Neural Networks 130 (2020) 22–32

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Heart sound classification based on improved MFCC features and convolutional recurrent neural networks

Muqing Deng [a,*], Tingting Meng [b], Jiuwen Cao [b], Shimin Wang [c], Jing Zhang [d], Huijie Fan [e,**]

[a] School of Automation and Guangdong Key Laboratory of IoT Information Technology, Guangdong University of Technology, Guangzhou, China
[b] Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, China
[c] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada
[d] School of Computer Science, University of Sydney, Sydney, Australia
[e] Center of Preventive Disease, The People's Hospital of Yangjiang, Yangjiang, China

ARTICLE INFO

Article history:
Received 3 September 2019
Received in revised form 2 May 2020
Accepted 19 June 2020
Available online 23 June 2020

Keywords:
Heart sound classification
Convolutional neural network
Recurrent neural network
Improved MFCC features

ABSTRACT

Heart sound classification plays a vital role in the early detection of cardiovascular disorders, especially for small primary health care clinics. Despite that much progress has been made for heart sound classification in recent years, most of them are based on conventional segmented features and shallow structure based classifiers. These conventional acoustic representation and classification methods may be insufficient in characterizing heart sound, and generally suffer from a degraded performance due to the complicated and changeable cardiac acoustic environment. In this paper, we propose a new heart sound classification method based on improved Mel-frequency cepstrum coefficient (MFCC) features and convolutional recurrent neural networks. The Mel-frequency cepstrums are firstly calculated without dividing the heart sound signal. A new improved feature extraction scheme based on MFCC is proposed to elaborate the dynamic characteristics among consecutive heart sound signals. Finally, the MFCC-based features are fed to a deep convolutional and recurrent neural network (CRNN) for feature learning and later classification task. The proposed deep learning framework can take advantage of the encoded local characteristics extracted from the convolutional neural network (CNN) and the long-term dependencies captured by the recurrent neural network (RNN). Comprehensive studies on the performance of different network parameters and different network connection strategies are presented in this paper. Performance comparisons with state-of-the-art algorithms are given for discussions. Experiments show that, for the two-class classification problem (pathological or non-pathological), a classification accuracy of 98% has been achieved on the 2016 PhysioNet/CinC Challenge database.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Cardiovascular disease remains the leading cause of death and poses a serious threat to the global population due to its burstiness and recidivity. Hence, investigating early preventive methods for cardiac disorders becomes important and significant (Members, et al., 2010, 2016).

The heart sound signal, carries early pathological information of cardiovascular disorders (Yuenyong et al., 2011), has been demonstrated to be effective for the early detection of latent cardiovascular diseases. Traditionally, heart sound is used by the medical specialists for cardiac disorder detection through auscultation. It is non-invasive, cost-effective and requires minimal equipment, making it very suitable for cardiac examination, especially in small primary health care clinics. In reality, however, the heart sound auscultation relies heavily on physicians' clinical experience and examination skills. The accuracy rate of auscultation by cardiologists is about 80% (Strunic et al., 2007), whereas the accuracy by primary care physicians is about 20%–40% (Lam, et al., 2005). Attempts to resolve this dilemma have resulted in the development of computer-based automatic heart sound analysis and classification.

The standard process for the computer-based heart sound analysis can be summarized into the following steps: (1) preprocessing (segmentation and filter); (2) feature extraction; (3) classifier design (Gupta et al., 2007; Yang, et al., 2016). In the past decades, fruitful methods have been reported for each step of the aforementioned heart sound analysis process (Emmanuel, 2012; Monali & Aparana, 2012). To name a few, in Sun et al.

* Corresponding author at: School of Automation and Guangdong Key Laboratory of IoT Information Technology, Guangdong University of Technology, Guangzhou, China.
** Corresponding author.
E-mail addresses: mqdeng@gdut.edu.cn (M. Deng), 15521107492@163.com (H. Fan).

https://doi.org/10.1016/j.neunet.2020.06.015
0893-6080/© 2020 Elsevier Ltd. All rights reserved.

(2014), an automatic heart sound signal segmentation method based on Hilbert transform was proposed. In Springer et al. (2016), the hidden semi-Markov model (HSMM) method was extended with logistic regression to achieve the signal segmentation in noisy environment. In Yuenyong et al. (2011) and Deng and Han (2016), the authors proposed methods for detecting heart sound anomalies without signal segmentation. In Zabihi et al. (2016), forty time–frequency features were extracted from undivided heart sound signals and heart sound anomaly detection was performed. Different kinds of spectral features, including spectral parameter models, instantaneous frequency and amplitude (IFA) and octave power, were extracted in Schmidt et al. (2015) to characterize the time–frequency attribute. Fast wavelet decomposition was used in Kumar, et al. (2006) to extract high frequency heart sound feature. As for heart sound classification, decision trees algorithms were used in Stasis et al. (2003) for diagnosis task. In Sh-Hussain, et al. (2017), a hidden Markov model was constructed for heart sound classification and promising experimental results were achieved. Wang et al. used a combination of hidden Markov model (HMM) and MFCC features to classify abnormal heart sound signals (Wang et al., 2007). A random forest composed of 70 tree-structured classifiers was used to avoid the over-fitting problem (Balili et al., 2015). Three integration techniques, namely Bagging, AdboostM1 and Random Subspace were adopted to improve the recognition rates based on low-performance classifiers (Ali et al., 2017).

With the rapid development of artificial intelligence algorithms, deep learning neural networks (DNN) have been explored for human heart sound classification in recent years. In contrast to conventional heart sound classification algorithms, the notable merit of deep learning algorithms is attributed to its feature extraction functions from complex heart sound signals. In Bozkurt et al. (2018), three different kinds of heart sound features, including Mel-spectral map, MFCC and sub-band envelope were extracted and compared with various feature fusion and segmentation strategies based on feedforward CNNs. In Chen, et al. (2016), the Short Time Fourier Transform (STFT) magnitude spectrum was used as the input of CNNs. The authors optimized the model complexity by varying the number of convolutional layers. It was found that the CNN with two convolutional layer achieved the best results. The authors in De Vos and Blanckenberg (2007) used time–frequency features and energy characteristics in 12 frequency bins at 10 equally-spaced time intervals over each heart cycle. Power spectral density (PSD) of the raw heart sound signal was extracted in Nilanon et al. (2016) and fed to the CNN for automatic heart sound classification. A robust method for heart sound classification that combines a deep CNN-based feature extractor and a support vector machine (SVM) was presented and evaluated in Tschannen et al. (2016). In Thomae and Dominik (2016), end-to-end deep neural networks were developed, in which RNN was constructed as the convolutional front end. Latif et al. proposed a RNN-based abnormal heartbeat detection algorithm in Latif et al. (2018). The authors explored the performance and computational complexity of four RNN models, namely Long Short-term Memory (LSTM) (Ergen & Kozat, 2017; Hochreiter & Schmidhuber, 1997), Gated Recurrent Units (GRU) (Cho, et al., 2014; LeCun et al., 2015), Bidirectional Long Short-Term Memory (BLSTM) (Zeng et al., 2014) and Bidirectional Gated Recurrent Units (BGRU) (Zhang et al., 2015).

Although extensive works on heart sound classification have been presented, most of them suffered from a degraded performance due to the complicated and changeable cardiac acoustic environment (Clifford, et al., 2017). Since early cardiac disorder detection is significant for patients to take precaution to reduce the harm caused by latent disorders, in this paper, we present a novel heart sound classification algorithm for disorder detection based on improved MFCC features of the heart sound signals and convolutional recurrent neural networks (CRNN).

The main contributions of this paper are threefold: (1) The Mel-frequency cepstrums are extracted without dividing the heart sound signal, reducing the computational complexity, and the standard MFCC, the first and second order differential parameters of MFCC feature, are combined for the dynamics representation of heart sound signal. (2) A heart sound classification algorithm that combines 2D-CNN and LSTM for feature extraction and classification, and uses the MFCC-based features as the input is constructed. The proposed algorithm takes advantage of the encoded local characteristics extracted from the CNN and the long-term dependencies captured by the RNN. (3) Comprehensive studies on different network parameters, different network connection strategies, and comparisons with state-of-the-art algorithms, are presented in the paper. To demonstrate the effectiveness of the proposed algorithm, experiments on the PhysioNet Computational Cardiology (CinC) 2016 Challenge Database (Liu, et al., 2016) are conducted.

2. The proposed classification algorithm

In this section, a detailed description of our proposed algorithm is given in three parts: heart sound signal preprocessing, MFCC-based signal representation, and CRNN classification model. The overall flowchart of the proposed method is depicted in Fig. 1. In the training phase, the original heart sound signal is preprocessed to remove the low-frequency artifacts, baseline wandering and high-frequency interference. The standard MFCC, the first order MFCC (ΔMFCC) and the second order MFCC (Δ_2MFCC) features of heart sound signals for pathological and healthy subjects are extracted to construct the training database. The labeled MFCC-based features are inputted into the CRNN model and used as benchmark samples for classifier learning and training. In the testing phase, the testing signal is preprocessed to extract the MFCC features. The category of a testing signal is decided using the pre-trained CRNN model.

2.1. Signal preprocessing

Heart sound signal preprocessing is the basis of the whole classification algorithm. Using the fifth order Butterworth bandpass filter (pass-band: 25–400 Hz) mentioned in Refs. Gaikwad and Chavan (2014), low-frequency artifacts, baseline wandering and high-frequency interference are removed from the original collected heart sound signals. Despite that signal preprocessing, segmentation and morphology analysis are the main contents for an automatic heart sound analysis system, we do not focus on preprocessing and signal segmentation techniques in this paper, as they are quite well established techniques. Special emphasis is given to feature extraction and pattern recognition methodologies applied to the preprocessed signals instead. Figs. 2 and 3 show the original signals and preprocessed signals for normal heart sound and abnormal heart sound, respectively.

2.2. MFCC-based signal representation

The heart sound signal carries early pathological information of cardiovascular disorders, which is helpful for the early detection of latent cardiovascular diseases. To derive an effective feature representation for the cardiac disorder detection, we exploit the Mel-frequency cepstrums on heart sound signal and proposed an improved MFCC-based feature extraction algorithm.

The Mel-frequency cepstrums reflect the nonlinear relationship between the human ear and the frequency of the sound

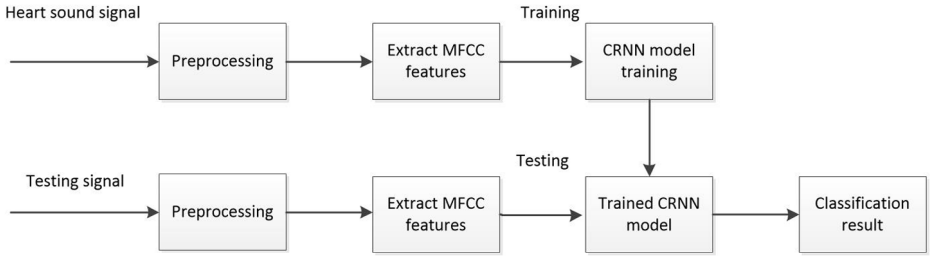

Fig. 1. The flowchart of the proposed heart sound classification algorithm.

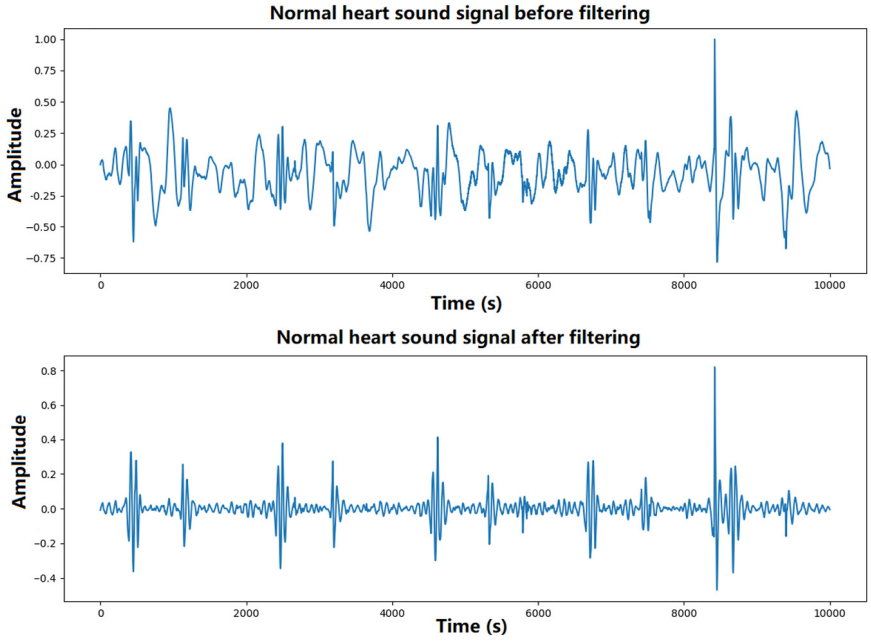

Fig. 2. The original and preprocessed signals of normal heart sound.

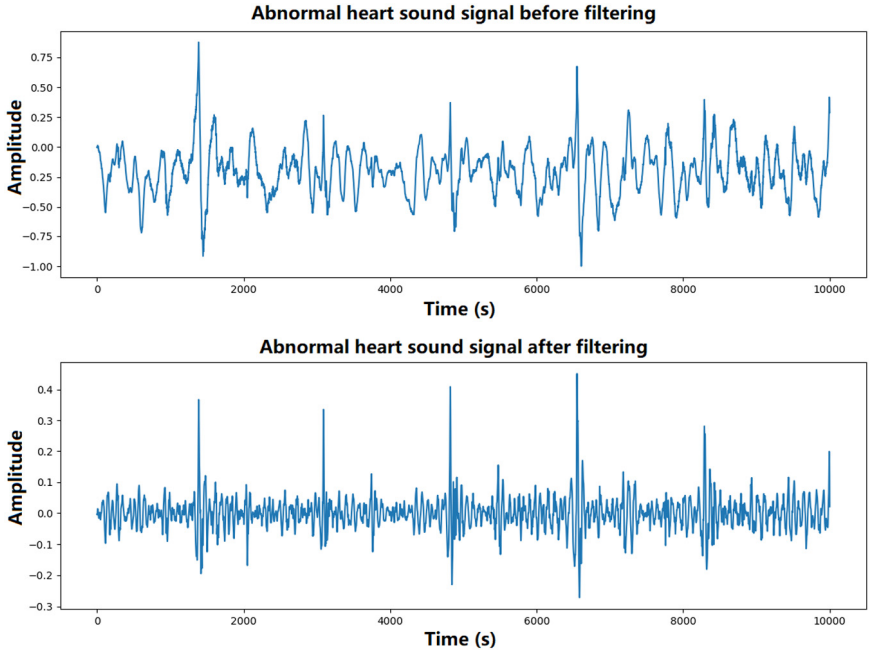

Fig. 3. The original and preprocessed signals of abnormal heart sound.

heard, which can be expressed by using the following relationship of heart sound signal:

$$Mel(f) = 2595lg(1 + f/700) \tag{1}$$

where $Mel(\cdot)$ is the Mel scale frequency, and f is the actual frequency of heart sound signal. The specific process of MFCC feature extraction can be presented as follows:

1. Pre-emphasis

The heart sound signal is pre-emphasized to amplify the high frequency components. After passing by the pre-emphasis filter $H(z)$, the random noises can be effectively suppressed. The equation of the filter $H(z)$ is as follows:

$$H(z) = 1 - \mu z^{-1} \tag{2}$$

where the coefficient μ is usually within (0.9,1).

2. Window framing

Despite that heart sound signal belongs to a kind of non-stationary signal, the signal reaches stable state between 24 ms and 256 ms. Therefore, the heart sound signal is first framed, and the feature extraction is performed in a short frame. In order to achieve a smooth transition between frames, 50% overlaps between consecutive frames are employed. The Hamming window function $W(n)$ is adopted for signal framing as it can effectively overcome the leakage phenomenon (Astuti et al., 2012). The detailed equation of Hamming window function $W(n)$ can be given as follows:

$$W(n) = (1 - \alpha) - \alpha cos(2\pi n/(N-1)), 0 \le n \le N-1 \tag{3}$$

where α is empirically set to be 0.46 (Trang et al., 2014); N represents the number of samples in each frame.

3. Discrete Fourier transform (DFT)

To obtain a spectrum $X(k)$, we transform the time domain heart sound signal $x(n)$ into frequency domain signal through DFT, that is

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}, 0 \le n, k \le N-1 \tag{4}$$

where $x(n)$ is the heart sound signal pre-treated by noise reduction, windowing and framing, N is the number of samples in each frame.

4. Power spectrum calculation

Taking the signal spectrum $X(k)$ as the square of its modulus, the power spectrum $P(k)$ is obtained as follows:

$$P(k) = \frac{1}{N}|X(k)|^2 \tag{5}$$

5. Mel frequency filter

The power spectrum $P(k)$ is passed through a set of Mel-scale triangular filter banks to obtain a Mel spectrum. At each frequency, the product of $P(k)$ and filters $H_m(k)$ is calculated. Define a triangular filter bank with M filters, the frequency response of the triangular filter $H_m(k)$ is calculated as:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \le k \le f(m+1) \\ 0, & k > f(m+1) \end{cases} \tag{6}$$

where $f(m)$ is the center frequency of the Mel triangle filter. Here, we have

$$\sum_{m=0}^{M-1} H_m(k) = 1 \tag{7}$$

6. The logarithmic spectrum

The logarithm energy spectrum $S(m)$ at each frame is then obtained by using a logarithmic operation:

$$S(m) = \ln[\sum_{k=0}^{N-1} P(k) H_m(k)], 0 \le m \le M \tag{8}$$

where P_m is the power spectrum, $H_m(k)$ is the filter bank, and M is the number of filter banks.

7. Discrete cosine transform (DCT)

The above logarithmic spectrum is subjected to discrete cosine transform (DCT) to obtain the MFCC coefficients $C(n)$:

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos(\pi n(m - 0.5)/M), n = 1, 2, \ldots, L \tag{9}$$

where L represents the order of MFCC coefficients, M is the number of filter banks.

8. Dynamic MFCC feature extraction

With above description, it is noticed that the obtained MFCC coefficients only reflect the static characteristics of the heart sound signal. Since human ear is more sensitive to the dynamic characteristic of acoustic signal, the dynamic information of heart sounds spectrum also contains abundant information, which can be used to further improve the classification accuracy. In order to reflect the dynamic information of the heart sound signal, the ΔMFCC, Δ_2MFCC are extracted by the first and the second difference of MFCC. The first difference coefficients of MFCC (ΔMFCC) can be calculated by the following formula:

$$D(n) = \frac{1}{\sqrt{\sum_{i=-k}^{i=k} i^2}} \sum_{i=-k}^{i=k} i \cdot C(n+i) \tag{10}$$

where $C(n+i)$ is a frame of MFCC parameters. The k value is set to be 2.

By substituting the results D_n in Eq. (10), the second difference MFCC parameters (Δ_2 MFCC) can be obtained:

$$D_2(n) = \frac{1}{\sqrt{2\sum_{i=-k}^{i=k} i^2}} \sum_{i=-k}^{i=k} i \cdot D(n+i) \tag{11}$$

$D_2(n)$ is the second difference of MFCC coefficients.

Fig. 4 shows the curves of MFCC features for normal and abnormal heart sound signals. Each line in Fig. 4 represents one-dimensional coefficient in the extracted MFCC features. Figs. 5 and 6 show the curves of first and second difference MFCC features for normal and abnormal heart sound signals, respectively. We use the MFCC features, the first difference vector in (10) and the second difference vector in (11) to construct a 39 dimensional feature vector for further feature learning through the proposed CRNN model.

2.3. CRNN classification model

In the past few years, convolutional neural network (CNN) has been investigated and shown to be effective in large scale and high-dimensional data learning. And recurrent neural network (RNN) is shown to be effective in temporal sequence learning and long-term dependencies capturing. In this paper, we combine the CNN and RNN to perform feature learning on the MFCC-based representation derived from the heart sound signal, which takes the advantage of the encoded local characteristics extracted from the CNN and the long-term dependencies captured by the RNN. Two network architectures, namely convolutional recurrent neural networks (CRNN) and paralleling recurrent convolutional neural network (PCRNN) are investigated in the proposed paper for heart sound signal classification.

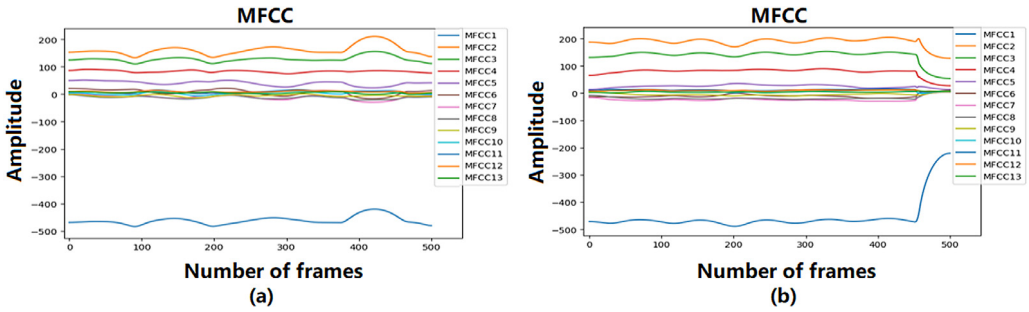

Fig. 4. MFCC feature curves: (a) normal heart sound signal (b) abnormal heart sound signal.

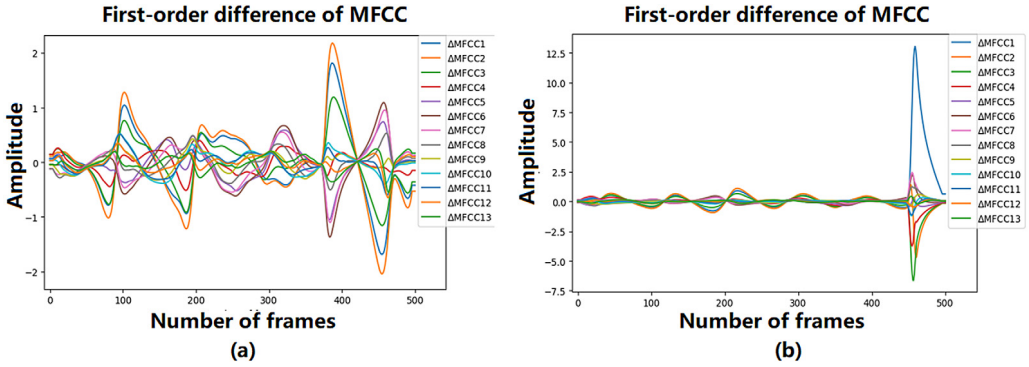

Fig. 5. The Δ MFCC feature curves: (a) normal heart sound signal (b) abnormal heart sound signal.

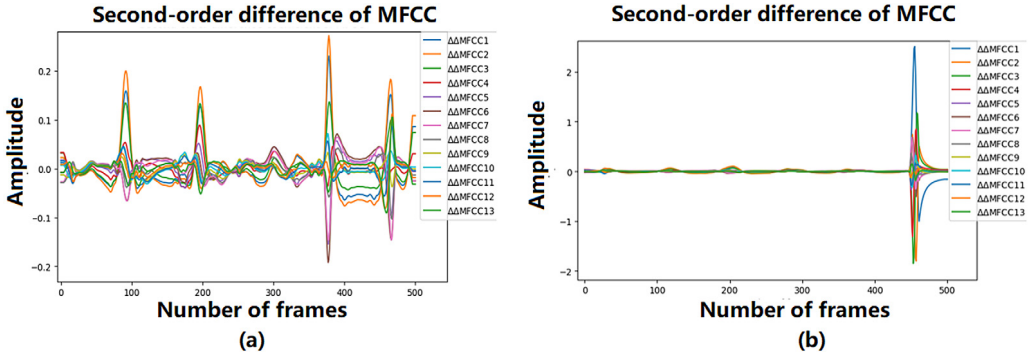

Fig. 6. The Δ_2 MFCC feature curves: (a) normal heart sound signal (b) abnormal heart sound signal.

2.3.1. Convolutional recurrent neural network (CRNN)

The network structure of CRNN proposed in this paper is shown in Fig. 7. Three convolutional layers, the first layer with 32 kernels, the second layer with 32 kernel and the third layer with 64 kernel, are used in CNN blocks. The learnable kernel size in each layer is set to be 3×3, and the popular ReLu activation function is used in each convolutional layer. Followed by each convolutional layer, a max-pooling is employed where 2×2 windows are used and the stride is 2×2. Three residual blocks for the corresponding convolutional and pooling layers are used. Within a residual block, a batch normalization (BN) layer and a dropout layer are constructed followed by a max-pooling layer. BN layer normalizes each mini-batch throughout the entire network, reducing the internal covariate shift caused by progressive transforms (Sigtia et al., 2016), and the dropout layer can reduce the number of neurons and prevent overfitting. Hence, for each input sample, a feature map can be obtained. After the convolutional and max-pooling layers, a Long Short-term Memory (LSTM) layer is applied to learn the temporal features among the obtained feature maps, and a fully connected (FC) layer with 64 neurons is performed to learn the global features. Finally, a softmax layer is adopted to derive the probability distribution across two classes corresponding to normal and abnormal heart sounds.

2.3.2. Paralleling recurrent convolutional neural network (PRCNN)

The aforementioned CRNN model uses the RNN for temporal features learning based on the outputs of CNNs, however, the temporal features underlying the original heart sound signal cannot be preserved. In order to capture the spatial and temporal features of the original signal, this section further investigates another network architecture, namely paralleling recurrent convolutional neural network (PRCNN), for heart sound signal classification.

As shown in Fig. 8, the proposed PRCNN architecture is divided into four blocks. We utilize the extracted MFCC features as the input of the network. The MFCC features are fed parallelly into CNN and RNN blocks for feature learning. The proposed PRCNN model takes advantage of spatial characteristics extracted from the CNNs and the temporal characteristics by the RNN. The outputs of

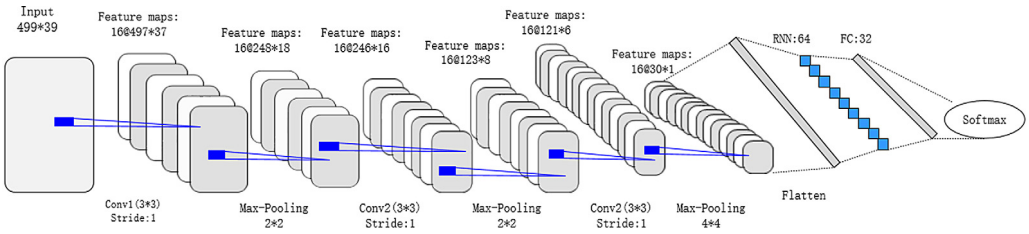

Fig. 7. The network structure of CRNN.

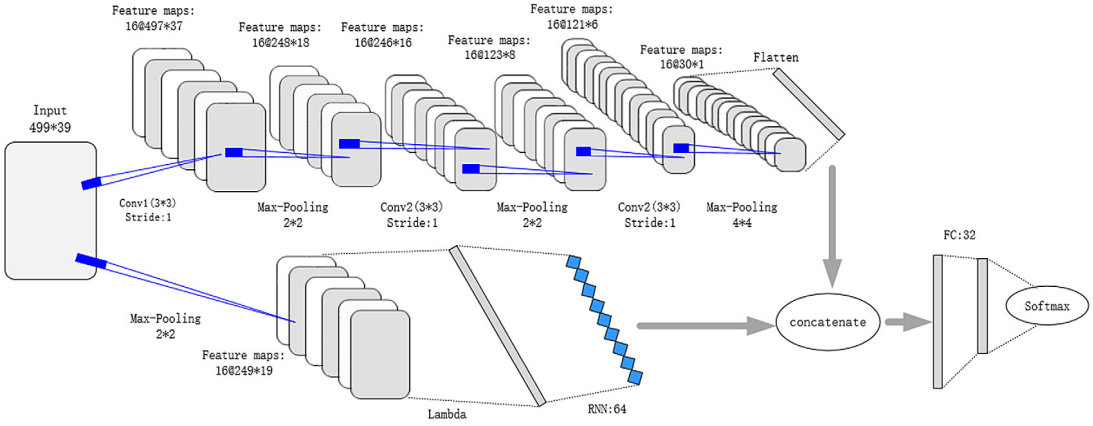

Fig. 8. The network structure of PRCNN.

the two parallel blocks are then fused into one uniform feature vector for classification. A fully connected layer is applied and a softmax layer is adopted to derive the probability distribution across different classes.

Three convolutional layers are used in the CNN block, the first layer with 16 kernels, the second layer with 32 kernels and the third layer with 64 kernels. Followed by convolutional layers, three max-pooling layers are employed, the first layer with 2×2 window, the second layer with 2×2 window and the third layer with 4×4 window. In the RNN block, we adopt a max-pooling layer to perform dimensionality reduction and use a LSTM layer for heart sound signal temporal feature learning. We concatenate the outputs of the CNNs block and RNN block into one feature vector. A network that consists of fully connected layers and softmax layers is developed for further feature learning and classification on the concatenated feature vectors.

3. Experiments

Experiments are carried out in this section to evaluate the performance of the proposed heart sound classification algorithm for cardiac disorder detection. The results on three testing scenarios are reported. The first experiment tests the parameter sensitiveness of the proposed CRNN based classification algorithm. The second experiment shows the performance comparisons on different network connection strategies. The last experiment shows the performance comparisons with algorithms based on state-of-the-art heart sound signal feature extraction and machine learning algorithms.

The heart sound signals used in our experiments are from the PhysioNet/CinC challenge 2016 database (Liu, et al., 2016), publicly available on the PhysioNet website.[1] The database consists of nine subdatabases from different research groups, containing a total of 3,240 original heart sound recordings from healthy subjects and patients with various heart diseases. These heart sound recordings are collected by using various electronic stethoscopes (e.g. Welch Allyn Meditron, Littmann E4000), ranging from 5 s to 120 s in length, and the sampling frequency is 2000 Hz. The heart sound samples are intercepted to 5 s and divided into three different sets of mutually exclusive populations, using 75% of them to train the network, 15% for validation and 10% to test the network. Since that the number of normal and abnormal heart sound recordings in the dataset is not balanced, we use the k-means SMOTE algorithm (Georgios et al., 2018) to perform resample on the training set, and then extract the MFCC features for each sample.

For the evaluation of the proposed method, four measurements are used for each experiment: Accuracy, Recall, Precision, and F1. These measurements are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{13}$$

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall} \tag{17}$$

where, TP is the number of true positive results, TN is the number of true negative results, FP is the number of false positive results and FN is the number of false negative results.

[1] The CinC challenge 2016 database, Available: http://www.physionet.org/physiobank/database/challenge/2016/

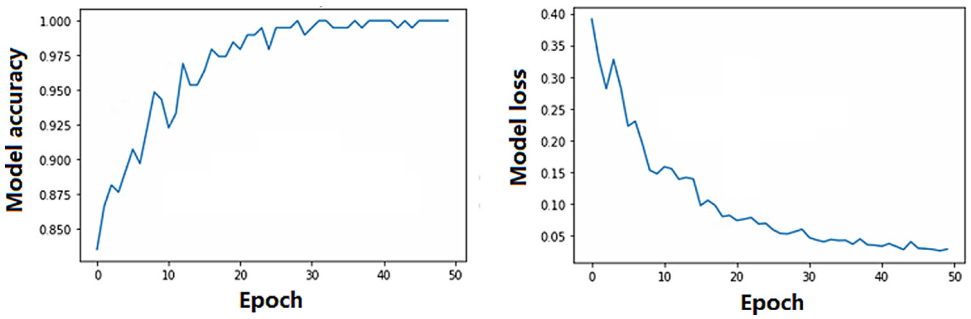

Fig. 9. Performance improvement of the model with increasing of training epochs.

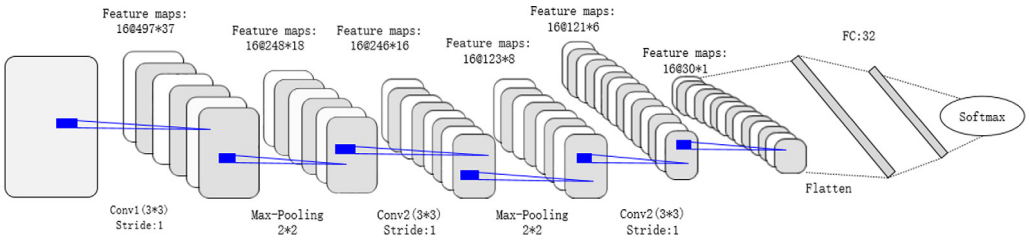

Fig. 10. The network structure of CNN.

Table 1
Classification accuracy on different dropout rates.

Dropout rate	0.2	0.4	0.5	0.6	0.8	No Dropout
Accuracy	0.9501	0.9535	0.9701	0.9634	0.9401	0.9468
F1	0.9501	0.9527	0.9698	0.9634	0.9391	0.9470

3.1. Performance on network parameters

In this section, we test the parameter sensitiveness of the proposed MFCC-based heart sound classification algorithm using CRNN and PRCNN, including the dropout rate, the learning rate and training epoch, etc. The adaptive moment estimation (Adam) optimizer, a type of stochastic gradient descent (SGD) algorithm, is adopted for the network training and weight parameters optimization. An initial learning rate of 0.01 and exponential decay rates of 0.9, 0.999 for the first and second moment estimates are used.

We test the classification performance on the different dropout rates. The dropout rate values are taken from the interval [0.2, 0.8] and the epoch is set to be 50 for all trials. Table 1 depicts the classification accuracy on the CinC challenge 2016 database with respect to different dropout rates. It is shown that the best performance can be obtained by using the dropout rate of 0.5 in the experiment. In addition, the performance improvement of the model with increasing of training epochs is shown in Fig. 9, which indicates that setting the parameter for training epoch to 50 is sufficient for the algorithm to converge.

3.2. Performance on different network architectures

To further illustrate the classification performance, we test the proposed algorithm on different deep learning network architectures, where CNN, RNN, CRNN and PRCNN are adopted (Table 2). The pure CNN model (Fig. 10) and RNN model (Fig. 11) are constructed for performance comparison. As shown in Table 2, Conv(x,y,z,n) stands for a convolutional layer with x filters, where x and y are the width and height of 2D filter window, n is the stride. MaxPool(x,y) stands for a Max Pooling layer where x and y represent the pool sizes. BN stands for a batch normalization

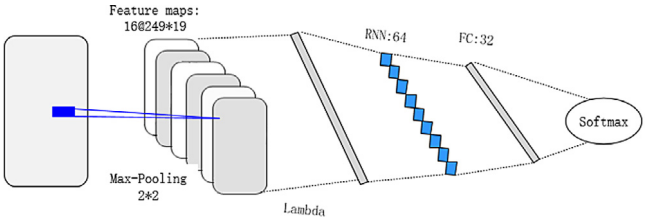

Fig. 11. The network structure of RNN.

layer. FC(x) stands for a fully connected layer with x nodes. LSTM(x) stands for an LSTM layer where x is the dimensionality of the output space. GRU(x) stands for a gated recurrent unit (GRU) layer where x is the dimensionality of the output space. Drop(x) stands for a dropout layer with a dropout coefficient equal to x. The commonly-used Rectified Linear Units (ReLU) is used as the activation function, with the exception of the last layer with softmax activation. Categorical cross entropy is used as the loss function and batch Stochastic Gradient Descent (SGD) with Adam is adopted for the optimization process.

Fig. 12 shows the experimental results obtained by the proposed MFCC-based classification algorithm with different network architectures. It can be observed that the models CRNN-a and PRCNN-a achieve the best results in terms of accuracy and F1. The CRNN-a model attains an accuracy of 98.34%, a recall of 98.66%, a precision of 98.01% and F1 score of 98.34%. The PRCNN-a model attains an accuracy of 97.34% and F1 score of 97.33%. Particularly, to show the detailed accuracy, we give the confusion matrix of the classification accuracy on each model with the proposed MFCC-based features in Fig. 13. From the experimental results, we can see that the model with three convolutional layer followed by one LSTM layer achieves the best results. This section further reports the experimental results of the proposed method under 2-fold, 5-fold and 10-fold cross-validation. For the 2-fold (5-fold, 10-fold) cross-validation, the whole database can be divided into two (five, ten) subsets. For each experiment, one of the subsets is randomly drawn for the test set and the remaining subsets are used as the training set. The final classification

Table 2
Details of the adopted deep learning architectures in the experiments.

Model	Architecture details
CNN	Conv(16,3,3,1)-MaxPool(2,2)-BN-Conv(32,3,3,1)-MaxPool(2,2)-BN-Conv(64,3,3,1)-MaxPool(2,2)-BN-FC(32)-Drop(0.5)-FC(2)
RNN	LSTM(64)-FC(64)-Drop(0.5)-FC(2)
CRNN-a	Conv(16,3,3,1)-MaxPool(2,2)-BN-Conv(32,3,3,1)-MaxPool(4,4)-BN-Conv(64,3,3,1)-MaxPool(2,2)-BN-LSTM(64)-FC(32)-Drop(0.5)-FC(2)
CRNN-b	Conv(16,3,3,1)-MaxPool(2,2)-BN-Conv(32,3,3,1)-MaxPool(4,4)-BN-Conv(64,3,3,1)-MaxPool(2,2)-BN-GRU(64)-FC(32)-Drop(0.5)-FC(2)
PRCNN-a	Conv(16,3,3,1)-MaxPool(2,2)-BN-Conv(32,3,3,1)-MaxPool(4,4)-BN-Conv(64,3,3,1)-MaxPool(2,2)-BN; MaxPool(2,2)-LSTM(64); FC(32)-Drop(0.5)-FC(2)
PRCNN-b	Conv(16,3,3,1)-MaxPool(2,2)-BN -Conv(32,3,3,1)-MaxPool(4,4)-BN-Conv(64,3,3,1)-MaxPool(2,2)-BN; MaxPool(2,2)- GRU(64); FC(32)-Drop(0.5)-FC(2)

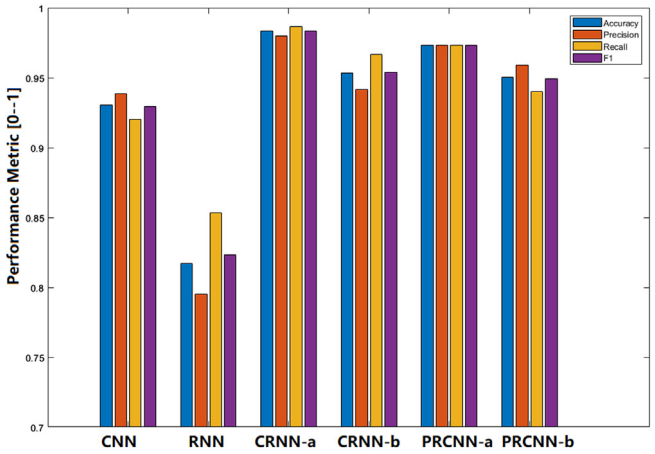

Fig. 12. Classification performance of the proposed algorithm on different network architectures.

accuracies are the average of these experiments. By using 2-fold, 5-fold and 10-fold cross-validation, the correct classification rates are reported to be 91.46%, 95.95% and 98.63%, respectively. It is shown that the proposed method achieves good performance under different gallery sizes, which indicates that the proposed method has a promising discriminatory ability in the classification of heart sound signals between healthy subjects and patients with various heart diseases.

3.3. Performance comparison with state-of-the-art methods

In this section, we compare the classification performance of the proposed algorithm with several algorithms based on state-of-the-art heart sound signal classification methods. Since the hidden testing set of the PhysioNet/CinC challenge is lacked, convincing and quantitative comparisons with different methods are almost impossible to achieve. Therefore, qualitative comparisons with the related works are firstly investigated in this part. Note that, if multiple classifiers or features are used in the related work, only the best results are reported. As shown in Table 3, qualitative comparisons with existing other works on heart sound classification using deep learning techniques are given.

Secondly, to comply with the convincing quantitative comparison and the development period requirements, we implement the LeNet-5, Modified LeNet-5, Default AlexNet and Modified AlexNet models in Dominguez-Morales et al. (2018) and present the experimental results in Fig. 14. According to the comparison results, the proposed method is not inferior to other methods.

Furthermore, based on the same MFCC-based features, the traditional popular support vector machines (SVM), K-Nearest Neighbor (KNN) and extreme learning machine (ELM) methods are selected as baseline models for comparison. For SVM, three kernels including linear, polynomial, and radial basis function

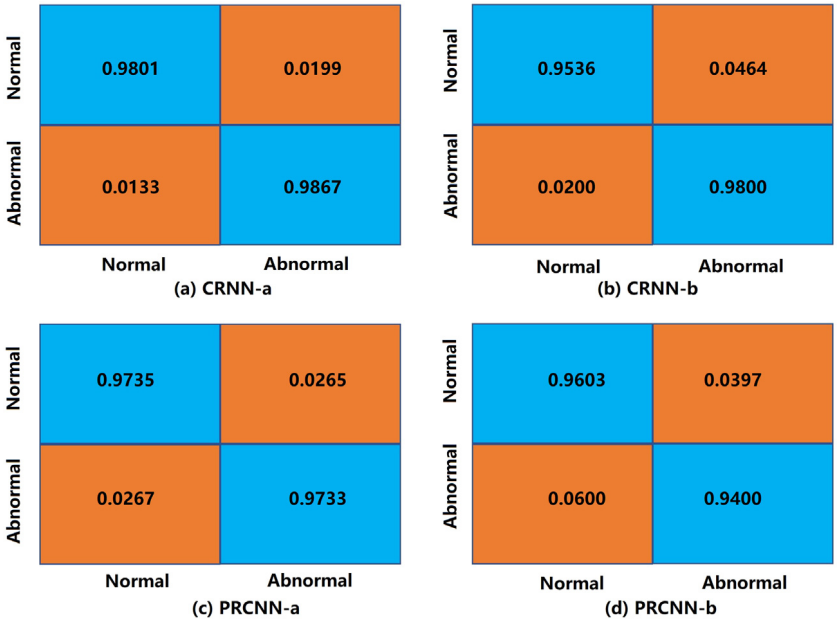

Fig. 13. The confusion matrix of the classification accuracy using the CRNN and PRCNN models.

Table 3
Qualitative performance comparisons with state-of-the-art algorithms. In this table, the number of training and test patterns used is different from each of the other methods.

Method	Sensitivity	Specificity	Accuracy
AdaBoost and CNN (Potes et al., 2016)	94.24%	77.81%	–
ANNs (Zabihi et al., 2016)	86.91%	84.90%	–
Wavelet-based CNN (Tschannen et al., 2016)	85.50%	85.90%	–
CNN (Rubin, et al., 2017)	72.78%	95.21%	–
DNNs (Nassralla et al., 2017)	63.00%	82.00%	–
Proposed CRNN-a method	98.66%	98.01%	98.34%
Proposed PRCNN-a method	97.33%	97.33%	97.34%

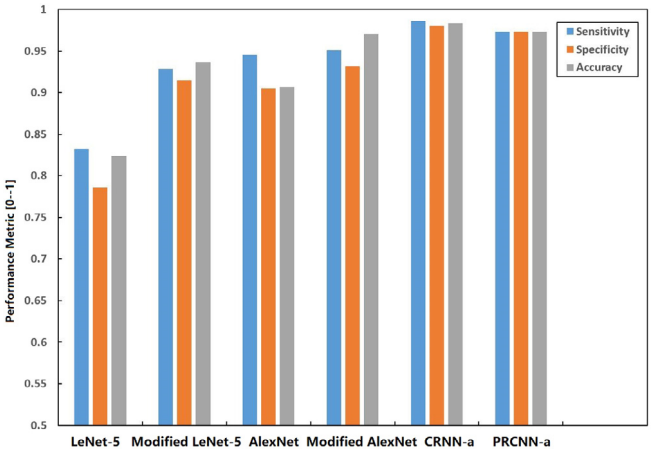

Fig. 14. Quantitative performance comparisons with state-of-the-art algorithms.

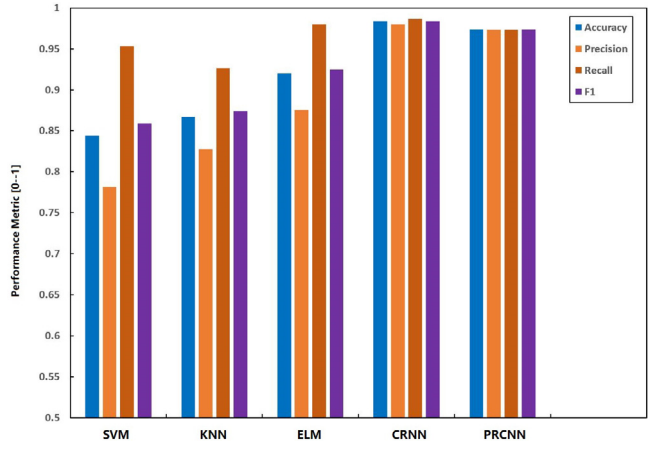

Fig. 15. Performance comparison with other classifiers based on MFCC features.

(RBF) are adopted. For ELM, the sigmoid function is used as the activation function and the number of hidden nodes is optimized from 50 to 2000. Fig. 15 shows the detailed experimental results. It is noted that the proposed CRNN and PRCNN methods significantly outperform other classifiers. Additionally, we apply the log-spectrogram as heart sound features, and implement the extraction of feature sets in Potes et al. (2016), Zabihi et al. (2016) and Nassralla et al. (2017). Based on the same CRNN classification scheme, we present the experimental results in Fig. 16. It can be clearly seen that the proposed MFCC-based features perform better than other feature sets and achieve promising performance.

3.4. Computational complexity analysis

The proposed algorithm is implemented on an Intel Core i5 CPU, 3.5 GHz, 8.00 GB RAM PC by using an Anaconda environment. For the heart sound classification task, preprocessing, MFCC feature extraction, the construction and calculation of deep learning networks are needed. Compared with the neural computation load of convolutional recurrent neural networks construction, the complexity of data preprocessing and MFCC features extraction is almost negligible. Specifically, the proposed heart sound classification algorithm consists of an off-line training phase and an on-line classification phase. In the off-line deep learning training phase, it takes on average 3 h in the CRNN training, and 3.37 h in the PRCNN training. In the on-line classification phase, the average classification time is 2.5 s and 1.9 s for CRNN and PRCNN, respectively. To further improve the real-time performance, the computational complexity can fortunately be reduced by using the Graphical Processing Units (GPU), 1080Ti. The average training time can be reduced to 1.52 h for CRNN training, and 1.77 h for PRCNN training. For the on-line classification, the average classification time is 1.2 s and 0.9 s for CRNN and PRCNN models.

4. Discussion

From the results obtained above, the following observations can be obtained:

The proposed framework facilitates the applications of automatic heart signal classification in practice. An “abnormal” result in the proposed method can be the focus of the physicians attention especially in small primary health care clinics. A “normal” result can be used for early discharge of healthy subjects, which reduces unnecessary waste of medical resources.

The proposed method achieves superior performance compared with existing other acoustic feature extraction methods. An enhanced feature extraction algorithm based on MFCC has been developed, in which the dynamic variations underlying the time-varying heart sound signals are explored. The fusion of three different MFCC features can provide a comprehensive characterization of heart sound dynamics. One of the most important reasons for the promising experimental results of our method is that our method makes full use of acoustic static and dynamics information for optimal classification performance.

The proposed deep learning framework is designed to enhance the classification accuracy of MFCC-based signal representation. The CNN and RNN are designed to complement each other in the heart sound classification process. The proposed method takes advantage of the spatial characteristics extracted from the CNN and the temporal characteristics extracted from the RNN, leading to a superior performance in the combined use of spatial and temporal features. The results using the proposed CRNN classification scheme are better than those using other classifiers. The combined use of the MFCC-based signal representation and CRNN classification scheme improves the classification accuracy.

It can be clearly seen that the proposed CRNN and PRCNN models outperform existing other deep learning models on heart sound classification with a considerable improvement. The proposed algorithm is shown to be effective in heart sound classification.

5. Conclusions

In clinical practice, it is difficult to arouse attention and vigilance of people with latent diseases, therefore, many suspected patients miss their best treatment time. To construct an efficient early detection system for cardiac disorders, we investigated the heart-sound-signal-based diagnostic tool in this paper. In fact,

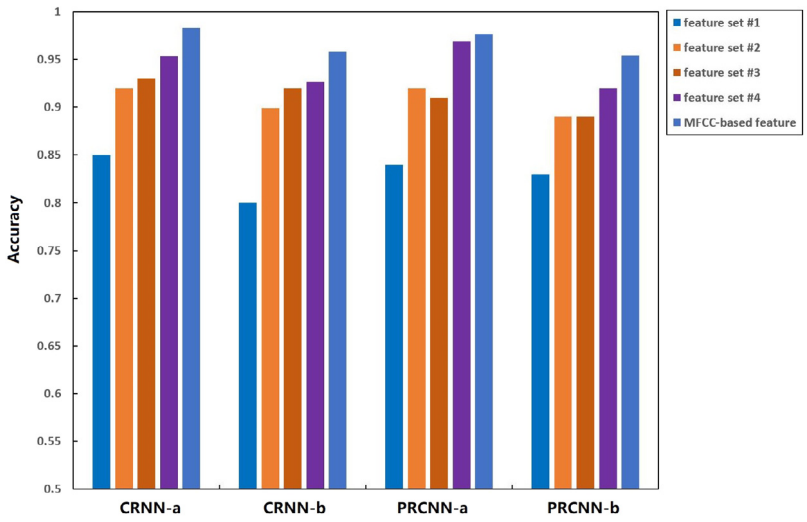

Fig. 16. Performance comparison with other heart sound features based on the proposed CRNN classification scheme. Here, feature set #1, set #2, set #3 and set #4 represent the log-spectrogram, the feature sets in Potes et al. (2016), Zabihi et al. (2016) and Nassralla et al. (2017), respectively.

the heart sound examination still remains the main noninvasive tool that cardiologists use first in today's clinical practice. A new improved MFCC feature extraction scheme is presented to characterize the dynamics of acoustics, in which the first and second order difference features of MFCC, are extracted to elaborate the dynamic feature of acoustics among consecutive heart sound signals. A heart sound classification algorithm that combines CNN and RNN for feature learning and classification is proposed, which takes advantage of the encoded local characteristics extracted from the CNN and the long-term dependencies captured by the RNN. Compared with the classification results in existing other methods, the proposed method achieves superior performance. An abnormal result in the proposed heart sound classification algorithm may be the focus of the physicians attention especially in the case of asymptomatic heart diseases. These suspected patients can choose to accept following CT or coronary angiography examination. The proposed method is expected to provide a complementary diagnostic tool in the first assessment of suspected patients.

Future research and development directions can be summarized as the following aspects. A further analysis of the proposed method along with a larger database is needed. More representative sound features should be extracted for performance improvement. In addition, different heart diseases are frequently encountered and the multi-class classification problem for the heart sound signals is another important issue in the diagnostic system.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61803133, U1909209, 61806062.

References

Ali, S., Adnan, S., Nawaz, T., Ullah, M. O., & Aziz, S. (2017). Human heart sounds classification using ensemble methods. *University of Engineering and Technology Taxila. Technical Journal*, *22*(1), 113.

Astuti, W., Sediono, W., Aibinu, A., Akmeliawati, R., & Salami, M. (2012). Adaptive short time fourier transform (STFT) analysis of seismic electric signal (SES): A comparison of hamming and rectangular window. In *2012 IEEE symposium on industrial electronics and applications* (pp. 372–377). IEEE.

Balili, C. C., Sobrepena, M. C. C., & Naval, P. C. (2015). Classification of heart sounds using discrete and continuous wavelet transform and random forests. In *2015 3rd IAPR Asian conference on pattern recognition* (pp. 655–659). IEEE.

Bozkurt, B., Germanakis, I., & Stylianou, Y. (2018). A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection. *Computers in Biology and Medicine*, *100*, 132–143.

Chen, Q., Zhang, W., Tian, X., Zhang, X., Chen, S., & Lei, W. (2016). Automatic heart and lung sounds classification using convolutional neural networks. In *2016 Asia-pacific signal and information processing association annual summit and conference* (pp. 1–4). IEEE.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., & Schwenk, H. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Clifford, G. D., Liu, C., Moody, B., Millet, J., Schmidt, S., & Li, Q. (2017). Recent advances in heart sound analysis. *Physiological Measurement*, *38*(8), E10–E25.

De Vos, J. P., & Blanckenberg, M. M. (2007). Automated pediatric cardiac auscultation. *IEEE Transactions on Bio-Medical Engineering*, *54*(2), 244–252.

Deng, S.-W., & Han, J.-Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Generation Computer Systems*, *60*, 13–21.

Dominguez-Morales, J. P., Jimenez-Fernandez, A. F., Dominguez-Morales, M. J., & Jimenez-Moreno, G. (2018). 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.

Emmanuel, B. S. (2012). A review of signal processing techniques for heart sound analysis in clinical diagnosis. *Journal of Medical Engineering and Technology*, *36*(6), 303–307.

Ergen, T., & Kozat, S. S. (2017). Online training of LSTM networks in distributed systems for variable length data sequences. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(10), 5159–5165.

Gaikwad, K. M., & Chavan, M. S. (2014). Removal of high frequency noise from ECG signal using digital IIR butterworth filter. In *Wireless Computing and Networking* (pp. 121–124). IEEE.

Georgios, D., Fernando, B., & Felix, L. (2018). Improving imbalanced learning through a heuristic oversampling method based on K-means and SMOTE. *Information Sciences*, *465*, 1–20.

Gupta, C. N., Palaniappan, R., Swaminathan, S., & Krishnan, S. M. (2007). Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing*, *7*(1), 286–297.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Kumar, D., Carvalho, P., Antunes, M., Henriques, J., Eugenio, L., & Schmidt, R. (2006). Detection of S1 and S2 heart sounds by high frequency signatures. In *International conference of the IEEE engineering in medicine biology society*. (pp. 1410–1416).

Lam, M., Lee, T., Boey, P., Ng, W., Hey, H., & Ho, K. (2005). Factors influencing cardiac auscultation proficiency in physician trainees. *Singapore Medical Journal*, *46*(1), 11–14.

Latif, S., Usman, M., Rana, R., & Qadir, J. (2018). Phonocardiographic sensing using deep learning for abnormal heartbeat detection. *IEEE Sensors Journal*, *18*(22), 9393–9400.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., & Chorro, F. J. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, *37*(12), 2181–2213.

Members, W. G., Lloyd-Jones, D., Adams, R. J., Brown, T. M., Carnethon, M., & Dai, S. (2010). Heart disease and stroke statistics–2010 update: a report from the american heart association. *Circulation*, *127*(1), 143–152.

Members, W. G., Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., & Blaha, M. J. (2016). Heart disease and stroke statistics-2016 update: A report from the American Heart Association. *Circulation*, *133*(4), e38–e360.

Monali, U. M., & Aparana, R. S. (2012). Review on heart sound analysis technique. In *International conference on internet computing and information communications, (vol. 216)*. (pp. 93–101).

Nassralla, M., El Zein, Z., & Hajj, H. (2017). Classification of normal and abnormal heart sounds. In *2017 Fourth international conference on advances in biomedical engineering* (pp. 1–4). IEEE.

Nilanon, T., Yao, J., Hao, J., Purushotham, S., & Liu, Y. (2016). Normal/abnormal heart sound recordings classification using convolutional neural network. In *2016 Computing in cardiology conference* (pp. 585–588). IEEE.

Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016). Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *2016 Computing in cardiology conference*, (pp. 621–624). IEEE.

Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., & Sricharan, K. (2017). Recognizing abnormal heart sounds using deep learning. arXiv preprint arXiv:1707.04642.

Schmidt, S. E., Holst-Hansen, C., Hansen, J., Toft, E., & Struijk, J. J. (2015). Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2611–2619.

Sh-Hussain, H., Mohamad, M., Zahilah, R., Ting, C.-M., Ismail, K., & Numanl, F. (2017). Classification of heart sound signals using autoregressive model and hidden Markov model. *Journal of Medical Imaging and Health Informatics*, *7*(4), 755–763.

Sigtia, S., Benetos, E., & Dixon, S. (2016). An end-to-end neural network for polyphonic piano music transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(5), 927–939.

Springer, D. B., Tarassenko, L., & Clifford, G. D. (2016). Logistic regression-HSMM-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, *63*(4), 822–832.

Stasis, A. C., Loukis, E. N., Pavlopoulos, S. A., & Koutsouris, D. (2003). Using decision tree algorithms as a basis for a heart sound diagnosis decision support system. In *International conference on information technology applications in biomedicine*. (pp. 354–357).

Strunic, S. L., Rios-Gutierrez, F., Alba-Flores, R., Nordehn, G., & Burns, S. (2007). Detection and classification of cardiac murmurs using segmentation techniques and artificial neural networks. In *2007 IEEE symposium on computational intelligence and data mining* (pp. 397–404). IEEE.

Sun, S., Jiang, Z., Wang, H., & Fang, Y. (2014). Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform. *Computer Methods and Programs in Biomedicine*, *114*(3), 219–230.

Thomae, C., & Dominik, A. (2016). Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound. In *2016 Computing in cardiology conference* (pp. 625–628). IEEE.

Trang, H., Loc, T. H., & Nam, H. B. H. (2014). Proposed combination of PCA and MFCC feature extraction in speech recognition system. In *2014 International conference on advanced technologies for communications* (pp. 697–702). IEEE.

Tschannen, M., Kramer, T., Marti, G., Heinzmann, M., & Wiatowski, T. (2016). Heart sound classification using deep structured features. In *2016 Computing in cardiology conference* (pp. 565–568). IEEE.

Wang, P., Lim, C. S., Chauhan, S., Foo, J. Y. A., & Anantharaman, V. (2007). Phonocardiographic signal analysis method using a modified hidden Markov model. *Annals of Biomedical Engineering*, *35*(3), 367–374.

Yang, X., Yang, F., Gobeawan, L., Yeo, S. Y., Leng, S., & Zhong, L. (2016). A multimodal classifier for heart sound recordings. In *2016 Computing in cardiology conference* (pp. 1165–1168). IEEE.

Yuenyong, S., Nishihara, A., Kongprawechnon, W., & Tungpimolrut, K. (2011). A framework for automatic heart sound analysis without segmentation. *Biomedical Engineering Online*, *10*(1), 1–23.

Zabihi, M., Rad, A. B., Kiranyaz, S., Gabbouj, M., & Katsaggelos, A. K. (2016). Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In *2016 Computing in cardiology conference* (pp. 613–616). IEEE.

Zeng, D., Liu, K., Lai, S., Zhou, G., & Zhao, J. (2014). Relation classification via convolutional deep neural network. In *2014 International conference on computational linguistics*. (pp. 2335–2344).

Zhang, S., Zheng, D., Hu, X., & Yang, M. (2015). Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th pacific asia conference on language, information and computation*. (pp. 73–78).