Computer Science and Information Systems 20(4):1483–1501 https://doi.org/10.2298/CSIS230418059H

Heart Sounds Classification using Adaptive Wavelet Threshold and 1D LDCNN

Jianqiang Hu[1,3,⋆], Qingli Hu[2], and Mingfeng Liang[1,3]

[1] School of Computer and Information Engineering, Xiamen University of Technology,
361024 Xiamen, P. R. China
hujianqiang@tsinghua.org.cn
[2] iFlytek Research, iFlytek Co. Ltd.,
230088 Hefei, P.R. China
huqingli2014@outlook.com
[3] Key Laboratory of Internet-of-Things Applications of Fujian Province,
Xiamen University of Technology, 361024 Xiamen, P.R. China
lmfanny115@hotmail.com

Abstract. Heart sounds classification plays an important role in cardiovascular disease detection. Currently, deep learning methods for heart sound classification with heavy parameters consumption cannot be deployed in environments with limited memory and computational budgets. Besides, de-noising of heart sound signals (HSSs) can affect accuracy of heart sound classification, because erroneous removal of meaningful components may lead to heart sound distortion. In this paper, an automated heart sound classification method using adaptive wavelet threshold and 1D LDCNN (One-dimensional Lightweight Deep Convolutional Neural Network) is proposed. In this method, we exploit WT (Wavelet Transform) with an adaptive threshold to de-noise heart sound signals (HSSs). Furthermore, we utilize 1D LDCNN to realize automatic feature extraction and classification for de-noised heart sounds. Experiments on PhysioNet/CinC 2016 show that our proposed method achieves the superior classification results and excels in consumption of parameter comparing to state-of-the-art methods.

Keywords: heart sounds classification, adaptive wavelet threshold, lightweight deep convolutional neural network.

1. Introduction

Population aging is the trend of population development in the world. According to China Cardiovascular Health Index (2019), the mortality rate of residents from cardiovascular diseases (CVD) accounts for all disease mortality more than 85% of the total, and the trend is increasing [28]. Most seriously, heart disease is one of the biggest challenges of cardiovascular disease in China, and there currently are 11 million coronary heart diseases, 5 million pulmonary heart diseases, 4.5 million heart failures, 2.5 million rheumatic heart diseases, and 2 million congenital heart diseases. Heart sounds contain a large number of biomedical signals of cardiac activity. Heart sound classification is one of the most economical and effective non-invasive diagnostic methods for various cardiac abnormalities,

[⋆] Corresponding Author

and is also of great significance for primary screening and early treatment of cardiovascular diseases.

Heart sounds classification based on manual auscultation has greater uncertainty and delay in diagnosis, and it is difficult to meet the increasing patients. Since hospitals, community hospitals, nursing centers and other places are crowded with people, manual auscultation is easily affected by many factors such as the surrounding environment, the quality of the stethoscope, and the doctor's experience. Currently, digital stethoscopes based on IoT (Internet of Things) are developing rapidly. Three products of digital stethoscopes are currently available in the market: Eko Core [8], Thinklabs[35] and Hefei Huake Electronic HKY-06C [15] . Digital stethoscopes are rapidly entering family healthcare monitoring. However, automatic heart sounds classification is insufficient in terms of objectivity and effectiveness, which restricts the popularization of digital stethoscopes. Currently, automatic heart sounds classification is developing in the direction of high accuracy, lightweight deployment, and real-time response, so as to effectively support family health monitoring and clinical applications, which is becoming a research hot issue.

In general, signal preprocessing, feature extraction and classification are the mainly steps of heart sound signals (HSSs) diagnosis.

(i) In the first step, signal pre-processing includes noise removal and signal segmentation. Empirical Mode Decomposition (EMD) [1], STFT (short-time Fourier Transform)[40], Hidden Markov Model [19] and Hibernate transform, wavelet threshold de-noising method [3] were usually used to measure cardiac cycle durations and de-noise the signals. The above methods are mainly unsupervised heart sound de-noising algorithms, which needs to manually set thresholding parameters and decomposition levels. Signal segmentation plays a crucial role for feature extraction and classification. Heart sound signals is segmented into series of fundamental heart sounds (FHSs), and each FHS includes a number of the first (S1), the second (S2), systolic and diastolic hear sounds. For example, an event detection approach with deep recurrent neural networks (DRNNs) [24] was proposed for heart sound segmentation, i.e. the detection of the state-sequence first heart sound (S1)-systole-second heart sound (S2)-diastole. In order to accurately segment PCG signals, most of PCG segmentation algorithms need synchronous ECG (Electrocardiograph) as reference signals. However, it is not convenient to collect heart sounds and their reference ECG signals at the same time and ensure their synchronization in practice.

(ii) In the second step, extracted features can be divided into three major types: time-domain, frequency-domain and time-frequency domain-based features. Generally, it is relatively easy to extract the time-domain features or frequency-domain domain, but it is difficult to calculate the features in the time-frequency domain, because these features are difficult to represent discriminative features. Extracted features can also be divided into handcrafted features and deep features. Hand-crafted feature refers to extracting the discriminative features from HHSs, such as MFCC (Mel Frequency Cepstrum Coefficient), LPC (Linear Prediction Coefficient), and LPCC (Linear Prediction Cepstrum Coefficient) features. For example, STFT (Short-time Fourier Transform), Wavelet transform and S-transform method can be adopted to transform and represent signals in different time-frequency-domain. Extraction of handcrafted features still is a challenging task because of the non-stationary and diversity of heart sound signals. Besides, it is easy to be subjectively affected and produces actual deviations. Deep feature refers to extract features from HSSs through specific model which is obtained by learning and training. Owing to the strong feature representation power of deep learning technologies, deep learning has recently been used for exploratory applications in heart sounds, such as Depth Recurrent Neural network (DRNN), ShuffleNet [41], and 1D CNN (Convolutional Neural Network) [39].

(iii) In the last step, the classifier is trained over the extracted features in order to the prediction results of each heart sound signal. Various classifiers, such as Artificial Neural Network (ANN)[7], twin Support Vector Machine(tSVM)[20], and improved duration-dependent HMM, have been used to classify heart sounds. Feature extraction and classification are inseparable in deep learning-based heart sound classification. Heart sounds classification based deep convolution neural network (DCNN), such as 1D CNN , 1D DCNN [32], DS-CNN [13] , require a large scale of annotated data. Furthermore, DCNN model with heavily parameters consumption relies on high-performance GPU and parallel processing technology.

To sum up, the main challenges of heart sounds classification under are as follows: (i) The quality of heart sound is affected by the complex noise of internal physiological changes and external environmental changes. Besides, de-noising algorithms for HSSs can erroneously remove meaningful heart sound components due to manual setting of parameters, and even lead to heart sound distortion. (ii) Most of heart sound segmentation algorithms ideally assume that heart sounds are collected under strictly constrained environment. In practice, it is difficult to capture the state sequence S1-systole-S2-diastole, resulting in insufficient segmentation accuracy. (iii) The size of deep neural networks is not suitable for deployment on digital stethoscopes with strict constraints on memory and computational budget.

In this paper, we develop an automated heart sound classification method using adaptive wavelet threshold and 1D LDCNN (one-dimensional Lightweight Deep Convolutional Neural Network) with low parameters and high accuracy. In this method, wavelet transform with an adaptive threshold is used to de-noise heart sound signals. The denoised heart sound is segmented by a 3s sliding window and then fed into 1D LDCNN for automatic feature extraction and classification. Compared with several related work in heart sound classification methods, the proposed 1D LDCNN obtains the better classification performances with an accuracy of 97.92%, a sensitivity of 98.20%, an F1-score of 0.9859 and the lowest parameters consumption of 0.02M.

The key contributions of our work are as follows:

(i) We propose a wavelet transform with an adaptive threshold which de-noises the heart sound signal and avoids filtering out the approximate components of heart sound in the process of wavelet transform decompose.

(ii) We build a new 1D LDCNN which includes tem blocks, dense blocks and transition blocks. Among them, a point-wise convolution and a depth-wise separable convolution are used to effectively reduce the amount of parameters in dense blocks. The channel attention mechanism is introduced to recalibrate feature maps and further increase representation power in transition blocks.

(iii) Experiments demonstrate the superiorities of the proposed architecture with other state-of-the art CNN-based methods in terms of classification performance and parameters consumption. Besides, a heart sound acquisition system is implemented, which includes acquisition module of heart sounds and a mobile application. It deploys the proposed architecture to achieve automated heart sounds classification.

This paper is framed in different sections. Section 2 introduces related works of automatic heart sound classification. Section 3 proposes the framework of automatic heart sound classification and discusses the architecture of 1D LDCNN. Section 4 explains experimental results and deploys 1D LDCNN on mobile system. In Section 5, we draw the conclusions and discuss the future work.

2. Related Works

HSSs have periodicity, randomness and non-stationarity, and many of their features are recessive. Due to strong feature representation ability, deep learning is suitable for HSSs. Most of research works use deep neural networks and end-to-end architecture to learn and classify heart sound signals. Related works are discussed as follows:

(i) Deep neural networks are only used for heart sound segmentation. In response to the problem of insufficient utilization of cardiac cycle duration information, a Duration Long-short Term Memory network [5] was exploited to address heart sound segmentation by incorporating the duration features. Ma et al. [23] proposed a diagnosis method for congenital heart disease-related pulmonary arterial hypertension. This method first utilized a double-threshold adaptive method to segment heart sound. And then, deep learning features and time-frequency domain features were combined to form the fusion feature. Finally, XGBoost was used to classify heart sounds. Chen et al.[3] proposed a method for heart sounds classification that combined an improved frequency slice wavelet transform with CNN. This method converted 1D cardiac signal into a 2D time-frequency picture, and selected appropriate classifiers by SampEn (sample entropy) threshold to determine whether the heart sound recordings is normal. Besides, Humayun et al.[16] proposed a classification framework, consisting of a CNN with 1D CNN time-convolutional layers. In addition, representation learning was utilized to generate features. Finally, SVM and LDA (linear discriminant analysis) classifiers were exploited to classify heart sounds. Similarly, Li et al.[21] utilized convolution module to extract frequency-domain features and recurrent module to extract the time-domain features, and finally implemented heart sounds classification based on the fusion features.

(ii) Deep neural networks are only used for heart sound classification. In[26] Markov switching autoregressive model (MSAR) was exploited to segment heart sound and further a continuous-density HMM with Gaussian mixtures was utilized to classify heart sounds. Oh et al. [27] exploited deep WaveNet model to classify heart sounds, which includes fives Heart valve diseases (HVD) as follows: mitral valve prolapse (MVP), mitral stenosis (MS), mitral regurgitation (MR) and aortic stenosis (AS), normal (N). In addition, Ismail et al. [17] introduced a hybrid network-based heart sounds classification using transfer learning.

(iii) Deep neural networks are used for feature extraction and classification with end-to-end architecture. Xiao et al. [39] proposed an automatic heart sound classification method using deep learning, which includes pre-processing, heart sound classification of patches using CNN with attention mechanism, and majority voting for heart sounds classification. Raza et al.[29] depended on band filter removed the noise from HSSs, and further exploited RNN that is based on LSTM, Dropout, Dense and Softmax layer to classify heart sound recordings. Due to the low SNR, 497 features were extracted and then fed these features into the CNN, performing heart sounds classification. Ghosh et al. [12]

proposed a time-frequency-domain (TFD) deep neural network approach for automated FHSA detection using PCG signals. Xiang el al.[37] proposed a heart sound classification using two-dimensional features, which transferred heart sound classification into image classification. An end-to-end Le-LWTNe, which embedded the trainable CNN into the lifting wavelet transform (LWT), is proposed for automatic abnormality detection of heart sounds [11]. Wang et al. [36] proposed an automatic approach for heart failure typing based on heart sounds and one-dimensional CNN (1D CNN). Guo et al. [13] developed a dual-stream convolutional Neural Networks (DS-CNN) to detect abnormal from heart sound recordings.

(iv) Deep neural networks are used for feature extraction and classification (not end-to-end architecture). Shukla et al. [33] proposed an efficient method for automatic segmentation detection. Furthermore, a supervised ANN model is exploited to detect S1-S2 and non-S1-S2 segments of the cardiac cycle. Finally, a CNN model is used to automatically diagnose the heart diseases based on heart sounds. Rubin et al.[32] captures the time-frequency distribution of signal energy and classifies normal and abnormal heat maps using DCNN. A combination of WT and WPT energy-based features followed by a deep recurrent neural network (RNN) model was proposed for recognizing heart sounds [18]. Ren et al. [30] proposed deep attention-based neural networks for heart sounds classification, which exploited attention mechanisms to a CNN and an RNN to capture feature and context information.

Compared to the above methods, we can highlight the contributions of our proposed method. (i) The use of wavelet transform with an adaptive threshold is more benefit to remove noise and enhance the quality of HSSs than the other methods. In practice, it is not sufficient to only use the frequency domain filtering method, such as elliptic filter [6] and band filter, and wavelet threshold to remove the noise from HSSs. This is because that the main frequency of heart sound signal overlaps with the main frequency of the noise signal[4]. Besides, the parameters of wavelet transform in this paper are adaptive thresholds for their superior effect in the de-noising of HSSs. (ii) Deep learning methods for heart sounds classification, such as [21] [20], etc., are getting deeper and wider which bring a mass of trainable parameters and need to consume a lot of memory and computing resources. 1D LDCNN is a kind of lightweight models, which is more conducive to large-scale application of heart sounds classification.

3. Proposed Framework

In this section, we will give a detailed description about our proposed method as follows: HSSs pre-processing phase, de-noising the heart sound signal based on wavelet transform with an adaptive threshold; HSSs classification phase recognizing normal and abnormal heart sounds based on 1D LDCNN which constitutes stem block, three simplified dense blocks and transition blocks, and Softmax layer. A GAP (Global Average Pool) layer is followed by FC (Fully Connected Layer) and a Softmax. FC is usually used before the classification layer is replaced with a GAP to obtain global information about the feature map and avoid overfitting. The method increases the types of distinguishable heart sounds and improves the performance without affecting the accuracy while reducing its computational complexity.

3.1. Pre-processing of HSSs

Heart sounds have the following characteristics: (i) HSSs have periodicity, randomness and non-stationarity; (ii) HSS has obvious common characteristics and weak individual characteristics; (iii) The important information of heart sound is concentrated in the frequency of 25Hz-400Hz; (iv) The primary murmurs of heart sound span between 30Hz and 700Hz. Heart sound signals often contain noise such as lung sound and internal body noise. Therefore, effective filtration is of critical important to enhance the heart sounds signal by reducing the influence of background noise and removing spike noise. In this paper, first, the Butterworth bandpass filter is used to filter out frequencies above 400Hz and frequencies below 25Hz of heart sound signal. Butterworth bandpass filter can eliminate most of the noise signal, and reduce the calculations of subsequence wavelet transform. Then, because the main frequency of heart sound overlaps with the main frequency of the noise signal, wavelet transform is used for secondary noise eliminated. When wavelet transform decomposes HSSs, only the low-frequency part is further decomposed, and the high-frequency part, that is, the detailed part of the signal, is no longer decomposed [34]. Wavelet coefficients with relatively small amplitude values are mostly noise, while the wavelet coefficients are relatively large for the effective signal of heart sounds [42]. The threshold is set on the basis of this property. The wavelet coefficients below the selected threshold are zeroed or smoothed by threshold quantization processing to suppress the influence of high-frequency noise, while the coefficients not below the selected threshold are retained.

In general, the hard and soft threshold function method was proposed as follows: Mini-max threshold, Sqtwlolg threshold, and Rigrsure threshold[25]. Mini-max threshold is directly related to the length of HSSs. Sqtwlolg threshold is a hard threshold which the reconstructed signal after de-noising processing is very rough. When the heart sound signals are too long, the Mini-max threshold is too larger to filter out the most of the wavelet coefficients and reconstructed signal will be easily lose useful signal. Rigrsure threshold relies on Stein's unbiased risk estimate to obtain adaptive threshold of wavelet coefficients of decomposed layers. Rigrsure threshold is continuous, we first calculate the square value of each element in the signal S, and then sort from the largest to the smallest as a new sequence M =$\{M_1$,M_2,$\cdots$,$M_L\}$,and finally calculate the risk estimate for each element in M according to formula (1). Let M_k be square root of the smallest element k_0 in the risk estimate, and λ is used as the threshold.

$$R_k = \frac{L - 2k + \sum_{i=1}^{k} M_k + (L-k)\,M_{L-k}}{L} \qquad k = 1, 2, ..., L \tag{1}$$

$$\lambda{=}\sqrt{\sigma M_{k_0}} \tag{2}$$

Where L is the length of the signal, and k represents the index in corresponding to the element currently calculated.

Stein's unbiased risk estimate mainly calculates the threshold based on the variance of the high-frequency coefficients decomposed in the first layer, and then uses the threshold to process the wavelet coefficients of other layers. It does not take into account the problem of high-frequency component reduction and results in removing useful heart sound components. Wavelet decomposition is performed in accordance with the high and low frequency coefficients. For this reason, this paper introduces adaptive factor of the number of decomposition layers, and its formula is as follows:

$$\lambda_j = \frac{\lambda}{\alpha_j} \tag{3}$$

$$\alpha_j = \ln\left(j + \frac{j}{J}\right) \tag{4}$$

where j is the threshold of the $j - th$ layer, J is the number of decomposition layers, α_j denotes the adaptive factor, and λ_j is related to the number of layers of wavelet decomposition. The threshold increases and decreases with the number of layers. When we calculate the high-frequency coefficient, a larger threshold can be obtained. When we calculate the low-frequency coefficient, a slight smaller threshold can be obtained. The detailed algorithm of de-noising is shown as Algorithm 1.

Algorithm 1: Heart sound signals de-noising based on wavelet transform with an adaptive threshold

Input: Heart sound signals
Output: The reconstructed heart sound signal

1. Using the Butterworth bandpass filter to filter out frequencies above 400Hz and frequencies below 25Hz of heart sound signal.
2. Using wavelet transform to further eliminate noise.
3. Selecting the appropriate wavelet function to suppress the influence of high frequency noise.
4. Calculating the Stein's unbiased risk estimate for each element in sequence according to formula (1).
5. Introducing the layer number adaptive factor on the original basis, according to fomula (4).
6. The processed wavelet coefficients are inversely transformed to obtain the final de-noised heart sound signal.
7. Return the reconstructing of heart sound signal.

Heart sound signals de-noising based on wavelet transform with an adaptive threshold has some advantages as follows:

(i) The effect of wavelet transform with hard threshold remains rough because it neglects to processes wavelet coefficients larger than hard threshold and results in de-noising distortion. Wavelet transform with an adaptive threshold compensates for the deficiency of hard threshold by taking into account high-frequency and low-frequency coefficients.

(ii) The effect of wavelet transform with rigrsure threshold improves smooth because it performs continuous compression on wavelet coefficients and results in filtering out the approximate components of heart sound signals. Wavelet transform with an adaptive threshold in this paper retains a large coefficient and avoids de-noising distortion in the process of wavelet decomposition.

3.2. An Architecture of 1D LDCNN

In order to reduce extra computing of raw input signals, we employ the sliding windows to split the Heart sound recordings into a series of patches with fixed length, i.e., 3s length and 1s stride (from empirically selected). On one hand, segmenting heart sound recordings into FHSs accurately is very difficult. On the other hand, sliding windows can extends the scale of training set. This method does not need to extract features in advance, so as to avoid the loss of features due to improper design and affect the classification effect. We construct a 1D LDCNN to automatically learn the discriminative features of heart sounds. The network first uses the stem block to enhance the characterization ability of features, and then uses simplified dense blocks and transition blocks to extract deep features, and finally uses softmax as the heart sound classifier. In particular, deep separable convolution in dense blocks can reduce the amount of network parameters. Furthermore, channel attention in transition blocks highlights the channel features with high contribution. The detail architecture of 1D LDCNN is as follows:

(i) Stem Block

We design stem block which can be effectively increase representation power while increasing a small amount of computational cost. At the beginning, the first convolutional layer uses a 1×3 kernel size, stride 2, followed by batch normalization (BN) and Rectified Linear Unit (ReLU), the output feature map can be obtained. In order to enhance the richness of features, we use a 2-way convolutional layer to get different scales of receptive fields. One way of the layer uses a Cov 1×1 , stride 1 and Cov 1×3 , stride 2, followed by BN and ReLU, respectively. The other way of the layers uses max-pooling 1×2, stride 2. The output feature map G_2 and G_3 can be obtained respectively. Finally, in order to finally connect in the channel dimension, we use convolution to compress the amount of channels to 24, and the output feature map G_4 can be obtained. In order to reduce the computational complexity, the maximum pooling compression feature dimension with a step size of 2 is used to obtain the final output feature map. The calculation of the entire stem block is as follows

$$G_1 = F\left[\sum_{i=1}^{24}(W_{1\times 3_i} \times X_{1\times w})\right] \tag{5}$$

$$G_2 = F\left\{\sum_{j=1}^{24}\left[W_{1\times 3_j} \times F\left[\sum_{i=1}^{12}(W_{1\times 1_i} \times G_1)\right]\right]\right\} \tag{6}$$

$$G_3 = M\text{axPool}(G_1)_{1\times 2} \tag{7}$$

$$G_4 = F\left\{\sum_{i=1}^{24}\left[W_{1\times 1_i} \times \text{Cat}\left(G_2, G_3\right)\right]\right\} \tag{8}$$

$$G_{out} = M\text{axPool}[G_4]_{1\times 3} \tag{9}$$

$$F = \text{Re}LU\left[BN\left(\cdot\right)\right] \tag{10}$$

where $X_{1\times w}$ indicates the input of stem block; w is the dimension of input data; $W_{1\times 3}$, $W_{1\times 1}$ are the 1×3 convolutional kernel and 1×1 convolutional kernel, respectively;

$Cat()$ denotes concatenation operation; $MaxPool()$ denotes the max-pooling neural network operation.

(ii) Simplified dense block

Dense concatenation is an important feature reuse in DenseNet. Since DenseNet allows all previous feature maps are used to as input to the subsequent layer of the network. This method can make the network simple, but it also causes the memory access cost to increase quadratically with network depth and in turn leads to computation cost. In order to reduce the amount of original dense connections, each layer of the network is directly connected to all previous layers with only retaining the reuse of low-level features, thereby reducing redundant connections. In the simplified dense network, the output of each block is divided into two parts: one part is used for the input of the next block to extract higher-level features; the other part is used for the final concatenation operation, so that low-level features can be obtained so as to improve feature expression ability. The formula of a simplified dense block is as follows

$$G_{dense} = cat\left(H_1, H_2, \cdots, H_k\right) \tag{11}$$

where G_{dense} denotes output of a simplified dense block, $H_k(k > 1)$ is a composite function which includes a series of operations, i.e., BN, H_SWISH, Dropout and convolutional layers.

In order to further lower the parameter consumption of simplified dense block, a separable convolution is utilized to extract features. DWconv can effectively reduce the amount of parameters and computation cost by separating spatial features and channel features. Besides, the H-Swish activation function is used instead of the ReLU activation function to further reduce the computation cost. The formula of H_k is as follows

$$H_k = \begin{cases} J\left\{S\left[J\left[\sum\limits_{i=1}^{C}\left(W_{1\times 1_i} \times G_{k-1}\right)\right]\right]\right\}, stride = 1, \\ cat\left(J\left[S\left(G_{k-1}\right)\right], J\left\{S\left[J\left[\sum\limits_{i=1}^{C}\left(W_{1\times 1_i} \times G_{k-1}\right)\right]\right]\right\}\right), stride = 2. \end{cases} \tag{12}$$

$$J = H\_SWISH\left[BN\left(\cdot\right)\right] \tag{13}$$

where G_{k-1} denotes the output of dense block, $G_{k-1} = G_{out}(k = 1)$; i denotes $i - th$ convolution kernel; C denotes the current number of convolution kernels; S denotes the depth-wise separable convolution.

(iii) Transition block

The output G_{dense} of simplified dense block can be seen that the amount of output channels is very high and cannot be directly used as the input of the next dense module. Transition block is used to squeeze the dimensionality of the output map of simplified dense block. In order to improve representation power, attention mechanism module is introduced into transition blocks. The transition blocks with pooling layers are introduced to divide the networks into two blocks processing feature maps at different resolutions. Using a convolution 1×1 instead of fully connected layer, the number of output channels is a half of the original dense block. In the attention mechanism module, it exploits global maximum pooling and global average pooling (GAP) to simultaneously generate average-pooled features and max-pooled features respectively. And then, both features are forward to a shared two-layer convolution, where the amount of output channels of the first convolution is a half of that of the second convolution. The two output features are added and the corresponding weights are obtained through Sigmoid. In short, the channel attention is computed as

$$G_{att} = \sigma \left\{ Conv \left[AvgPool \left(G'_{dense} \right) \right] + Conv \left[MaxPool \left(G'_{dense} \right) \right] \right\} \tag{14}$$

where σ denotes the sigmoid activation function; G'_{dense} represents the two-layer convolutions; $AvgPool(G'_{dense})$ and $MaxPool(G'_{dense})$ represent the global average pooling calculation and global maximum pooling calculation, respectively.

3.3. Heart sound classification

Heart sounds classification is the last step of the model. Heart sounds are mainly divided into two categories: Normal and Abnormal, with 0 for Normal and 1 for Abnormal. The GAP layer averages the features extracted by the network and maps the features to 2 channel dimensions through the FC (fully connected) layer, and finally uses softmax layer to calculate the probabilities of the 2 channels, and takes the corresponding index of the maximum probability as the final output of the network. Softmax is an effective way that handles multi-class classification problems in which output represents in categorical ways. The activation function of softmax is defined as

$$S(y) = \frac{e^{y_i}}{\sum_{k=1}^{K} e^{y_k}} \tag{15}$$

where y, S are the input and the output, respectively. The Softmax function is used in the last layer of the neural network to obtain the probabilities of the category class of each input.

4. Experimental Setup and Analysis

4.1. Experimental Setting

We conduct the experiments on publicly available heart sound dataset which provided by PhysioNet/CinC Heart Sound Classification Challenge held in 2016. The heart sound records of PhysioNet/CinC 2016 data set [22] are collected from different clinical and non-clinical real environments, including clean and noisy heart sound records. The targets of its collection are both healthy subjects and pathological patients, including children and adults. The database includes six sub-data sets a-b-c-d-f, which are integrated from data sets provided by different research organizations such as MIT, AAD, AUTH, TUT, UHA, etc., and are strictly labeled to divide the data into normal and abnormal. There are two types of normal heart sound records from healthy subjects, and abnormal records from pathological patients who have been diagnosed with heart disease. There are a total of 3240 heart sound records from 764 subjects. The shortest record is only 5s and the longest lasts over 120s. This experiment exploits python and PyTorch (Deep Learning Framework) to build the proposed network. The training environment of the networks as follows: CPU (Intel i5-10400F), GPU (Nvidia RTX 2070Super), and 8GB video memory. The operating system is Ubuntu 18.04.

4.2. Evaluation Metrics

In order to compare with state-of-the art heart sound classification method, we choose Accuracy (Acc), Sensitivity (Se), Specificity (SP), Precision (Pr), F_1 score and $MAcc$ as several evaluation metrics, which are defined as follows

$$Acc : \frac{TP + TN}{TP + FP + TN + FN} \tag{16}$$

$$Se : \frac{TP}{FP + TN} \tag{17}$$

$$Sp : \frac{TN}{FP + TN} \tag{18}$$

$$Pr : \frac{TP}{TP + FP} \tag{19}$$

$$F_1 : 2 \times \frac{Se \times \mathrm{Pr}}{Se + \mathrm{Pr}} \tag{20}$$

$$MAcc : \frac{Se + Sp}{2} \tag{21}$$

where TP, TN, FP, FN denotes true-positive, true-negative, false-positive and false-negative.

4.3. Pre-Processing of HSSs

The a0074 record is de-noised by a 5-layer db6 wavelet decomposition. The choice of threshold rules will have a certain effect on the noise reduction effect. Fig. 1 shows the noise reduction effect of wavelet transform with Rigrsure threshold of a0074 record. Although the noise is basically eliminated, some meaningful components of heart sound signal heart are also excessively eliminated. Fig. 2 shows the noise reduction effect of wavelet transform with an adaptive threshold. Compared to Fig. 1, more meaningful components are retained in Fig. 2.

4.4. Classification Effect of 1D LDCNN

In order to fully learn the potential features of the provided dataset, we need train the proposed architecture of 1D LDCNN as much as possible. Therefore, we can divide the provided dataset (PhysioNet/CinC 2016) into three parts, a training set/a validation set/a test set with a percentage of 8:1:1 respectively. The training set is used to train the model, the validation set to optimize the model and the test set to evaluate and check the performance of the model. Our proposed model is trained with a batch size of 64. The WCE (Weighted cross-entropy) with rate 1 to 0.25 (Abnormal to Normal) is chosen as loss function. Adam optimizer function uses momentum and adaptive learning rates to converge faster.

Table 1 summarizes the experimental results. Our proposed method is mainly compared with five state-of-art methods. It can be seen that the proposed method outperforms other methods except Sp. In terms of Acc, Se and Pr, our proposed method obtains an

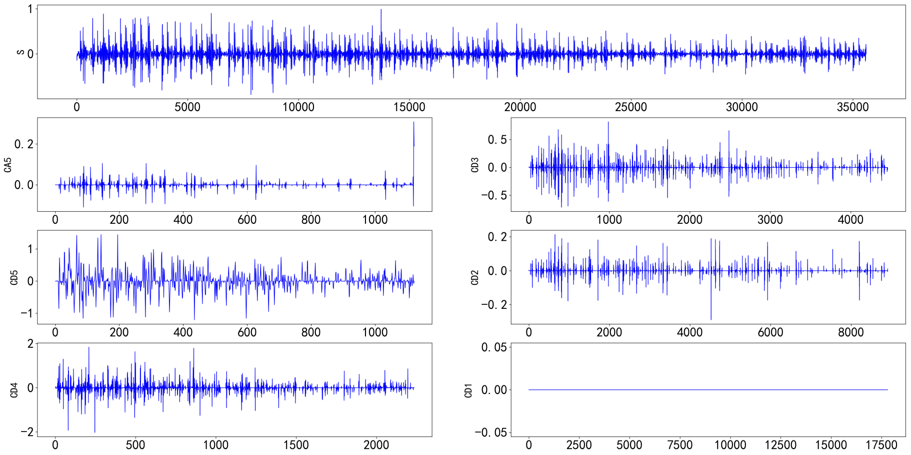


Fig. 1. The de-noising effect of 5-layer db6 wavelet decomposition with Rigrsure threshold

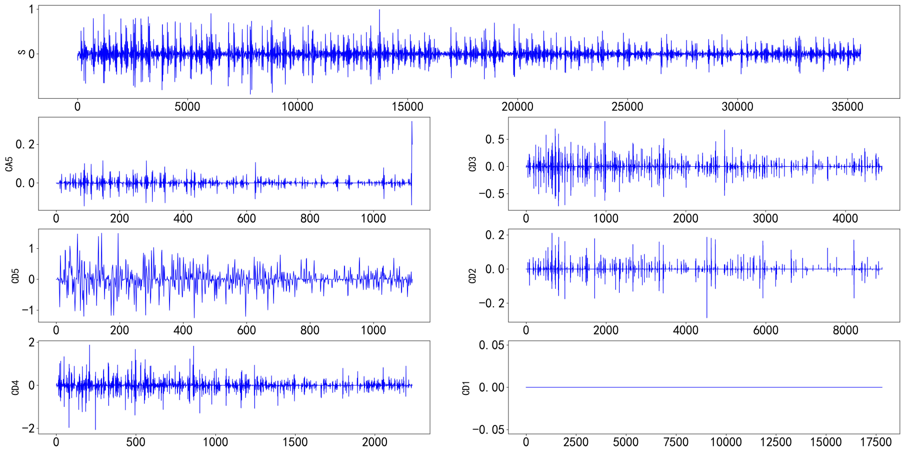


Fig. 2. The de-noising effect of 5-layer db6 wavelet decomposition with an adaptive threshold

accuracy of 97.92%, a sensitivity of 98.20% and a precision of 98.98%. It obtains the first place. The effect of MFCC-CNN [31] method is obtained with an accuracy of 93.31%, a specificity of 95.16%, and a sensitivity of 82.66%, which uses MFCC-based features to transform 1D to 2D time-frequency heat map, and exploits MFCC heart maps using CNN to classify heart sounds recordings. In addition, a modified AlexNet model [10] achieves an accuracy of 97.05%, a specificity of 93.20% and a sensitivity of 95.12%. A cross-wavelet assisted AlexNet model [9] obtains an accuracy of 97.89%, a specificity of 97.12% and an F_1 score of 0.9421, which exploits thresholding-based wavelet transform to remove noise, and convolution neural network (Alex Net architecture) recognizes abnormal/normal PCG signals. Notably, 1D Clique [39], 1D Dense [38] and 1D LDCNN, all the three focus on feature reuse and parameter efficiency to automatic heart sounds classification. Compared with 1D Clique and 1D Dense, 1D LDCNN provides the best sensitivity (98.20% vs. 86.21% vs. 85.29%), which indicates that the positive feature of 1D LDCNN is to avoid missed diagnosis as much as possible. Besides, 1D LDCNN provides the lowest specificity (92.22% vs. 95.16% vs. 95.73%), which demonstrates that 1D Clique and 1D Dense overemphasize to avoid misdiagnosis at the expense of missed diagnosis. Table 3 also shows that only our proposed method and cross-wavelet assisted AlexNet employ WT-HSEGAN and wavelet transform for de-noising, respectively.

In addition, 1D LDCNN, 1D Clique and 1D Dense all put raw heart sound data into the network to automatically extract features and perform classification. Among them, the model sizes, namely the trainable parameters (Params) are 0.02M, 0.19M, and 0.11M, respectively. And interestingly, model size of 1D LDCNN without deep separable convolution is 0.023M, i.e., deep separable convolution can further reduce parameters by 0.003M, which is more conducive to use in resource-constrained terminals.

Table 1. Evaluation results for the proposed method in comparisons with state-of-art methods

Methods	$Acc(\%)$	$Se(\%)$	$Pr(\%)$	$Sp(\%)$	F_1	$Parms$(M)
MFCC-CNN	93.31	82.66	95.38	95.16	0.8857	-
Modified AlexNet	97.05	95.12	-	93.20	-	-
1D Dense	93.56	85.29	96.09	95.73	0.9037	0.11
1D Clique	93.28	86.21	96.27	95.16	0.9096	0.19
Cross-wavelet assisted AlexNet	97.89	97.12	-	-	0.9421	-
1D LDCNN	97.92	98.20	98.98	92.22	0.9859	0.02

In order to verify the effect of different modules on the improvement performance of proposed model, this paper constructs a basic DCNN model in which the stem module and separable convolution are replaced by conventional convolution operations with convolution kernel sizes of 7 and 3, respectively. The attention mechanism is removed from the transition module. Then, stem module, channel attention mechanism, and separable convolutions are added to the basic DCNN model, and the network structure used by each model remains the same. Finally, under the same data set conditions, the network is trained and tested, and the final results are shown in Table 2. It can be observed that adding each module in sequence has a relatively obvious improvement effect on the model, and can obtain a high F_1 score. The deep neural network model

$(DCNN + stem\_block + attention + DWconv)$ uses the stem structure to enhance the model's initial feature presentation ability for original heart sound data, and reuses low-level features in subsequent dense modules to further improve the network. The feature presentation ability of the transition module introduces the channel attention mechanism to highlight the channel features with large contributions, which makes the extracted features more distinguishable, and is more conducive to the classification and recognition of heart sounds.

Table 2. Performance comparison of different modules on the performance of the model

Methods	Acc(%)	Se(%)	Pr(%)	Sp(%)	F_1	MAcc
DCNN	89.13	95.27	91.26	66.77	0.9322	0.8102
DCNN+stem_block	91.76	94.54	94.93	81.60	0.9473	0.8807
DCNN+stem_block+attention	94.89	96.74	96.74	88.13	0.9674	0.9244
DCNN+stem_block+attention+DWconv	97.70	98.20	98.98	92.22	0.9859	0.9521

4.5. A Real-Time Heart Sound Detection System

The real-time heart sound detection system is developed by our group, which includes acquisition module of heart sounds and a mobile application. In Fig. 3, acquisition module of heart sounds consists of a transducer (acquisition probe), microcontroller, analog signal processing, audio AD module, power amplifier, and communication module. The sensitivity of a transducer is $-36db \pm 3d$. The acoustic vibration generated during the cardiac activity is output through the transducer, amplifier circuit, detection circuit and serial port in the form of vibration wave.

The following points should be noted:

(i) During the heart sound acquisition, the patient should keep the probe relatively still, and try to avoid holding the acquisition probe with hands.

(ii) The patient collects heart sounds as far as possible in a temperature-friendly and quiet environment, keeping relaxed and breathing evenly. The mobile application is used to display real-time HSSs, preprocess and classify HSSs transmitted from via serial port.

We deploy 1D LDCNN on a resource constrained device with CUP (Qualcomm Snapdragon 865), GUP (Adreno 650), Memory (12GB) and 256GB storage capacity[14]. The deployment process of 1D LDCNN is as follows:

(i) Convert 1D LDCNN to an ONNX model, and further convert an ONNX model to a NCNN model.

(ii) Build an Android application package (APK) using a NCNN model.

(iii) Migrate the APK to mobile phone for installation and operation. The deployed mobile application is relatively simple, mainly including three buttons of preprocessing, de-noising and classification.

Fig. 4 shows two screenshots of heart sounds classification in a mobile phone. In Fig. 4(a), the normal heart sound has an SNR of 20.358465dB and a positive predictive value (PPV) of 31.8356%. Correspondingly, the abnormal heart sound has an SNR of 19.945709dB and a PPV of 99.96306% in Fig. 4(b). The results verify the feasibility of deployment.

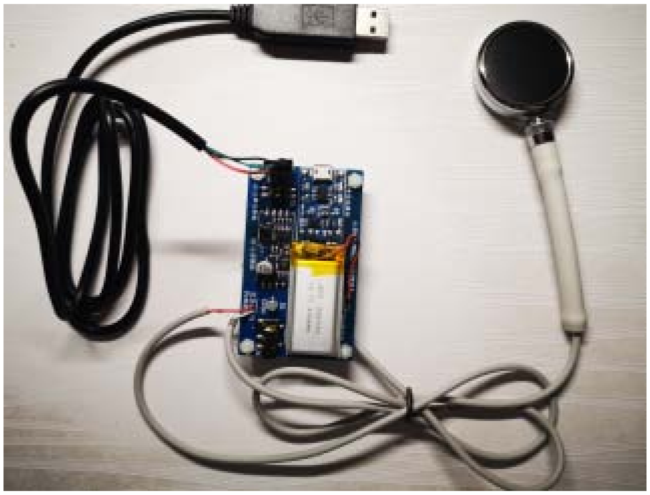


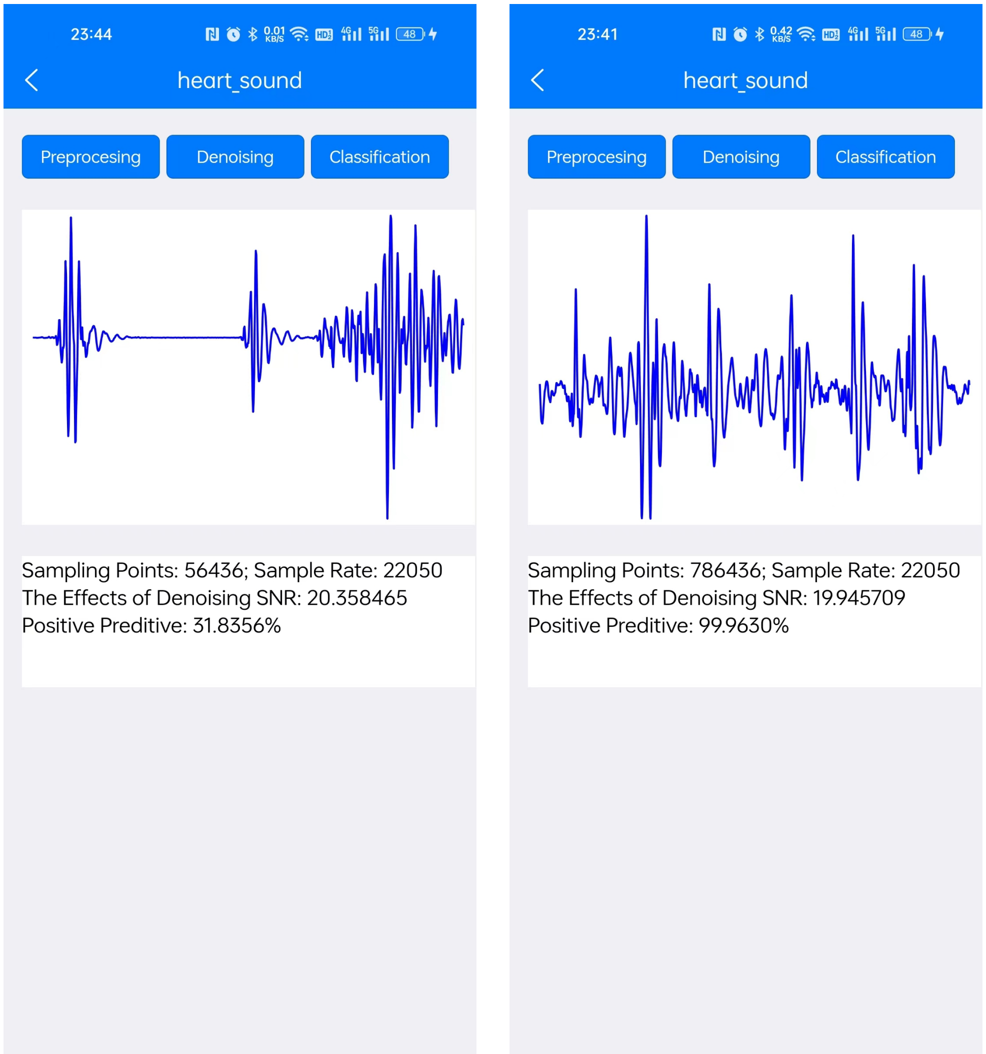
Fig. 3. Acquisition module of heart sounds

5. Conclusions

In this paper, a novel heart sound classification method using adaptive wavelet threshold and 1D LDCNN is proposed. Taking advantages of wavelet transform with an adaptive threshold, the noise of HSSs can be effectively removed. More importantly, it can avoid filtering out meaningful component in the process of wavelet transform decomposition. Furthermore, 1D LDCNN is exploited to realize the automatic feature extraction and final classification, which uses simplified dense blocks and attention mechanism to reduce parameters. Experiment results on PhysioNet/CinC 2016 Challenge database show that our proposed method achieves better performances in terms of classification performance and parameters consumption. To a certain extent, easy lightweight deployment of the proposed method also promotes the application of digital stethoscopes in unconstrained environment.

In our future works, we will explore more efficient de-noising method for heart murmurs and environments noises. We will take into consider to build more efficient architecture suitable for deployment in resource-constrained terminals. Additionally, 1D LDCNN effectively captures hidden patterns of HSSs in Euclidean space, but we hope to achieve better prediction results using graph neural network (GNN) [2][43], provided that sufficient training data is available.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant No. 62173285; Natural Science Foundation of Fujian Province under Grant No.2023J011426; The New Generation Information Technology Innovation Project of China University Industry-University-Research Innovation Fund under Grant No. 2020ITA03015; Research Climbing Program of Xiamen University of Technology under Grant No. 2020044.



Sampling Points: 56436; Sample Rate: 22050
The Effects of Denoising SNR: 20.358465
Positive Preditive: 31.8356%

Sampling Points: 786436; Sample Rate: 22050
The Effects of Denoising SNR: 19.945709
Positive Preditive: 99.9630%

(a) Normal

(b) Abnormal

Fig. 4. Screenshot of heart sounds classification in a mobile phone.

References

1. Altuve, M., Suárez, L., Ardila, J.: Fundamental heart sounds analysis using improved complete ensemble emd with adaptive noise. Biocybernetics and Biomedical Engineering 40(1), 426–439 (2020)
2. Cen, Y., Hou, Z., Wang, Y., Chen, Q., Luo, Y., Yao, X., Zeng, A., Guo, S., Yang, Y., Zhang, P., et al.: Cogdl: toolkit for deep learning on graphs. arXiv preprint arXiv:2103.00959 (2021)
3. Chen, P., Zhang, Q.: Classification of heart sounds using discrete time-frequency energy feature based on s transform and the wavelet threshold denoising. Biomedical Signal Processing and Control 57, 101684 (2020)
4. Chen, P., Zhang, Q.: Classification of heart sounds using discrete time-frequency energy feature based on s transform and the wavelet threshold denoising. Biomed. Signal Process. Control. 57 (2020)
5. Chen, Y., Lv, J., Sun, Y., Jia, B.: Heart sound segmentation via duration long–short term memory neural network. Applied Soft Computing 95, 106540 (2020)
6. Coskun, H., Deperlıoğlu, Ö., Yığıt, T.: Classification of extrasystole heart sounds with mfcc features by using artificial neural network. In: 2017 25th Signal Processing and Communications Applications Conference (SIU). pp. 1–4. IEEE (2017)
7. Das, S., Pal, S., Mitra, M.: Supervised model for cochleagram feature based fundamental heart sound identification. Biomedical Signal Processing and Control 52, 32–40 (2019)
8. Devices, E.: Eko core digital stethoscope, [Online]. Available: https://ekodevices.com/(current June 2023)
9. Dhar, P., Dutta, S., Mukherjee, V.: Cross-wavelet assisted convolution neural network (alexnet) approach for phonocardiogram signals classification. Biomedical Signal Processing and Control 63, 102142 (2021)
10. Dominguez-Morales, J.P., Jimenez-Fernandez, A.F., Dominguez-Morales, M.J., Jimenez-Moreno, G.: Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. IEEE transactions on biomedical circuits and systems 12(1), 24–34 (2017)
11. Fan, J., Tang, S., Duan, H., Bi, X., Xiao, B., Li, W., Gao, X.: Le-lwtnet: A learnable lifting wavelet convolutional neural network for heart sound abnormality detection. IEEE Transactions on Instrumentation and Measurement (2023)
12. Ghosh, S.K., Ponnalagu, R., Tripathy, R.K., Panda, G., Pachori, R.B.: Automated heart sound activity detection from pcg signal using time–frequency-domain deep neural network. IEEE Transactions on Instrumentation and Measurement 71, 1–10 (2022)
13. Guo, Z., Chen, J., He, T., Wang, W., Abbas, H., Lv, Z.: Ds-cnn: Dual-stream convolutional neural networks based heart sound classification for wearable devices. IEEE Transactions on Consumer Electronics (2023)
14. Hu, J., Wu, K., Liang, W.: An ipv6-based framework for fog-assisted healthcare monitoring. Advances in Mechanical Engineering 11(1), 1687814018819515 (2019)
15. Huake, H.: Hky-06c heart sound sensor, [Online]. Available: http://www.hfhuake.com/(current June 2023)
16. Humayun, A.I., Khan, M., Ghaffarzadegan, S., Feng, Z., Hasan, T., et al.: An ensemble of transfer, semi-supervised and supervised learning methods for pathological heart sound classification. arXiv preprint arXiv:1806.06506 (2018)
17. Ismail, S., Ismail, B., Siddiqi, I., Akram, U.: Pcg classification through spectrogram using transfer learning. Biomedical Signal Processing and Control 79, 104075 (2023)
18. Karhade, J., Dash, S., Ghosh, S.K., Dash, D.K., Tripathy, R.K.: Time–frequency-domain deep learning framework for the automated detection of heart valve disorders using pcg signals. IEEE Transactions on Instrumentation and Measurement 71, 1–11 (2022)

19. Kui, H., Pan, J., Zong, R., Yang, H., Wang, W.: Heart sound classification based on log mel-frequency spectral coefficients features and convolutional neural networks. Biomedical Signal Processing and Control 69, 102893 (2021)
20. Li, J., Ke, L., Du, Q.: Classification of heart sounds based on the wavelet fractal and twin support vector machine. Entropy 21(5), 472 (2019)
21. Li, S., Li, F., Tang, S., Luo, F.: Heart sounds classification based on feature fusion using lightweight neural networks. IEEE Transactions on Instrumentation and Measurement 70, 1–9 (2021)
22. Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I., Johnson, A.E., et al.: An open access database for the evaluation of heart sound algorithms. Physiological measurement 37(12), 2181 (2016)
23. Ma, P., Ge, B., Yang, H., Guo, T., Pan, J., Wang, W.: Application of time-frequency domain and deep learning fusion feature in non-invasive diagnosis of congenital heart disease-related pulmonary arterial hypertension. MethodsX p. 102032 (2023)
24. Messner, E., Zöhrer, M., Pernkopf, F.: Heart sound segmentation-an event detection approach using deep recurrent neural networks. IEEE transactions on biomedical engineering 65(9), 1964–1974 (2018)
25. Naing, H., Hidayat, R., Hartanto, R., Miyanaga, Y.: Discrete wavelet denoising into mfcc for noise suppressive in automatic speech recognition system. International Journal of Intelligent Engineering and Systems 13(2), 74–82 (2020)
26. Noman, F., Salleh, S.H., Ting, C.M., Samdin, S.B., Ombao, H., Hussain, H.: A markov-switching model approach to heart sound segmentation and classification. IEEE Journal of Biomedical and Health Informatics 24(3), 705–716 (2019)
27. Oh, S.L., Jahmunah, V., Ooi, C.P., Tan, R.S., Ciaccio, E.J., Yamakawa, T., Tanabe, M., Kobayashi, M., Acharya, U.R.: Classification of heart sound signals using a novel deep wavenet model. Computer Methods and Programs in Biomedicine 196, 105604 (2020)
28. Organization, W.H., et al.: World health statistics overview 2019: monitoring health for the sdgs, sustainable development goals. Tech. rep., World Health Organization (2019)
29. Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.S., On, B.W.: Heartbeat sound signal classification using deep learning. Sensors 19(21), 4819 (2019)
30. Ren, Z., Qian, K., Dong, F., Dai, Z., Nejdl, W., Yamamoto, Y., Schuller, B.W.: Deep attention-based neural networks for explainable heart sound classification. Machine Learning with Applications 9, 100322 (2022)
31. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., Sricharan, K.: Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In: 2016 Computing in cardiology conference (CinC). pp. 813–816. IEEE (2016)
32. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., Sricharan, K.: Recognizing abnormal heart sounds using deep learning. arXiv preprint arXiv:1707.04642 (2017)
33. Shukla, S., Singh, S.K., Mitra, D.: An efficient heart sound segmentation approach using kurtosis and zero frequency filter features. Biomedical Signal Processing and Control 57, 101762 (2020)
34. Taranenko, Y.K.: Efficiency of using wavelet transforms for filtering noise in the signals of measuring transducers. Measurement Techniques 64(2), 94–99 (2021)
35. Thompson, J.: Thinklabs digital stethoscopes, electronic stethoscope systems (2013)
36. Wang, H., Guo, X., Zheng, Y., Yang, Y.: An automatic approach for heart failure typing based on heart sounds and convolutional recurrent neural networks. Physical and Engineering Sciences in Medicine 45(2), 475–485 (2022)
37. Xiang, M., Zang, J., Wang, J., Wang, H., Zhou, C., Bi, R., Zhang, Z., Xue, C.: Research of heart sound classification using two-dimensional features. Biomedical Signal Processing and Control 79, 104190 (2023)

38. Xiao, B., Xu, Y., Bi, X., Li, W., Ma, Z., Zhang, J., Ma, X.: Follow the sound of children's heart: a deep-learning-based computer-aided pediatric chds diagnosis system. IEEE Internet of Things Journal 7(3), 1994–2004 (2019)
39. Xiao, B., Xu, Y., Bi, X., Zhang, J., Ma, X.: Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption. Neurocomputing 392, 153–159 (2020)
40. Yang, Y., Guo, X.M., Wang, H., Zheng, Y.N.: Deep learning-based heart sound analysis for left ventricular diastolic dysfunction diagnosis. Diagnostics 11(12), 2349 (2021)
41. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6848–6856 (2018)
42. Zhang, Y., Ding, W., Pan, Z., Qin, J.: Improved wavelet threshold for image de-noising. Frontiers in neuroscience 13, 39 (2019)
43. Zhao, J., Dong, Y., Ding, M., Kharlamov, E., Tang, J.: Adaptive diffusion in graph neural networks. Advances in Neural Information Processing Systems 34, 23321–23333 (2021)

Jianqiang Hu is an associate professor in school of computer and information engineering, Xiamen University of Technology, China. He once worked as a postdoctoral researcher at Tsinghua University. He received his Ph.D. degree in computer science and engineering from National University of Defense Technology, China, in 2005. He is the author of more than 60 articles, and more than 8 inventions. His current research interests include Edge Computing, Biomedical Signal Processing, and Big Data Analytics.

Qingli Hu is a senior researcher at iFlytek Research, iFlytek Co. Ltd.. He received B.S. and M.S. degrees from Anhui Jianzhu University and Xiamen University of Technology, China, in 2017 and 2021, respectively. His current research interests include Speech Enhancement and Big Data Analytics.

Mingfeng Liang is a master student at school of computer and information engineering, Xiamen University of Technology, China. She received her B.S. degree from Shenzhen University in 2019. Her interest include Biomedical Signal Processing and Big Data Analytics.

Received: April 18, 2023; Accepted: July 15, 2023.