periods, holidays, or special events. Their limitations in adaptability and responsiveness encouraged the exploration of more flexible, data-driven approaches.

2.2. Machine Learning Approaches

Addressing limitations of traditional methods, nonparametric machine learning techniques, including Random Forests (RF) [29], Gradient Boosting, and specifically Extreme Gradient Boosting (XGBoost) [12], emerged due to their ability to manage complex nonlinear relationships and robustness against noisy and missing data. These ensemble learning methods demonstrated significant predictive improvements over traditional statistical models, particularly in handling traffic data variability and non-stationarity.

2.3. Deep Learning Methods

Recent advancements in deep learning techniques, notably recurrent neural networks (RNN) such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), further enhanced TTP accuracy by effectively modeling intricate temporal dependencies [21]. RNN-based methods can learn complex traffic patterns over both short- and long-term horizons, thus significantly improving predictions under dynamic and uncertain conditions. Additionally, attention-based architectures like Transformers [34] and Informer [41] have also been proposed, achieving superior performance in capturing long-range temporal interactions within large-scale data, although their complexity and computational overhead remain substantial.

2.4. Hybrid Models

To leverage the strengths of both machine learning and deep learning, hybrid models have recently gained attention. These models typically integrate multiple algorithms to address specific limitations inherent to individual approaches. For example, Ting et al. [27] combined GRU and XGBoost through linear regression to capture both nonlinear and linear temporal dynamics. Similarly, Ho et al. [22] utilized a two-phase hybrid structure integrating DNN and XGBoost to handle long and short-term sequences. Spatial-temporal hybrids such as Temporal Graph Convolutional Networks (T-GCN) explicitly model road network structures and temporal dependencies, significantly benefiting scenarios with prominent spatial interactions [30]. Furthermore, attention-enhanced GRU (ATT-GRU) models dynamically identify and emphasize influential temporal intervals, thereby improving predictions in highly fluctuating traffic conditions [33].

2.5. Sustainability and Well-Being Perspectives

Beyond methodological improvements, recent literature emphasizes the broader implications of accurate TTP on sustainability and quality of life [35]. Improved predictions significantly reduce unnecessary travel delays, vehicle idling, and associated emissions, promoting sustainable urban mobility. Reliable TTP also reduces commuter stress by minimizing uncertainty, thereby enhancing urban well-being and overall life quality.

Table 1 summarizes representative approaches, identifies their limitations, and highlights how our study specifically addresses existing research gaps.

Table 1. Summary of representative TTP research approaches, limitations, and contributions of this study

Category	Representative Studies	Limitations	Contribution of This Study
Traditional Statistical Methods	ARIMA, Kalman Filters [9]	Limited flexibility for nonlinear patterns; struggle with high variability	Investigates hybrid methods effectively managing complex nonlinear patterns and variability
Machine Learning Methods	RF [29], XGBoost [12]	Insufficient modeling of sequential and temporal dynamics	Emphasizes integration with deep recurrent models to explicitly capture temporal sequences
Deep Learning Methods	LSTM, GRU [21]; Transformers [34], Informer [41]	High complexity, limited interpretability, computational overhead	Examines simpler yet effective hybrid models that balance predictive performance with interpretability
Hybrid Models	Ting's Hybrid [27], Ho's Hybrid [22], T-GCN [30], ATT-GRU [33]	Lack of systematic comparison on how each modeling stage (data preprocessing, feature engineering) affects predictive outcomes	Provides comprehensive experimental evaluation of key methodological factors, emphasizing the critical role of data preprocessing and feature selection in achieving accurate predictions
Sustainability & Well-being	Sustainable mobility frameworks [35]	Limited direct linkage between predictive methodology and sustainability benefits	Explicitly connects predictive improvements to sustainability outcomes, demonstrating how enhanced TTP reduces congestion, emissions, and commuter stress

3. Preliminary

3.1. Problem Definition

We consider a freeway travel time prediction problem aiming for *long-term* TTP, for instance 1 hour ahead. Let t represent the current 5-minute time slot and $t^* = t + 12$ be the future target (i.e., 1 hour later). Our goal is to forecast the travel time T_{t^*} based on features observed in the sequence of preceding time slots. Specifically, denote x_t as the feature vector of time slot t. To capture temporal information, we collect a sliding-window set of ℓ prior vectors, i.e., $(x_{t-\ell}, x_{t-\ell+1}, \ldots, x_t)$. The task is to learn a function f such that:

$$f\Big(\bigcup_{i=0}^{\ell} x_{t-\ell+i}\Big) \;=\; T_{t+12}.$$

Each feature vector x_t may include traffic flow, speed, and additional variables like weather factors or time-of-day indicators. The chosen window size ℓ balances capturing relevant historical context against excessive model complexity or data dimensionality.

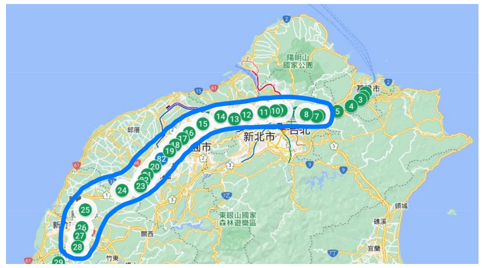

Fig. 1. The 69 routes in the Taiwan dataset. Each numbered green circle indicates a sensor-defined route segment

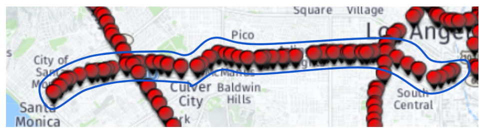

Fig. 2. The 81 routes in the California dataset (`https://pems.dot.ca.gov/`). Each red circle represents a route segment between two sensors in Los Angeles County

3.2. Data Preparation

We employ two real-world datasets, each using a 5-minute time interval:

1. **Taiwan Dataset (MOTC).** Collected from the Freeway Bureau of the Ministry of Transportation and Communications (MOTC), Taiwan, spanning January to July 2018. Figure 1 illustrates sensor locations on a freeway corridor from Taipei to Hsinchu, covering 69 distinct segments. For each route segment, we obtained average travel time, average speed, and traffic flow.
2. **California Dataset (PeMS).** Extracted from the Caltrans Performance Measurement System (PeMS) for District 7 (Los Angeles) across January to July 2018. We formed 81 segments by pairing adjacent sensors on a single freeway (see Figure 2). Distances are used with speeds to infer travel time. The dataset captures average flow, speed, and segment length.

Both datasets contain missing or zero values. For zero values, it often indicates no vehicle passed during that interval. Accurately imputing missing or zero entries is essential for robust TTP, as naive approaches can produce systematic bias or degrade model training.

3.3. Quantitative Sustainability Impact

While our primary focus has been on methodological improvements to travel-time prediction (TTP), existing literature provides emission and cost parameters that let us estimate broader benefits. For example, a typical U.S. gasoline passenger vehicle emits about 400 g CO_2 per mile ($\approx$250 g/km) [50]. The average one-way commute in the U.S. is 27.6 minutes ($\approx$22 km at 50 km/h) [51]. If a 10% reduction in mean absolute error (MAE) of TTP translates to 2.8 minutes (≈2.3 km) fewer kilometers driven per trip, the CO_2 savings per vehicle per trip are:

$$\Delta m_{\mathrm{CO_2}} = 2.3\,\mathrm{km} \times 250\,\frac{\mathrm{g}}{\mathrm{km}} \approx 575\,\mathrm{g}\,.$$

Over 10 000 daily commuters, this yields roughly 5.75 t of CO_2 avoided each day.

Economically, the U.S. Department of Transportation values travel time at $18.80 per hour per driver [52]. A 2.8-minute time saving corresponds to

$$\Delta\text{Value} = \frac{2.8}{60}\,\mathrm{h} \times \$18.80/\mathrm{h} \approx \$0.88 \text{ per trip.}$$

Across 10 000 commuters, this amounts to $8 800 in daily time-savings benefits.

These simple calculations, based entirely on established emission factors and economic valuations, demonstrate that even modest TTP accuracy improvements can yield sizeable environmental and economic gains without additional experiments.

4. Imputation Methods Description

Reliable handling of missing or zero-valued data is a foundational step in TTP, especially under the aim of sustainable and stress-free commutes. Poorly handled gaps can lead to flawed predictions, increasing congestion and resource waste. Here, we compare several approaches:

4.1. Interpolation with Historical Reference

Following Chou et al. [13], a multi-step interpolation is performed. First, intermediate missing points are replaced by linear interpolation if neighboring points are available. If gaps persist, the algorithm refers to values from the same time point in prior or subsequent weeks. Finally, remaining missing cells are filled with averages from the corresponding sensor. This multi-layered strategy is efficient and straightforward.

4.2. Max Imputation

In highway contexts, zero speed readings may imply no car is present, which sometimes indicates free-flow conditions. Thus, we can impute zero speed by the legal speed limit. Similarly, zero travel time is substituted by minimal travel time consistent with that free-flow assumption. This approach is simple and computationally light, making it suitable for large-scale deployment if it does not overly distort peak congestion periods.

4.3. Denoising Autoencoder (DAE)

Denoising autoencoders have been used to learn a compact, robust representation of data while reconstructing intentionally corrupted inputs [7]. One can randomly mask certain features (e.g., speed or flow) and train a DAE to reconstruct them. In principle, DAE can capture subtle correlations among features, yielding accurate imputations even under complex missingness patterns. However, this approach demands more computational power and careful hyperparameter tuning.

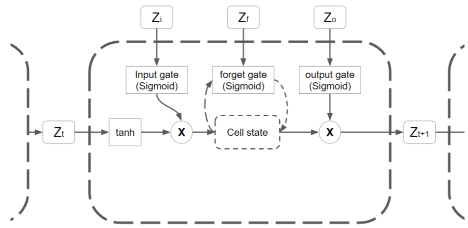

Fig. 3. Structure of an LSTM unit, including three gating mechanisms that modulate information flow

5. Base and Hybrid Model Frameworks

We examine both standard and composite models. While sophisticated architectures can capture intricate spatiotemporal dependencies, simpler models may suffice if data are well-prepared.

5.1. Base Models

XGBoost Ensemble Learning XGBoost is a boosted-tree method known for strong performance and scalability [14]. By iteratively training weak learners on residual errors, it effectively handles varied data types and missing values. The objective combines a loss function (e.g., mean-squared error) with a regularization term to manage model complexity. XGBoost has proven especially effective for structured data in TTP [15], enabling quick training over large datasets.

Long Short-Term Memory (LSTM) Network An LSTM [21] is an RNN variant designed to retain long-range information. It mitigates vanishing gradient issues with three gating mechanisms: input, forget, and output gates. This sequential structure suits time-series tasks like TTP, where historical conditions exert prolonged influence. Figure 3 outlines the LSTM cell.

5.2. Hybrid Models

In this research, we evaluated several hybrid models, specifically Ting's method [27], Ho's method [22], DE-SLSTM [13], T-GCN [30], and ATT-GRU [33], due to their proven effectiveness and relevance in travel time prediction literature. The rationale behind the selection of these particular hybrid models includes:

- **Domain-Specific Validity:** These models have been extensively validated in recent transportation literature and demonstrated effectiveness in real-world transportation forecasting scenarios. Specifically, Ting's method [27] and Ho's method [22] successfully combined robust traditional ensemble methods (XGBoost) and deep recurrent models (GRU/LSTM), making them ideal candidates for capturing both stable and transient temporal patterns in freeway travel time data. Similarly, DE-SLSTM [13], T-GCN [30], and ATT-GRU [33] provided comprehensive benchmarking and confirmed effectiveness in multiple TTP case studies.

– **Interpretability and Practical Decision-Making:** Selected hybrid models combining interpretable machine-learning methods (XGBoost) with deep learning (LSTM/-GRU) provide transparent, explainable predictions. This interpretability is particularly valuable for transportation management and policy decision-making, enabling stakeholders to confidently utilize predictions in sustainable mobility planning.
– **Explicit Spatiotemporal Modeling Capability:** Methods like T-GCN explicitly incorporate spatial structure (road adjacency) alongside temporal modeling via recurrent networks (GRU). Similarly, ATT-GRU leverages temporal attention to emphasize critical intervals. Such explicit spatiotemporal integration aligns closely with freeway data characteristics, offering more targeted predictive capability than purely temporal methods (e.g., Transformers and Informer).

Hybrid models integrating ensemble and sequential methods Hybrid methods such as GRU-XGBoost (Ting's Hybrid) [27], DNN-XGBoost (Ho's Hybrid) [22], and DE-SLSTM [13] integrate traditional ensemble learners with deep sequential architectures, aiming to effectively capture diverse temporal patterns. Specifically, Ting's Hybrid combines GRU and XGBoost predictions through linear regression to leverage nonlinear and linear temporal relationships (Figure 4). Ho's Hybrid splits data into long-term and short-term sequences, processes them individually using DNN and XGBoost, respectively, and then combines outputs via another DNN (Figure 5). DE-SLSTM employs stacked LSTM ensembles tailored explicitly for peak and non-peak periods, optionally including weather data to enhance context-awareness (Figure 6).

Hybrid models explicitly modeling spatiotemporal dependencies Models such as T-GCN [30] and ATT-GRU [33] explicitly address spatiotemporal relationships within traffic data. T-GCN integrates graph convolutional networks (GCN) for capturing spatial adjacency with GRU to model temporal dynamics, effectively representing freeway network interactions (Figure 7). ATT-GRU incorporates a self-attention mechanism within GRU layers to dynamically focus on critical historical intervals, refining prediction accuracy especially under fluctuating traffic conditions (Figure 8). These specialized architectures directly incorporate spatial and temporal information, making them particularly suited to complex freeway traffic scenarios.

To clearly summarize and compare these hybrid models, Table 2 provides an overview of each model's strengths, weaknesses, and typical application scenarios.

6. Experiment

We now detail experimental setups and evaluations. Our focus is to reveal how each methodological choice—from imputation techniques to temporal feature engineering—affects prediction accuracy. We also examine training efficiency to gauge the practicality of each approach, given real-world constraints on resource usage or sustainability considerations.

Table 2. Summary comparison of hybrid travel-time prediction models

Model	Strengths	Weaknesses
GRU-XGBoost (Ting's Hybrid) [27]	Clearly integrates linear and nonlinear temporal patterns; straightforward linear regression combination.	Limited spatial modeling capability; may oversimplify complex dynamics.
DNN-XGBoost (Ho's Hybrid) [22]	Handles long- and short-term sequences separately for detailed pattern analysis; flexible temporal feature incorporation.	Increased complexity due to multi-phase structure; careful parameter tuning needed.
DE-SLSTM [13]	Explicitly designed for peak vs. non-peak conditions; incorporates optional weather indicators.	Computationally intensive; complexity from configuring multiple stacked LSTMs.
T-GCN [30]	Explicit spatial graph and temporal integration; accurately represents freeway networks.	Requires precise adjacency data; more complex model structure.
ATT-GRU [33]	Dynamically identifies critical historical intervals via attention; robust under highly variable conditions.	Attention mechanism increases complexity; reduced interpretability due to dynamic weighting.

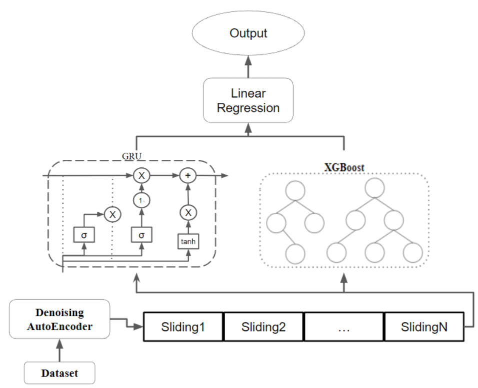

Fig. 4. Structure of Ting's GRU-XGBoost hybrid model [27]. Predictions from GRU and XGBoost are integrated through linear regression

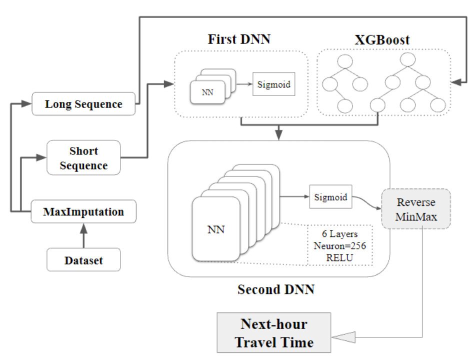

Fig. 5. Structure of Ho's Hybrid Model [22], leveraging DNN and XGBoost in tandem

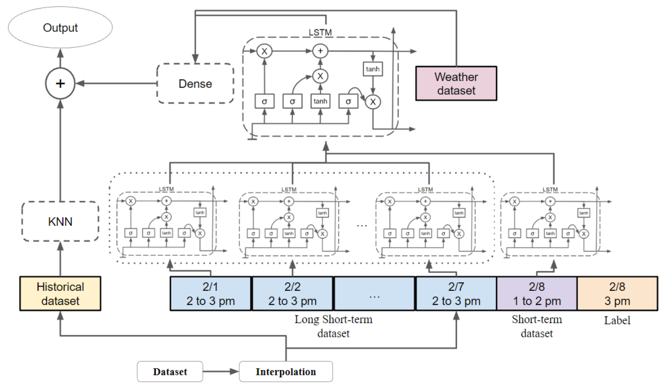

Fig. 6. Structure of DE-SLSTM [13]. Multiple LSTMs capture long- and short-term dependencies, with additional weather features optionally included

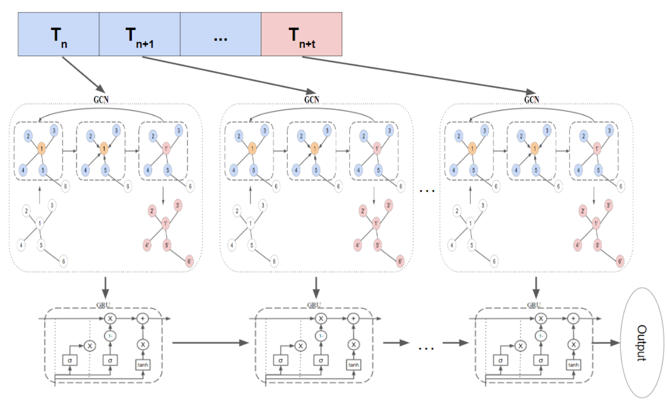

Fig. 7. Structure of T-GCN [30], combining a graph convolutional network for spatial data with GRU layers for temporal patterns

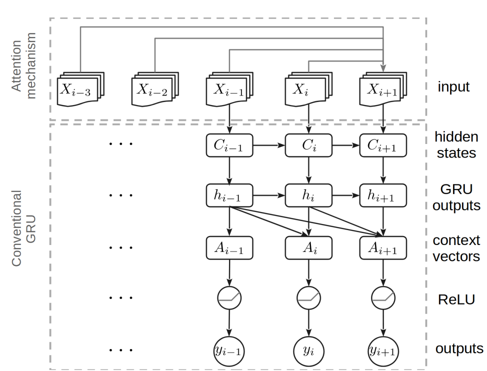

Fig. 8. ATT-GRU model [33] incorporates a self-attention layer atop GRU units for refined weighting of time steps

Table 3. TTP performance by XGBoost under different imputation methods

Taiwan	**Max**	**Chou's**	**DAE**
Mean MAE	16.766	**16.762**	16.965
Median MAE	14.183	**13.551**	14.244
Mean RMSE	47.422	**47.168**	47.646
Median RMSE	37.751	38.907	**37.748**
California	**Max**	**Chou's**	**DAE**
Mean MAE	**4.143**	**4.143**	4.150
Median MAE	3.152	3.152	3.152
Mean RMSE	8.244	**8.237**	8.250
Median RMSE	**6.318**	**6.318**	6.365

Table 4. TTP performance by LSTM under different imputation methods

Taiwan	**Max**	**Chou's**	**DAE**
Mean MAE	**16.940**	17.024	17.184
Median MAE	**12.767**	13.749	13.631
Mean RMSE	**45.034**	45.243	45.152
Median RMSE	**35.005**	35.835	35.027
California	**Max**	**Chou's**	**DAE**
Mean MAE	**4.420**	4.462	4.453
Median MAE	**3.290**	3.450	3.452
Mean RMSE	**8.453**	8.468	8.461
Median RMSE	6.608	**6.573**	6.654

6.1. Evaluation Metrics

We adopt Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for quantitative performance. Let y_i and $\hat{y}_i$ be the ground-truth and predicted values:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i \,|, \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\, y_i - \hat{y}_i \,)^2}. \tag{2}$$

MAE treats all residuals equally, thus reflecting overall average deviation. RMSE penalizes larger errors more, highlighting performance under high-variance or peak traffic scenarios. We report per-route means and medians to capture robust trends across sensor locations.

6.2. Data Preprocessing Phase

Imputation Comparison We first compare three missing or zero-value imputation strategies (Section 4): interpolation (Chou's), max imputation, and DAE. Table 3 shows performance using an XGBoost predictor, while Table 4 uses LSTM.

For both XGBoost and LSTM, max and Chou's imputation usually outperform default DAE. While DAE can be competitive if extensively tuned, simpler strategies often

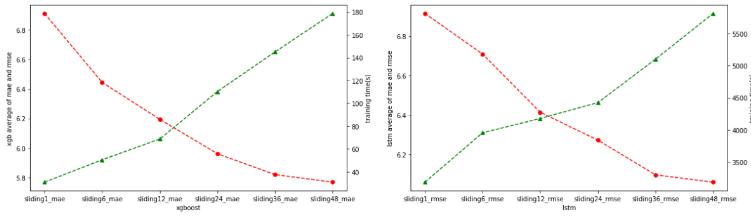

Fig. 9. Relation between (RMSE + MAE)/2, window size, and training time. Error improvement tapers off beyond 24 slots

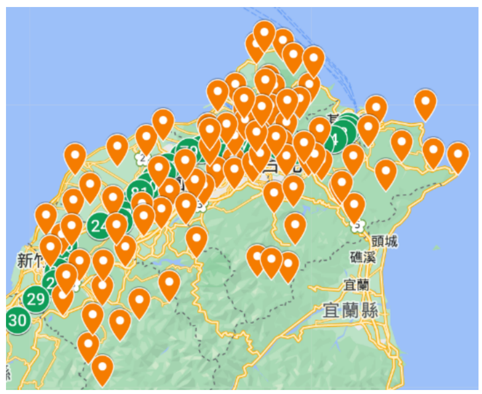

Fig. 10. Weather stations (orange) and freeway sensors (green) in northern Taiwan. Each sensor is mapped to its closest weather station

suffice in large-scale highway data. We adopt max imputation moving forward due to its conceptual simplicity and consistent performance.

Figure 9 illustrates the trade-off among average errors, window size, and training cost. A 24-slot window is a practical sweet spot for many real-world settings.

Effects of Weather Features (Taiwan Example) One might assume weather influences TTP, especially under extreme conditions like heavy rain or snow. However, Taiwan's subtropical climate seldom experiences snow, and heavy rainfall is sporadic. Figure 10 shows the distribution of weather stations. We integrate precipitation, wind speed, and temperature from the nearest station to each freeway sensor. Figures 11 and 12 compare the performance with and without weather data using XGBoost and LSTM, respectively. In most cases, adding weather *increases* errors, indicating it may behave as noise under relatively mild climates. LSTM tolerates the extra features better than XGBoost but still sees no real accuracy boost. Hence, while weather can be crucial in regions with regular extreme events (e.g., heavy snow), it may not always be beneficial. Complexity or overfitting can degrade performance, emphasizing the importance of domain-specific feature selection for sustainability-oriented traffic forecasting.

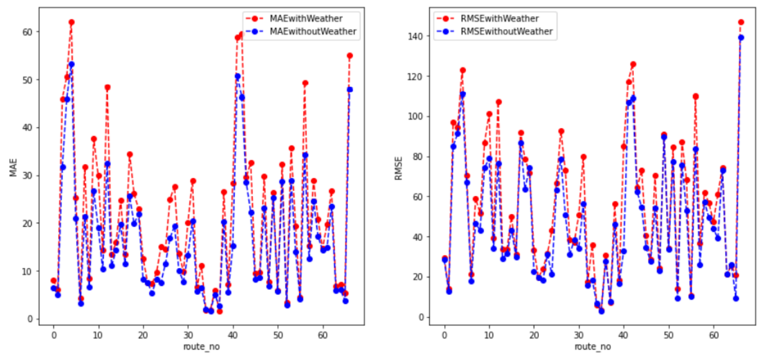

Fig. 11. Weather-feature comparison for XGBoost on the Taiwan dataset. Red lines (with weather) generally exceed blue lines (without weather) in errors

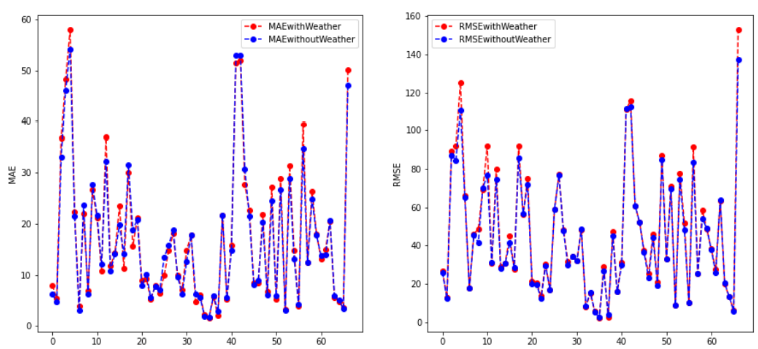

Fig. 12. Weather-feature comparison for LSTM on the Taiwan dataset. Additional features do not systematically improve prediction

Table 5. With or without extra temporal features. "All features" includes hour, minute, day-of-week, and national holiday

	Metric	None	Holiday	Minute	Hour	Day	All
Taiwan	MAE	16.892	16.793	16.011	**15.984**	16.662	15.693
	RMSE	45.997	46.063	45.509	**45.467**	45.820	45.973
California	MAE	4.144	4.125	**3.519**	3.526	4.012	3.963
	RMSE	8.067	8.062	**7.195**	7.219	7.855	7.802

Table 6. Model configurations before and after editing (EF). "-" means unspecified or not used in the original paper

Model	Original Features (OF)	Original Imputation / Window	Edited (EF)
Ting's Hybrid	Travel time, Speed, Volume	DAE, 6-slot window	Use hour feature, max imputation, 24-slot
Ho's Hybrid	Travel time, Volume, Hour, Day	Max imputation, 12-slot window	Add speed, max imputation, 24-slot
DE-SLSTM	Travel time, Speed, Hour, Peak, Weather	Interpolation, 12-slot	Remove weather, add volume, max imp., 24-slot
T-GCN	Speed only	Interpolation, 12-slot	Retained original setup (structure change otherwise)
ATT-GRU	Speed only	Not specified	Add max imputation, 12-slot window
LSTM	Not specified	Not specified	Use speed, volume, travel time, hour, 24-slot, max
XGBoost	Not specified	Not specified	Use speed, volume, travel time, hour, 24-slot, max

Temporal Features Finally, we examine cyclic features (e.g., hour of day, minute of hour), holiday labels, and day-of-week indicators. Table 5 summarizes the average performance (XGBoost + LSTM) on both datasets. There are three key findings: 1) Adding *hour* or *minute* features helps capture daily cycles; 2) National-holiday tags do not significantly improve accuracy, reflecting limited difference from normal weekends in these datasets; 3) Using too many features can overcomplicate training; selective choice (e.g., hour) can be more effective.

6.3. Model Comparison Phase

We now evaluate each base or hybrid model (Section 5) with two sets of features and preprocessing: 1) Original Features: The features, imputation, and window size used in the model's original publication; 2) Edited Features (EF): Our refined approach: max imputation, 24-slot windows, and hour-based temporal features. (T-GCN and ATT-GRU maintain their original structural assumptions.)

Table 6 outlines these adjustments, while Table 7 summarizes final results.

Observations:

Table 7. Comparison of all models on Taiwan (left) and California (right), with MAE and RMSE under both original features (OF) and edited features (EF). A dash "-" indicates the model's reference or default was not tested in that scenario

	Taiwan				**California**			
	MAE		RMSE		MAE		RMSE	
Model	OF	EF	OF	EF	OF	EF	OF	EF
Ting's Hybrid	19.743	16.871	52.205	49.024	4.723	3.687	9.290	7.598
Ho's Hybrid	23.844	21.800	55.150	54.972	**4.378**	4.284	8.627	8.505
DE-SLSTM	**17.235**	16.648	**44.902**	44.109	4.538	4.654	**8.022**	8.117
T-GCN	17.955	18.908	46.882	47.266	27.626	26.039	30.832	30.424
ATT-GRU	-	18.417	-	**42.207**	-	3.833	-	7.837
LSTM	-	16.041	-	44.128	-	3.577	-	**7.200**
XGBoost	-	**15.927**	-	46.807	-	**3.476**	-	7.239

1. *Importance of Data Editing.* Most models (Ting, Ho, DE-SLSTM) see improved accuracy after adopting our refined pipeline (EF), underscoring the high impact of data preprocessing.
2. *Base Models Often Excel.* Surprisingly, simpler base models—LSTM and XGBoost—deliver superior or comparable performance, despite the complexity of certain hybrid architectures (e.g., T-GCN). In particular, XGBoost yields the lowest mean MAE, whereas LSTM excels in RMSE.
3. *Context Matters.* T-GCN, which benefits from spatial adjacency, underperforms in California because we used data from a single freeway corridor with fewer complex interlinks. For large, well-connected networks, T-GCN might prove more advantageous.
4. *Sustainability Relevance.* When TTP is accurate, it not only enhances personal travel but also reduces wasted vehicle hours and emissions, contributing to environmental and well-being gains. Simpler models with well-prepared data may be especially useful for large-scale or resource-limited deployments (e.g., smaller city agencies).

7. Conclusion and Outlook

This work investigated how various data- and model-centric factors influence long-term travel time prediction (TTP) on freeways, with an eye toward environmental sustainability and human well-being. We compared multiple imputation methods, finding that simpler interpolation or max-value techniques can match or outperform more complex denoising autoencoders. We also demonstrated how carefully chosen temporal features (like hour-of-day) improve predictive accuracy, whereas additional signals (e.g., weather in Taiwan) may introduce noise. In the second phase, we examined seven modeling approaches—ranging from standard ensemble learners (XGBoost) and neural networks (LSTM) to hybrids (Ting's model, Ho's model, DE-SLSTM, T-GCN, ATT-GRU). Our results show that base models can rival or exceed more elaborate architectures if the data pipeline is well-structured. For practical deployment, these findings suggest that focusing effort on data cleanliness, the right feature set, and properly sized sliding windows is often more beneficial than adopting excessively complicated models.

For the future work, test the relative data-vs-model benefits on datasets from regions with markedly different traffic dynamics—such as cities with extreme weather (snowy or monsoon climates), high urban density, or rapidly developing suburbs—to assess the robustness and transferability of our "data over complexity" principle. On the other hand, given the demonstrated importance of accurate TTP for promoting sustainable mobility, we propose the following future directions for designing more accurate TTP algorithms that directly support sustainability goals:

1. **Dynamic and Context-Aware Feature Engineering for Sustainable Routing:** Future algorithms could incorporate adaptive feature selection mechanisms sensitive to real-time traffic conditions, weather patterns, and unusual events (e.g., accidents, large-scale gatherings). Such dynamic feature engineering approaches would enhance prediction reliability, enabling intelligent transportation systems to proactively alleviate congestion and lower emissions by directing traffic flows to more efficient routes during peak congestion periods or adverse weather conditions.

2. **Hybrid Attention-based Models for Eco-Friendly Commuting Decisions:** Integrating attention mechanisms (such as Transformer-based models or attention-enhanced GRU) with interpretable gradient-boosting methods could allow models to dynamically identify critical factors that most influence travel time variability. By clearly understanding these influential time intervals or external factors, city planners and travelers can make informed commuting decisions, effectively supporting eco-driving behavior, reducing fuel consumption, and lowering greenhouse gas emissions.
3. **Real-Time Spatiotemporal Graph Models for Enhanced Public Transport Efficiency:** Developing advanced graph-based neural network algorithms (e.g., enhanced Temporal Graph Convolutional Networks, T-GCN) that integrate real-time multimodal data (traffic sensors, GPS trajectories, social media incident reports, dynamic weather updates) could substantially improve prediction accuracy. Such improvements would directly support efficient scheduling and management of public transportation, reduce unnecessary idling times, and enhance reliability, thereby encouraging commuters to prefer public transit options over private vehicles.

In summary, future TTP algorithm research should explicitly integrate these predictive enhancements with sustainable mobility goals. Improved accuracy in TTP will reduce vehicle idling and emissions, optimize route planning, foster eco-friendly driving habits, and boost the reliability of public transit services. Consequently, such advancements will directly contribute to sustainable urban mobility, enhancing environmental outcomes and public well-being.

References

1. Schrank, D., Eisele, B., and Lomax, T. (2019). Urban Mobility Report 2019. *Texas A&M Transportation Institute*.
2. Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.
3. Li, Y., Zheng, Y., Zhang, H., Chen, L., Liu, Z., and Lu, X. (2015). Trend-based prediction of traffic congestion in urban areas. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 167-176.
4. Lv, Y., Duan, Y., Kang, W., Li, Z., and Wang, F. Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873.
5. J. Flum and M. Grohe, *Parameterized Complexity Theory*, Springer, New York, 2006.
6. C. Àlvarez, J. Gabarro and M. Serna, "Equilibria problems on games: Complexity versus succinctness," *Journal of Computer and System Sciences*, vol. 77, no. 6, pp. 1172–1197, 2011.
7. N. Abiri, B. Linse, P. Edén, and M. Ohlsson, "Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems," *Neurocomputing*, vol. 365, no. 6, pp. 137–146, 2019.
8. I. Ahmed, I. Kumara, V. Reshadat, A. S. M. Kayes, W. J. van den Heuvel, and D. A. Tamburri, "Travel time prediction and explanation with spatio-temporal features: A comparative study," *Electronics*, vol. 11, no. 1, article 106, 2021.
9. D. Billings and J. S. Yang, "Application of the ARIMA models to urban roadway travel time prediction-a case study," in *2006 IEEE International Conference on Systems, Man and Cybernetics (ICSMC 2006)*, vol. 3, Oct. 2006, pp. 2529–2534, doi:10.1109/ICSMC.2006.385244.
10. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

11. L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
12. C. M. Chen, C. C. Liang, and C. P. Chu, “Long-term travel time prediction using gradient boosting,” *Journal of Intelligent Transportation Systems*, vol. 24, no. 2, pp. 109–124, 2020.
13. C. H. Chou, Y. Huang, C. Y. Huang, and V. S. Tseng, “Long-term traffic time prediction using deep learning with integration of weather effect,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2019)*, vol. 3, Apr. 2019, pp. 123–135.
14. T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*, pp. 785–794, 2016, doi:10.1145/2939672.2939785.
15. Z. Chen and W. Fan, “A Freeway Travel Time Prediction Method Based on an XGBoost Model,” *Sustainability*, vol. 13, no. 15, article 8577, 2021.
16. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *arXiv preprint arXiv:1412.3555*, 2014.
17. S. Das, R. N. Kalava, K. K. Kumar, A. Kandregula, K. Suhaas, S. Bhattacharya, and N. Ganguly, “Map enhanced route travel time prediction using deep neural networks,” *arXiv preprint arXiv:1911.02623*, 2019.
18. Y. Duan, L. Yisheng, and F. Y. Wang, “Travel time prediction with LSTM neural network,” in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC 2016)*, pp. 1053–1058, 2016.
19. J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
20. Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *Proceedings of the 13th International Conference on International Conference on Machine Learning (ICML 1996)*, pp. 148–156, 1996.
21. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
22. M. C. Ho, Y. C. Chen, C. C. Hung, and H. C. Wu, “Deep ensemble learning model for long-term travel time prediction on highways,” in *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE 2021)*, pp. 129–130, 2021.
23. I. Islek and S. G. Ögüdücü, “Use of LSTM for short-term and long-term travel time prediction,” in *Proceedings of the CIKM 2018 Workshops co-located with 27th ACM International Conference on Information and Knowledge Management (CIKM 2018)*, 2018.
24. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Use of LSTM for short-term and long-term travel time prediction,” in *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS 2017)*, pp. 3146–3154, 2017.
25. W. Qiao, A. Haghani, and M. Hamedi, “Short-term travel time prediction considering the effects of weather,” *Transportation Research Record*, vol. 2308, no. 1, pp. 61–72, 2012.
26. B. Qiu and W. Fan, “Machine learning based short-term travel time prediction: Numerical results and comparative analyses,” *Sustainability*, vol. 13, no. 13, article 7454, 2021.
27. P.-Y. Ting, T. Wada, Y.-L. Chiu, M.-T. Sun, K. Sakai, W.-S. Ku, A. A.-K. Jeng, and J.-S. Hwu, “Freeway Travel Time Prediction Using Deep Hybrid Model – Taking Sun Yat-Sen Freeway as an Example,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8257–8266, 2020.
28. J.-S. Yang, “Travel time prediction using the GPS test vehicle and Kalman filtering techniques,” in *Proceedings of the 2005, American Control Conference, 2005*, pp. 2128–2133 vol. 3, 2005.
29. B. Yu, H. Wang, W. Shan, and B. Yao, “Prediction of bus travel time using random forests based on near neighbors,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 4, pp. 333–350, 2004, doi:10.1109/TVT.2020.2999358.
30. L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, “T-GCN: A temporal graph convolutional network for traffic prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.

31. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 3146–3154, 2017.
32. X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-based method with attention mechanism for travel time prediction", *Sensors*, vol. 19(4), pp. 861, 2019.
33. J.-R. Chughtai, I. U. Haq, and M. Muneeb, "An attention-based recurrent learning model for short-term travel time prediction", *PLoS One.*, vol. 17(12), pp. 1–20, 2022.
34. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need", *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, pp. 6000–6010, 2017.
35. A. Kandiri, R. Ghiasi, M. Nogal, and R. Teixeira, "Travel time prediction for an intelligent transportation system based on a data-driven feature selection method considering temporal correlation", *Transportation Engineering*, vol. 18, pp. 100272, 2024.
36. K. Fu, F. Meng, J. Ye, Z. Wang, "Compacteta: A fast inference system for travel time prediction", *The 26th ACM International Conference on Knowledge Discovery & Data Mining (SIGKDD)*, pp. 3337–3345, 2020.
37. M. Abdollahi, T. Khaleghi, K. Yang, "An integrated feature learning approach using deep learning for travel time prediction", *Expert Systems with Applications*, vol. 139, pp.112864, 2020.
38. H. Yuan, G. Li, Z. Bao, L. Feng, "Effective Travel Time Estimation: When Historical Trajectories over Road Networks Matter", *ACM SIGMOD International Conference on Management of Data*, pp. 2135–2149, 2020.
39. Y. Shen, C. Jin, J. Hua, "TTPNet: A neural network for travel time prediction based on tensor decomposition and graph embedding", *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, pp.4514–4526, 2020.
40. Boris Nikolaevich Oreshkin, Dmitri Carpov, N. Chapados, and Yoshua Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2019.
41. Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," *arXiv preprint arXiv:2012.07436*, 2021.
42. Bryan Lim, Sercan O. Arık, N. Loeff, and Tomas Pfister, "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting," *arXiv preprint arXiv:1912.09363*, 2020.
43. Haixu Wu, Jianmin Xu, Jian Wang, and Mingsheng Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *arXiv preprint arXiv:2106.13008*, 2021.
44. Slawek Smyl, "Hybrid methods and ensembles for time series forecasting," *International Journal of Forecasting*, vol. 36, no. 1, pp. 11–20, 2020.
45. Valentin Flunkert, David Salinas, and Jan Gasthaus, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *arXiv preprint arXiv:1704.04110*, 2017.
46. Huimin Han, Zehua Liu, Mauricio Barrios, Jiuhao Li, and Zhixiong Zeng, "Time Series Forecasting Model for Non-Stationary Series Pattern Extraction Using Deep Learning and GARCH Modeling," *Journal of Cloud Computing*, vol. 13, no. 3, pp. 45–61, 2024.
47. Hanjia Jiang, Leilei Song, Yu Zhang, Yue Yang, and Chao Li, "Graph Neural Networks for Traffic Forecasting: A Survey," *arXiv preprint arXiv:2007.01626*, 2021.
48. Kostas Benidis, Syama Sundar Rangapuram, and others, "Deep learning for time series forecasting: A survey," *arXiv preprint arXiv:2004.13408*, 2020.
49. John A. Miller, Mohammed Aldosari, Subas Rana, and Ninghao Liu, "A Survey of Deep Learning and Foundation Models for Time Series Forecasting," *arXiv preprint arXiv:2401.13912*, 2024.
50. U.S. Environmental Protection Agency, *Greenhouse Gas Emissions from a Typical Passenger Vehicle*, EPA-420-F-18-024, 2018. [Available: https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle]

51. U.S. Census Bureau, *Census Bureau Estimates Show Average One-Way Travel Time to Work Rises to All-Time High*, 2021. [Available: https://www.census.gov/newsroom/press-releases/2021/one-way-travel-time-to-work-rises.htm]
52. U.S. Department of Transportation, *Guidance on Treatment of the Economic Value of a Statistical Life in U.S. Department of Transportation Analyses*, January 2020.

Chuang-Chieh Lin received his Ph.D. degree in Computer Science and Information Engineering from National Chung Cheng University in 2011. He received his master degree in Computer Science and Information Engineering from National Chi Nan University in 2004, and his bachelor degree in Mathematics from National Cheng Kung University in 2002. In 2007–2008, he won the DAAD-NSC Sandwich Program Scholarship and visited RWTH Aachen University for the whole year. In 2011–2014, Chuang-Chieh served his Research and Development Alternative Service as a postdoctoral research fellow in Genomics Research Center of Academia Sinica. In 2014–2018, he joined the Computation Theory and Algorithms Group of the Institute of Information Science at Academia Sinica as a postdoctoral research fellow. In 2018–2020, he devoted himself into the fin-tech industry and focused on quantitative analysis of market micro-structures. In 2021–2024, he served as the faculty of Department of Computer Science and Information Engineering of Tamkang University. Started from August of 2024, he joined the faculty of Department of Computer Science and Engineering of National Taiwan Ocean University. His research interests include design and analysis of algorithms, game theory, machine learning, bioinformatics and quantitative finance.

Min-Chu Ho received the B.S. degrees in the Department of Management Information Systems and the Department of Foreign Languages and Literatures at National Chung Hsing University, Taiwan, in 2022. His interdisciplinary background combines technical training in data science with cross-cultural communication skills. His research interests include machine learning, deep learning, and their applications in intelligent information systems.

Chih-Chieh Hung received the Ph.D. degree in computer science from National Chiao Tung University, Taiwan, in 2010. He is currently an Associate Professor in the Department of Management Information Systems at National Chung Hsing University, Taiwan. He has held leadership positions including Chair (2024–2027) and Vice Chair (2021–2024) of the IEEE Computational Intelligence Society, Tainan Section, as well as Director and Secretary General of the Taiwan Association for Artificial Intelligence (TAAI). His research interests include data mining, spatiotemporal data analytics, intelligent transportation systems, and artificial intelligence. Dr. Hung has received several international awards, including Best Paper Awards at TAAI 2024 and 2023, the ACM Workshop on Location-Based Social Networks (2009), and a Gold Award and a Silver Award in Artificial Intelligence at the International Conference on Frontier Computing 2024 and 2025.

Received: January 26, 2025; Accepted: June 12, 2025.