

An Optimal Approach for Heart Sound Classification Using Artificial Neural Network

Yunendah Nur Fu'adah

Kumoh National Institute of Technology

Ki Moo Lim (✉ kmlim@kumoh.ac.kr)

Kumoh National Institute of Technology

Research Article

Keywords: Heart sound auscultation, optimal approach, artificial neural network, heart sound classification, Cardiovascular disease

Posted Date: July 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-754775/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Heart sound auscultation is one of the most widely used approaches for detecting cardiovascular disorders. Diagnosing abnormalities of heart sound using a stethoscope depends on the physician's skill and judgement. Several studies have shown promising results in the automatic detection of cardiovascular disorders based on heart sound signals. However, the accuracy performance needs to be improved as automated heart sound classification aids in the early detection and prevention of the dangerous effects of cardiovascular problems. In this study, an optimal heart sound classification method based on machine learning technologies for cardiovascular disease prediction is performed. It consists of three steps: pre-processing that sets the 5 s duration of the Physionet Challenge 2016 datasets, feature extraction using mel-frequency cepstrum coefficients (MFCC), and classification using an artificial neural network (ANN) with one hidden layer that provides low parameter consumption. Ten-fold cross-validation was used to evaluate the performance of the proposed method. The best model obtained 94% accuracy and 93% AUC score, which were assessed using 1626 test datasets. Taken together, the results show that the proposed method obtained excellent classification results and provided low parameter consumption, thereby reducing computational time to facilitate a real-time implementation.

Introduction

Cardiovascular disease is one of the main causes of mortality worldwide. In 2016, an estimated 17.9 million deaths occurred prematurely due to cardiovascular disease and accounts for 31% of all global deaths. Heart attacks and strokes are responsible for 85% of global deaths due to cardiovascular disease [1]. The high mortality rate is caused by cardiovascular problems that should be diagnosed early to avoid long-term complications and premature cardiac death.

Electrocardiograms (ECGs) and heart sound or phonocardiograms (PCGs) are commonly used to diagnose cardiovascular disease. PCG, which is a graphical representation of heart sound signals, can extract the heart valve opening time more accurate than ECG signals [2]. As a result, heart sound signals contain important physiological information of cardiac conditions that can be utilized to detect cardiac organ deformation as well as valve damage [3]. On the other hand, cardiac auscultation is determined by the physicians' abilities and subjective experiences [4]. Therefore, an objective and automatic computer-assisted analysis of heart sound signals is very important for the early diagnosis of cardiovascular diseases [5] which can potentially prevent premature death.

Automatic heart sound classification is currently a promising research field based on signal processing and artificial intelligence approaches. It is reliable to screen or monitor for cardiac diseases in a wide range of clinical settings [6], allowing for the reduction of costly and laborious manual examinations. Several studies have proposed algorithms to detect cardiovascular disorders based on heart sound signals.

Previously, Rubin et al. reported an 84% test accuracy using mel-frequency cepstral coefficients (MFCC) and two-dimensional convolutional neural network (2D CNN) [7]. In the preprocessing step, a 3 s duration of heart sound signals was selected. In addition, they extracted 13 MFCC features, and converted these into 2D heat map images as input for the 2D CNN. Nogueira et al. used 2D heat maps images as input for the support vector machine (SVM) classifier algorithm. This method achieved an accuracy of 82.33 % [8].

Meanwhile, in a study by Xiao et al. a validation accuracy of 93% was reported and verified via 10-fold cross validation using a 1D convolutional neural network (1D CNN) [9]. The preprocessing step required resampling at 2000 Hz, removing noise using a band-pass filter and sliding window with 3 s patches and 1s stride. The authors claimed that the proposed model provided extremely low parameter consumption. However, the proposed method did not use test datasets that were different from the training and validation datasets.

Li et al. proposed CNN as a classification method [10]. However, the method did not extract the features directly using CNN but required a separate feature extraction process using multi-feature extraction. The proposed model was tested using 831 test datasets and reported an accuracy of 86.8% using 5-fold cross validation. In a study by Khrisnan et al. a 6 s heart sound recording duration was used as input for a feed forward neural network (FNN) [11]. They verified the model using 5-fold cross validation and reported an accuracy of 85.65 % for 100 test datasets.

Al-Naami et al. reported the highest accuracy at 89% using high-order spectral analysis and adaptive neuro-fuzzy inference system (ANFIS) [12]. However, they only used 1837 heart sound recordings from folder 'a', 'b', and 'e' in Physionet Challenge 2016 datasets. Khan et al. reported a 91.39% accuracy using MFCC for feature extraction and LSTM as a classifier algorithm [13]. He et al. extracted 512 features using several feature extraction methods, such as the Hilbert envelope, homomorphic environment map, wavelet envelope, and power spectral density envelope as inputs for 1D CNN. This study reported an accuracy of 87.3% [14]. Jeong et al. used a 5 s heart sound recording duration and removed datasets that were less than 5 s [15]. They applied a short-time Fourier transform (STFT) to transform to the time-frequency domain and generated spectrogram images of heart sound signals as input for the 2D CNN. The proposed method obtained 91% accuracy for the 208 test datasets.

There are several challenges in developing automated cardiac disorders using Physionet challenge datasets, such as the imbalance datasets of normal and abnormal conditions, and the variation of the length data due to different recording procedures from clinical data around the world. The aforementioned studies reported promising results related to the performance of automated heart sound detection. However, the accuracy needs to be improved because the early diagnosis of heart sound is critical to save patient life. In addition to the accuracy performance, low consumption parameters must be considered by researchers. Although the deep network provides good accuracy and can extract the information from raw signals directly, the complex architecture provides high consumption parameters and a long computational time.

Considering the limitations of the previous studies, we proposed an algorithm that can detect cardiac abnormalities based on heart sound signals, which not only improves accuracy performance, but also produces low parameter consumption. To achieve this, we proposed the mel frequency cepstral coefficient (MFCC) as a feature extraction method that has many advantages, such as being capable of capturing important information contained in the audio signal and providing data as minimal as possible without losing information. Furthermore, the proposed classification method is a simple ANN that contains one hidden layer to classify heart sound signals as normal or abnormal conditions.

Results

We applied 10-fold cross validation using 10968 training datasets and 3833 validation datasets to find the best model. The proposed model was assessed using 1626 test datasets, 1236 normal conditions and 390 abnormal conditions that were not part of the training and validation datasets. After conducting several simulations, the best model was obtained from the first fold, epoch 46, and batch size 64 using the Adam optimizer with a learning rate of 0.001.

Another important term that affects the system performance is the consideration of features as input to the ANN. Based on the result, 128 MFCC features that were inputted into ANN provided the best performance, with 94% accuracy and 93% AUC score, compared with 13, 40, and 96 MFCC features as shown in Table 1. Furthermore, the performance measures such as precision, recall and F1-score were also provided to evaluate the performance of the proposed model. The model performance using 128 MFCC features obtained a precision of 97% and 85% for normal and abnormal conditions, respectively. The recall value obtained was 95% for normal conditions and 90% for abnormal conditions, and the F1-score values obtained for normal and abnormal conditions were 96% and 88% respectively.

Table 1
The detailed performance of selected MFCC features.

MFCC features	Condition	Precision	Recall	F1-score	Accuracy	AUC
13	Abnormal	77%	90%	88%	90%	87%
	Normal	94%	92%	93%		
40	Abnormal	81%	87%	84%	92%	90%
	Normal	96%	94%	95%		
96	Abnormal	80%	89%	84%	92%	91%
	Normal	96%	93%	95%		
128	Abnormal	85%	90%	88%	94%	93%
	Normal	97%	95%	96%		

Even though the proposed model is still affected by false detection as shown in Fig. 1a, most of the test datasets were successfully classified according to their class. In addition, the receiver operating characteristic (ROC) curve is a probability curve that plots the true positive rate (TPR) against the false positive rate (FPR). Meanwhile, the area under the curve (AUC) as a summary of the ROC curve measures a classifier's ability to distinguish between classes. The higher the AUC score, the better the model performs in classifying normal and abnormal heart sound conditions. As shown in Fig. 1b, the AUC score was 93%. Therefore, we can conclude that the proposed model can overcome the imbalanced datasets and generalize the test datasets well.

Discussion

In this study, an automated classification of the cardiovascular disorders based on heart sound was proposed using MFCC as a feature extraction method and ANN as a classification method. The results of the proposed method outperformed the aforementioned studies released by Physionet Challenge 2016, which reported the best accuracy of 86.02% [16]. Moreover, the results obtained in this study are superior to those of the latest studies that involve the same dataset as shown in Table 2. However, the proposed method was tested using datasets that are available on the Physionet Challenge 2016 website.

Table 2
Performance comparison with previous studies.

Authors	Preprocessing	Feature Extraction	Classifier	Accuracy
Rubin et al. (2016) [7]	Heart sound segmentation using springer algorithm, selected 3 s duration of heart sound	Extracted 13 MFCC features, converted into 2D feature maps	2D CNN	84%
Nogueira et al. (2019) [8]	Logistic regression-HSMM-based heart sound segmentation	Extracted MFCC features, converted into 2D feature maps	SVM	82.33%
Xiao et al. (2019) [9]	Resample 2000 Hz, BPF, sliding window with 3 s patches and 1 s stride	1D time series signal	1D CNN	93%
Li et al. (2020) [10]	HPF and HSMM	Multi feature extraction	1D CNN	86.8%
Khrisnan et al. (2020) [11]	Down-sampled to 500 Hz and selected 6 s heart sound duration	1D time series signal	FNN	85.65%
Al-Naami et al. (2020) [12]	Band pass notch filter, BPF, and selected 5 s heart sound duration	Higher order statistics	ANFIS	89%
Khan et al. (2021) [13]	Logistic regression-HSMM-based heart sound segmentation	MFCC features	LSTM	91.39%
He et al. (2021) [14]	Normalization, HPF and LPF	Hilbert envelope, homomorphic environment map, wavelet envelope, and power spectral density envelope (512 data points)	1D CNN	87.3%
Jeong et al. (2021) [15]	BPF, selected 5 s heart sound duration	STFT	CNN	91%
Our Method	Selected 5 s heart sound duration	Extracted 128 MFCC features	ANN	94%
BPF, band-pass filter; LPF, low-pass filter; HPF, high-pass filter; STFT, short time Fourier transform; MFCC, mel-frequency cepstral coefficient; CNN, convolutional neural network; F-NN, feed forward neural network; ANN, artificial neural network.				

In terms of preprocessing, we discussed the selection of the heart sound duration. An optimized heart sound duration can obtain the best performance accuracy. Khrisnan et al. and Liu et al. suggested that a data record of at least 5 s is required for reliable detection of heart abnormalities. Therefore, the duration was set to 5 s, similar to that in previous studies [12, 15]. Compared with the aforementioned studies that used the same duration length to detect abnormalities in heart sound signals, the proposed model obtained superior performance accuracy as shown in Table 2.

For feature extraction, the number of MFCC features selections was determined as shown in Table 1. The best performance was achieved using 128 MFCC features. Instead of directly using the MFCC features, previous studies used the heat map image of MFCC features as input for a 2D CNN, requiring a long computational time [7, 8]. Other studies used various methods to extract the features from heart sound signals; for example, Xiao et al. used a 1D time series consisting of 6000 data points as input for 1D-CNN. Li et al used multi feature extraction including 20 features in the time domain, 12 amplitude features, 42 features in the energy domain, 16 features in the higher order statistic domains, 65 features in the cepstrum domain, 308 features in frequency domain, 4 features in cyclostationarity domain, and 10 features in the entropy domain. Therefore, 497 features were obtained and served as input for 1D CNN. He et al. extracted 512 features from Hilbert envelope, homomorphic environment map, wavelet envelope, and power spectral density envelope. Meanwhile, Jeong et al. used STFT as a feature extraction method and used 217×334 pixels heart sound spectrogram images as input for 2D CNN. Compared with the previous studies, this study successfully extracted 128 MFCC features that provided fewer data points as feature results. However, the extracted features can represent normal and abnormal heart sound signals with reduced computational time.

In most studies on heart sound classification, machine learning and deep learning algorithms have the same goal of providing high accuracy and low parameter consumption, thereby reducing computational time. Deep learning is reliable for the automatic extraction of the heart sound characteristics from the raw data signals. On the other hand, the training process takes a long-time and requires a larger dataset to avoid overfitting. Moreover, a complex architecture will produce a larger number of parameters and require a long computation time. This study proposed a simple ANN that consists of 128 nodes as the input layer, one hidden layer with 256 nodes, and two nodes as the output layer, producing 33538 parameters, which is a low number parameter in machine learning that provides fast computational time.

The existing public heart sound datasets used in this study were imbalanced between normal and abnormal heart sound. This makes it challenging to accurately classify heart sound in real clinical applications. The model was therefore evaluated using the ROC curve to ensure that the suggested model was not underestimated or overestimated because of dataset imbalance. This is because the ROC does not account for differences in the dataset sizes between classes.

The characteristics of normal and abnormal heart sound time series data were captured by MFCC as a feature extraction method and successfully identified using a simple ANN employed in this study. The best model was selected via 10-fold cross validation and evaluated using 1626 test datasets. The 94%

test accuracy and 93% AUC score show that the proposed method can overcome the problem of an imbalanced dataset. Moreover, the low parameter consumption resulted in fast computational time for screening the abnormality of heart sound for heart disease identification. Therefore, it helps the cardiologist devise a faster and proper treatment plan for the patients.

Methods

In this study, we proposed an optimal heart sound classification system using MFCC as a feature extraction method and ANN as a classification method. A general block diagram of the proposed method is presented in Fig. 2a.

Dataset

This study uses heart sound signals that were collected from the PhysioNet Challenge 2016 datasets, which are the most widely used in the previous studies for heart sounds analysis [17, 18]. The datasets consisted of six folders from 'a' to 'f' with a total of 3160 heart sound recordings, 2495 are normal conditions and 665 are abnormal conditions such as mitral regurgitation, aortic stenosis, and valvular surgery. The datasets were provided by different medical institutions using different recording settings. The duration of each heart sound varied from 5 to 120 s with a 2000 Hz sampling frequency. However, recent studies have claimed that 5 s is the minimum duration required to detect heart sound abnormalities. Therefore, this study divided the heart sound into short recordings of 5 s using time series segmentation. On the other hand, this process is capable of expanding the heart sound datasets by generating 16251 more datasets, consisting 12418 normal conditions and 3833 abnormal conditions.

Feature extraction using MFCC

The feature extraction of heart sound using MFCC consists of several processes [8] as shown in Fig. 2b, namely pre-emphasis, framing, windowing, and fast Fourier transform (FFT) to transform the heart sound signal from the time domain to the frequency domain, which is then transformed to the mel frequency domain, generating the mel filter bank. The mel filter bank value shows the amount of energy in the frequency range of each mel filter. A non-linear transformation is performed to take the natural logarithmic value of each mel filter. In the last process, the discrete cosine transform (DCT) returns the heart sound signal to the time domain and produces MFCC features as input to the ANN. The number of MFCC features used in this study was 13, 40, 96 and 128 features.

Classification using ANN

An ANN is a fully connected structure that mainly comprises three layers of its architecture [19]: input, hidden and output layers (Fig. 2a). Input layer acquires input from the external resources. The feature extraction results of MFCC (128 features) were assigned as input for the ANN architecture. The hidden layers process the input from the previous layer and transfer the result to the output nodes. In this study, the hidden layers that were used in the ANN consisted of one hidden layer with 256 nodes. The Rel-U

activation function was applied to the hidden layer, and a sigmoid activation function was applied for the output layer, which consisted of two nodes that were represented as normal and abnormal conditions.

System Performance

The system performance was evaluated using a confusion matrix to obtain the accuracy, precision, recall, and F1-score. The following equations were used to calculate the parameters in order to measure the effectiveness of the system in diagnosing normal and abnormal heart sound conditions.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Recall = TP/(TP + FN) \quad (3)$$

$$F1 \text{ score} = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision} \quad (4)$$

In Equations (1), (2), and (3), a true positive (TP) is a result in which the model correctly predicts the abnormal condition as the abnormal condition (positive class). Conversely, a true negative (TN), on the other hand, is a result in which the model correctly predicts the normal condition as the normal condition (negative class). A false positive (FP) is a result in which data is normal but incorrectly classified as abnormal. Meanwhile, a false negative (FN) is a result in which data is abnormal but incorrectly classified as normal [20].

Declarations

Acknowledgments

This research was partially supported by the National Research Foundation (NRF) of Korea under the basic engineering research project (2016R1D1A1B0101440) and the EDISON (NRF-2011-0020576) Programs, and the Ministry of Science and ICT (MSIT), Korea, under the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the Institute for Information & communications Technology Planning & Evaluation (IITP).

Author contribution statement

This manuscript is an intellectual product of the entire team. Y.N.F. wrote the machine learning source code and manuscript, performed data analysis, and interpreted the results. K.M.L. designed the study, reviewed, and revised the entire manuscript based on the results. All authors have read and approved the final manuscript.

Data availability

Publicly available datasets were analyzed in this study. The datasets can be accessed here: <https://physionet.org/challenge/2016/>

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chen, W. *et al.* Deep learning methods for heart sounds classification: A systematic review. *Entropy* **23**, 1–18 (2021).
2. Elgendi, M. *et al.* The use of photoplethysmography for assessing hypertension. *npj Digit. Med.* **2**, 1–11 (2019).
3. Zhang, W., Han, J. & Deng, S. Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Syst. Appl.* **84**, 220–231 (2017).
4. Clark, D., Ahmed, M. I., Dell'Italia, L. J., Fan, P. & Mcgiffin, D. C. An argument for reviving the disappearing skill of cardiac auscultation. *Cleve. Clin. J. Med.* **79**, 536–538 (2012).
5. Narváez, P., Gutierrez, S. & Percybrooks, W. S. Automatic segmentation and classification of heart sounds using modified empirical wavelet transform and power features. *Appl. Sci.* **10**, (2020).
6. Herzig, J., Bickel, A., Eitan, A. & Intrator, N. Monitoring Cardiac Stress Using Features Extracted From S1 Heart Sounds. *IEEE Trans. Biomed. Eng.* **62**, 1169–1178 (2015).
7. Rubin, J. *et al.* Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. *Comput. Cardiol. (2010)*. **43**, 813–816 (2016).
8. Nogueira, D. M., Ferreira, C. A., Gomes, E. F. & Jorge, A. M. Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features. *J. Med. Syst.* **43**, (2019).
9. 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).
10. Li, F., Tang, H., Shang, S., Mathiak, K. & Cong, F. Classification of heart sounds using convolutional neural network. *Appl. Sci.* **10**, (2020).
11. Krishnan, P. T., Balasubramanian, P. & Umapathy, S. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Phys. Eng. Sci. Med.* **43**, 505–515 (2020).
12. Al-Naami, B., Fraihat, H., Gharaibeh, N. Y. & Al-Hinnawi, A. R. M. A Framework Classification of Heart Sound Signals in PhysioNet Challenge 2016 Using High Order Statistics and Adaptive Neuro-Fuzzy Inference System. *IEEE Access* **8**, 224852–224859 (2020).
13. Faiq Ahmad, K., Anam, A. & Muhammad Salman, K. Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. *Inst. Phys. Eng. Med.* **41**, 11–14 (2020).

14. He, Y. *et al.* Research on segmentation and classification of heart sound signals based on deep learning. *Appl. Sci.* **11**, 1–15 (2021).
15. Jeong, Y., Kim, J., Kim, D., Kim, J. & Lee, K. Methods for improving deep learning-based cardiac auscultation accuracy: Data augmentation and data generalization. *Appl. Sci.* **11**, (2021).
16. Potes, C., Parvaneh, S., Rahman, A. & Conroy, B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. *Comput. Cardiol. (2010)*. **43**, 621–624 (2016).
17. Liu, C. *et al.* An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* **37**, 2181–2213 (2016).
18. Classification of Heart Sound Recordings. *The PhysioNet Computing in Cardiology Challenge* <https://physionet.org/challenge/2016/> (2016).
19. Sairamya, N. J., Susmitha, L., Thomas George, S. & Subathra, M. S. P. *Hybrid approach for classification of electroencephalographic signals using time-frequency images with wavelets and texture features. Intelligent Data Analysis for Biomedical Applications: Challenges and Solutions* (Elsevier Inc., 2019). doi:10.1016/B978-0-12-815553-0.00013-6.
20. Novakovic, J., Veljovi, A., Ilic, S., Papic, Z. & Tomovic, M. Evaluation of Classification Models in Machine Learning. *Theory Appl. Math. Comput. Sci.* **7**, 39–46 (2017).

Figures

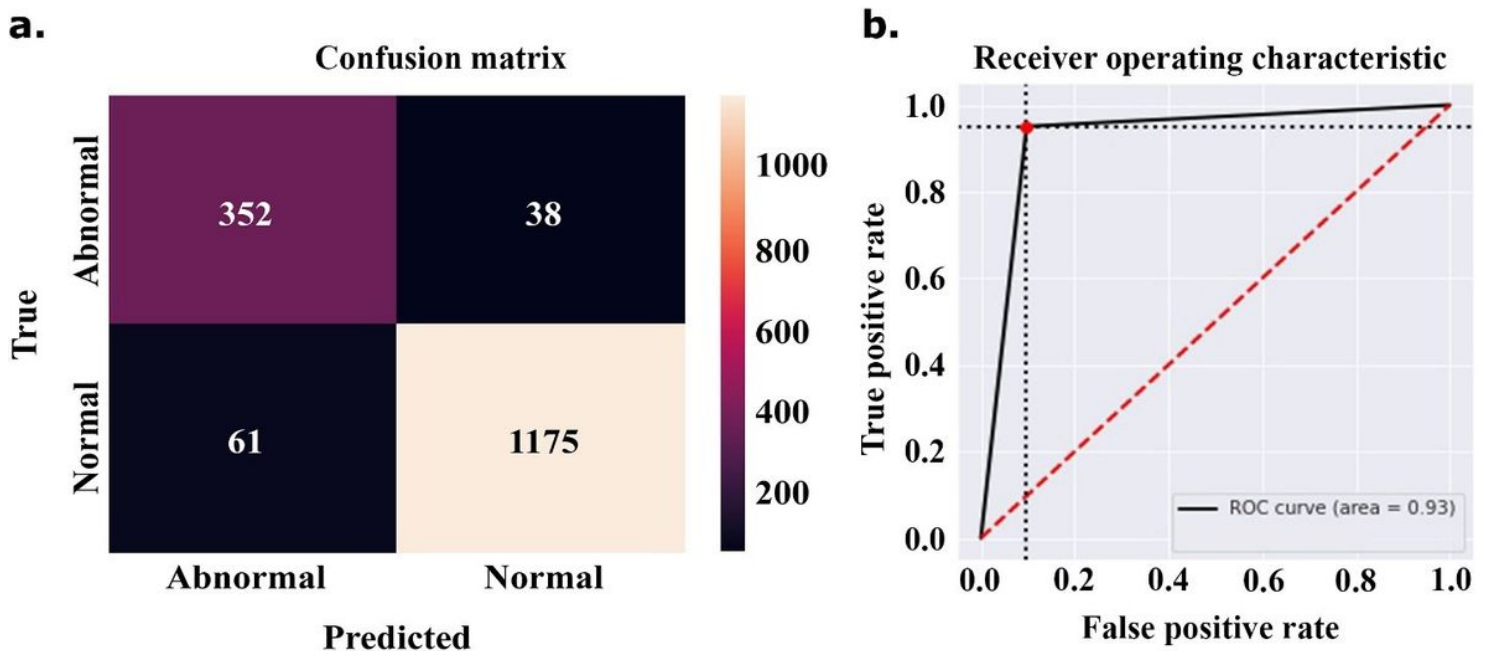


Figure 1

The proposed model performance. (a) The confusion matrix of test datasets; (b) The ROC curve of the ANN model.

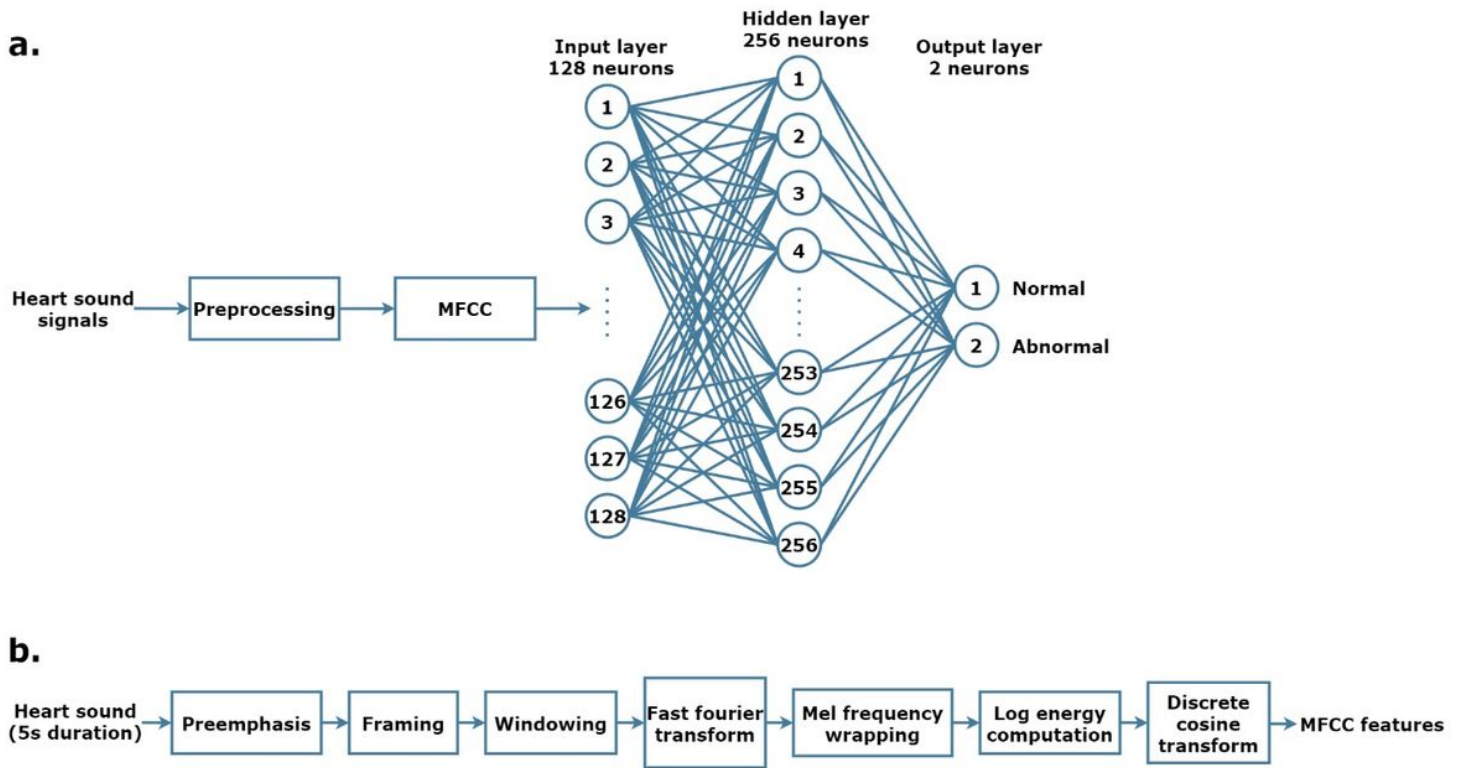


Figure 2

The proposed method. (a) General block diagram system; (b) Block diagram of MFCC method.