IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 27, 2019, accepted July 22, 2019, date of publication July 29, 2019, date of current version August 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931500

Automatic Cardiac Arrhythmia Classification Using Combination of Deep Residual Network and Bidirectional LSTM

RUNNAN HE[1], YANG LIU[1], KUANQUAN WANG[1], (Senior Member, IEEE), NA ZHAO[1], YONGFENG YUAN[1], QINCE LI[1], AND HENGGUI ZHANG[1,2,3,4]

[1]School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin 150001, China
[2]School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, U.K.
[3]SPACEnter Space Science and Technology Institute, Shenzhen 518117, China
[4]International Laboratory for Smart Systems and Key Laboratory of Intelligent of Computing in Medical Image, Ministry of Education, Northeastern University, Shengyang 110004, China

Corresponding authors: Qince Li (qinceli@hit.edu.cn) and Henggui Zhang (henggui.zhang@manchester.ac.uk)

This work was supported in part by the National Science Foundation of China (NSFC) under Grant 61572152 (to HZ), Grant 61571165 (to KW), Grant 61601143 (to QL), and Grant 81770328 (to QL), in part by the Science Technology and Innovation Commission of Shenzhen Municipality under Grant JSGG20160229125049615 and Grant JCYJ20151029173639477 (to HZ), and in part by the China Postdoctoral Science Foundation under Grant 2015M581448 (to QL).

ABSTRACT Cardiac arrhythmia is associated with abnormal electrical activities of the heart, which can be reflected by altered characteristics of electrocardiogram (ECG). Due to the simplicity and non-invasive nature, the ECG has been widely used for detecting arrhythmias and there is an urgent need for automatic ECG detection. Up to date, some algorithms have been proposed for automatic classification of cardiac arrhythmias based on the features of the ECG; however, their stratification rate is still poor due to unreliable features of signal characteristics or limited generalization capability of the classifier, and therefore, it remains a challenge for automatic diagnosis of arrhythmias. In this paper, we propose a new method for automatic classification of arrhythmias based on deep neural networks (DNNs). The two DNN models constitutive of residual convolutional modules and bidirectional long short-term memory (LSTM) layers are trained to extract features from raw ECG signals. The extracted features are concatenated to form a feature vector which is trained to do the final classification. The algorithm is evaluated based on the test set of China Physiological Signal Challenge (CPSC) dataset with F_1 measure regarded as the harmonic mean between the precision and recall. The resulting overall F_1 score is 0.806, F_{AF} score is 0.914 for atrial fibrillation (AF), F_{Block} score is 0.879 for block, F_{PC} and F_{ST} scores are 0.801 and 0.742 for premature contraction and ST-segment change, which demonstrates a good performance that may have potential practical applications.

INDEX TERMS Cardiac arrhythmia, electrocardiogram (ECG), deep neural networks (DNNs), deep residual network, bidirectional long short-term memory (LSTM).

I. INTRODUCTION

Cardiac arrhythmias are a group of conditions in which the electrical activity of the heart is irregular, manifesting faster or slower rhythm than that under normal condition [1]. Many types of arrhythmia are life threatening, and possibly caused by various cardiac diseases such as myocardial infarction, cardiomyopathy, and myocarditis [2]–[4]. Detailed analysis of the electrocardiogram (ECG) signal provides functional information on the patient's heart, which has been extensively used for arrhythmia detection in clinical practices. In general, analyzing ECG features contains two processing steps. The first step is to extract features of ECGs and the second one is to classify the ECGs into various conditions based on these extracted features [5]. As it is very time consuming and tedious for analyzing ECG features manually, it is necessary to develop automatic algorithms for ECG analysis.

In the past few years, numerous detection algorithms of cardiac arrhythmias have been proposed. These algorithms mainly consist of four main procedures including denoising [6]–[10], waveform detection [11]–[13], feature extraction and arrhythmia classification [14]–[28]. Among these

The associate editor coordinating the review of this manuscript and approving it for publication was Yongqiang Cheng.

four steps, feature extraction transforms the input ECG signal into a variety of features that play an important role in detecting most of cardiac arrhythmias. In recent studies, different features of ECGs have been extracted to characterize their properties that include Hermit coefficients [21], [24], [25], higher order statistical features [18], [25], morphological features [17], [26], [27], wavelet features and independent component analysis [12], [19], [20], [28]. After feature extraction, various machine detection algorithms such as linear discrimination analysis (LDA) [15], [24], self-organizing map (SOM) [21], conditional random field (CRF) [18], support vector machine (SVM) [12], [21], [25], artificial neural network (ANN) [20], [24], and ensemble methods [26] have been implemented for ECG classification.

Recently, deep neural networks (DNNs) have also been used in ECG classifications. As it is different from traditional methods, DNNs can learn a feature extraction function from the raw input based on the probability distribution of the dataset. Therefore, on the premise of sufficient training samples, the features extracted by a DNN model can be more comprehensive than those extracted by hand-crafted methods. For example, in the detection of ventricular arrhythmias, the stacked denoising autoencoders (SDAEs) was applied to learn a suitable feature mapping, after that, a softmax regression layer was added on the top of the resulting hidden representation layer to yield DNNs [29]. In parallel, a convolutional neural network (CNN) has also been used for automatic diagnosis of cardiac arrhythmias. In a previous study we have proposed CNN for atrial fibrillation (AF) detection [30]. Furthermore, a multiscale fusion of deep convolutional neural network (MS-CNN) was also proposed to screen out AF, which employs two-stream convolutional networks with different filter sizes to extract features of different scales [31]. In another study, a CNN with the residual network architecture to classify 12 rhythm classes was introduced [32]. CNNs have also been applied for beat-level arrhythmias classification where the length of model inputs are usually much shorter (e.g., hundreds of samples) [33], [34], and a nine-layer CNN was developed to automatically identify five different categories of heartbeats [35]. Moreover, other kinds of network structure used for ECG classification include restricted Boltzmann machine (RBM) [36]–[38] and Autoencoder [39], [40], for the latter of which a novel deep learning-based algorithm was introduced that integrates a long short-term memory (LSTM) based autoencoder network with SVM for arrhythmias classification [40]. With variable ECG lengths, an automated arrhythmia classification using a combination of CNN and LSTM was proposed [41]. Finally, a deep learning based ensemble network model was put forward for improving the performance, which can occur on a single network. Therefore, the model is designed to combine three single networks to capture features and classify [42].

Although some of the algorithms achieved reasonable performance for automatic ECG classification of arrhythmias, there are still unsolved challenges for practical uses. First, though extracted individual ECG features may provide some useful clinical information for automatic identification of cardiac arrhythmias, however, ECG signals of different patients under different physical conditions many have distinctive morphological and temporal features [43]. Therefore, ECG signal may vary its feature for each person at different timing, and different patients with the same disease may have diverse ECG morphologies. Furthermore, different cardiac diseases may have some common ECG features, forming a significant barrier for heart disease diagnosis by feature extraction and analysis [44]–[46]. Second, the current arrhythmia classification methods are developed based on limited training data sets, lacking consideration of the specific features of the rhythm of each patient that may be different to the training data. Therefore, the current algorithms may not perform well in practical condition. To address this problem, long-term ECG recordings for a patient can be obtained with the wide use of long-term ECG monitoring device, making it possible to develop automatic classification algorithms.

The objective of this study was to tackle of the aforementioned challenges by developing a reliable method for fully automated classification of arrhythmias combining deep residual network and bidirectional LSTM. The developed method is tested using ECG recordings provided by the China Physiological Signal Challenge (CPSC) [47]. An advantage of the proposed algorithm is that there is no need to manually extract features of ECGs compared with other traditional arrhythmia classification methods [14]–[28]. Based on the divided raw ECG signals, the proposed DNNs can automatically extract their features with proper training, based on which most of arrhythmia types can be identified. Moreover, without filtering, the proposed method is resistant to interferences, making it more suitable for practical applications. Thus, the method demonstrated high efficiency and averaged accuracy to classify different arrhythmias based on the CPSC database.

The rest of this paper is organized as follows. In Section 2, the ECG dataset and the proposed method are described in details. In Section 3, the proposed algorithm is evaluated by CPSC database and discussed. Finally, section 4 summarizes our study.

II. MATERIALS AND METHODS

The flowchart diagram of the proposed arrhythmia classification method is shown in Figure 1. It includes two stages, which are preprocessing as well as feature extraction and classification.

A. MATERIALS

The 12 leads ECG recordings from the CPSC database consist of about 10,000 recordings, which are used for the evaluation of arrhythmia classification allowing a direct comparison with other participant's results. Each recording has an uncertain length ranging from 6 to 60 seconds, which is sampled at 500 Hz (Fs=500Hz). The ECG recordings contain nine types such as atrial fibrillation (AF) and premature atrial contraction (PAC), etc. The training and test sets include 6,877 and

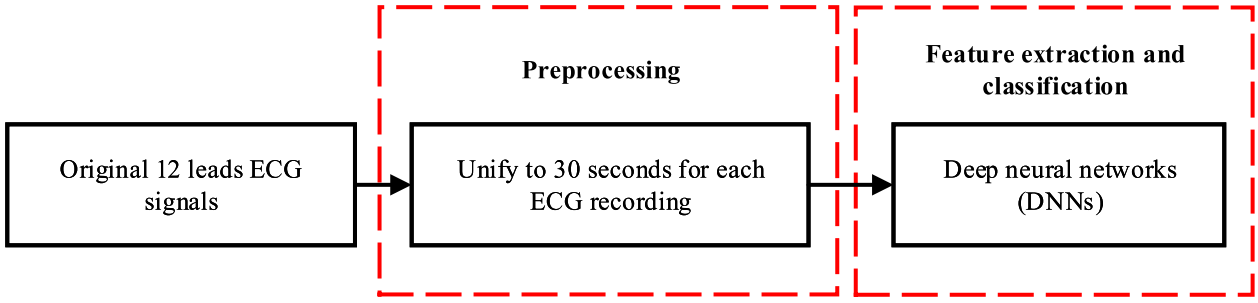


FIGURE 1. Flowchart diagram of the proposed algorithm for arrhythmia classification.

TABLE 1. The training set according to the 'Frist label' annotations. Mean, SD, Min, Median and Max in the header indicate the mean, standard deviation, minimum, median and maximum of the time lengths of the recordings, respectively [47].

Type	recording	Mean	SD	Min	Median	Max
Normal	918	15.43	7.61	10.00	13.00	60.00
Atrial fibrillation (AF)	1098	15.01	8.39	9.00	11.00	60.00
First-degree atrioventricular block (I-AVB)	704	14.32	7.21	10.00	11.27	60.00
Left bundle branch block (LBBB)	207	14.92	8.09	9.00	12.00	60.00
Right bundle branch block (RBBB)	1695	14.42	7.60	10.00	11.19	60.00
Premature atrial contraction (PAC)	556	19.46	12.36	9.00	14.00	60.00
Premature ventricular contraction (PVC)	672	20.21	12.85	6.00	15.00	60.00
ST-segment depression (STD)	825	15.13	6.82	8.00	12.78	60.00
ST-segment elevated (STE)	202	17.15	10.72	10.00	11.89	60.00
Total	6877	15.79	9.04	6.00	12.00	60.00

2,954 ECG recordings respectively. Every ECG signal has a labeled annotation. The most of the recordings only have one label (denoted as First label). However, some recordings have two or three labels. For the recordings which have more than one label, the classification result is considered right if one of them is correct. The details of training set are described in Table 1.

B. PREPROCESSING

A typical ECG heartbeat is characterized by a recurrent sequence of waves including P, QRS and T waves which represent the depolarization of the atria and ventricles, followed by repolarization of the ventricles [48]. In this study, the ECG signals are not filtered in the preprocessing stage considering two main factors. On one hand, in this database, all the 12 leads recordings are used and the amount of the data is large. Using the original data rather than filtering them, the computation cost is significantly reduced. On the other hand, as mentioned above, most of arrhythmia detection algorithms have applied filtering to process ECG signals [6]–[10]. However, in this study, the proposed model shows good performance in anti-noise interference due to no filtering, demonstrating a potential for practical applications. Detailed comparison between results of the model obtained with and in absence of filtering is presented in Section 3.2.

As shown in Table 1, the range of each recording length varies from 6 to 60 seconds. As it is not convenient for training the model with non-identical length of ECG recordings, therefore, in implementation, each recording is segmented into a length of 30 seconds. If the recording length is less than 30 seconds, we pad the recording into 30 seconds by assigning zero values at the beginning period. If the length is more than 30 seconds, we leave out the extra data after 30 seconds. We have also applied other fixed recording length (such as 10 and 20 seconds), however, it proves that 30 seconds is the optimal choice. It is possible that the padding of zero for recordings shorter of 30 seconds may have some influence on the performance of the model, however, it is necessary for solving the problem of different input length of data and convenient for training the model, as well as helpful to reduce the training time of model.

C. THE BALANCE OF DATASET

After the preprocessing, the balance of dataset is also important for the classification of arrhythmias. In this study, some attention needs to be paid to deal with imbalance training dataset. As we can see in Table 1, the number of LBBB is only 207, however the number of RBBB reaches 1695 which is nearly eight times as many as LBBB, and the number of other types of arrhythmias are also different. It is obvious that training dataset is not balanced. To address this issue, we have randomly divided five subsets for each class. After that, each class is copied to make the number equal to the class with the most recordings, which is RBBB in this dataset. This operation has addressed the issue of dataset imbalance and meanwhile dealt with five subsets to perform five cross-validation.

D. FEATURE LEARNING AND CLASSIFICATION

Feature extraction is a critical step in the classification of raw ECG signals. The relevance and representativeness of the extracted features will directly influence the performance of classification. Traditional feature extraction methods are usually based on hand-crafted algorithms, while DNN can learn features from the raw data by itself and generally have better performance [32], [49], [50]. In this study, we employ a novel kind of DNNs to learn features from the preprocessed ECG recordings. Our proposed network is composed of three parts, namely local features learning part, global features leaning part and classification part, as illustrated in Figure 2.

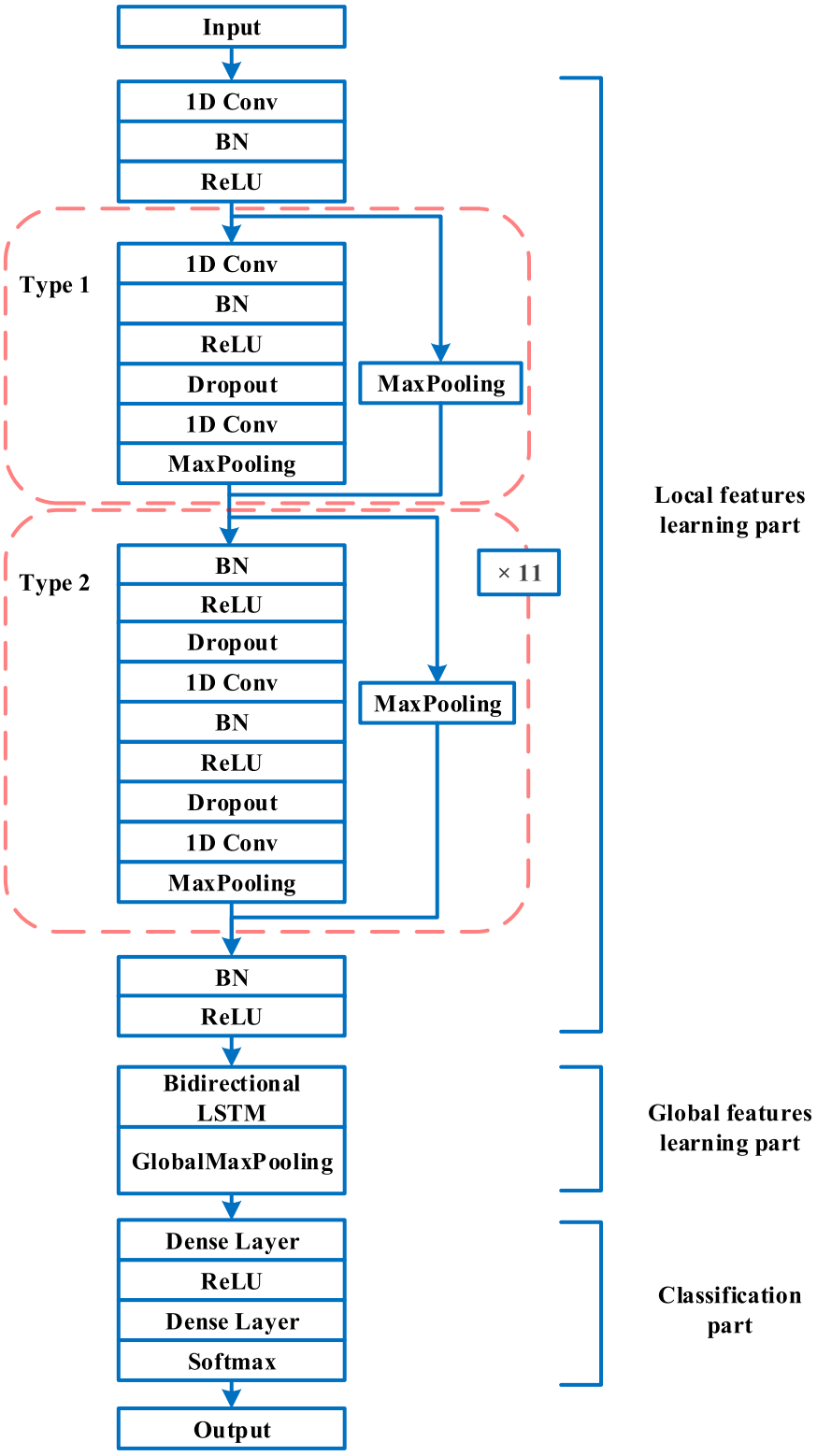


FIGURE 2. The overall structure of the proposed DNNs for cardiac arrhythmias classification.

In the local layer learning part, we utilize stacked residual convolutional modules to learn local features and compress the long course ECG signal into a much shorter sequence of local feature vectors. The input of this part is the raw ECG signal which is a 3-dimensional matrix with the dimensions determined as (batch size, 15000, 12). The batch size is set to

TABLE 2. The length/ number of convolution kernels and pool size of max-pooling layers in each residual modules.

Residual modules	Kernel length	Kernel number	Pool size
1^{st}	16	32	2
2^{nd}	16	32	1
3^{rd}	16	32	2
4^{th}	16	32	1
5^{th}	16	64	2
6^{th}	16	64	1
7^{th}	8	64	2
8^{th}	8	64	1
9^{th}	8	96	2
10^{th}	8	96	1
11^{th}	8	96	2
12^{th}	8	96	1

a large value as long as the GPU memory can accommodate the input and all the intermediate data. The length of the other two dimensions is the signal length and channel number (i.e., lead number) respectively. There are two types of residual convolutional modules adopted in our model from [32]. The first module in the network is in Type 1, while the follow-up 11 modules are in Type 2. As shown in Figure 2, both types of modules consist of 1-dimentional convolutional (1D Conv) layers, batch normalization (BN) [51] layers, rectified linear units (ReLU) [52] of activation layer, dropout [53] layers and max-pooling [54] layers. The difference between the two types of modules exists in the start positions of the residual connections: the module of Type 2 has three more layers (i.e., batch normalization, ReLU activation and dropout) than that of Type 1. Accordingly, there are a total of 25 convolutional layers and 12 max-pooling layers in this part. As the margin of the input will be lost during a convolutional operation, the input feature maps are padded before each convolutional layer so that the output has the same length as the original input. The feature maps are compressed in length only when they go through a pooling layer. As shown in Table 2, there are only 6 max-pooling layers (with pool size of 2) that actually work, others (with pool size of 1) have no effects to the feature maps. So, after the operations of local features learning part, output length is 1/64 of the input length. The convolution kernels also have different lengths and numbers among these residual modules, which can be found in Table 2.

After the local features learning part, the extracted feature vectors are input into a bidirectional LSTM [55] layer (the main body of the global features learning part) one by one, as indicated in Figure 3. As the ECG signals (also the learned local feature maps) represent the time course of heart electrical activity, a recurrent neural network (RNN) can be typically used to process the input along the time sequence in a parameters-sharing manner and utilizes their internal state to memorize the context. One of the successful implementations of RNN is LSTM. Due to the advantage of the gate designs of its unit, LSTM can maintain the properties of a longer temporal sequence spanning tens to hundreds of time steps in its internal state. The bidirectional layer in fact is composed of two LSTM layers in opposite directions: the forward LSTM and the backward LSTM. At each time step, the output of the forward LSTM summarizes local features from all previous steps, while the output of the backward LSTM summarizes local features from all subsequent steps. The outputs of the two LSTM layers are summed into a local-focused global feature vector which encapsulates features from the context of the current step in both forward and backward directions. The unit numbers of these two LSTM layers are all 64, which means each local-focused global feature vector is 64 in length. Then, all the local-focused global feature vectors are input into a global max-pooling layer to get a single global feature vector for the classification.

With the extracted global feature, the classification part learns a classifier to stratify recordings into different classes. The classification part consists of two dense layers and two activation layers. The first dense layer has 64 cells, while the second dense layer has 9 cells (corresponding to 9 classes respectively). The first activation layer, following the first dense layer, is a ReLU layer which enables the classification part to accelerate the back propagation of gradients. The second activation layer, also the final layer of the network, is a softmax layer which outputs the predicted probability distribution over the 9 classes.

III. RESULTS AND DISCUSSIONS

A. PERFORMANCE MEASURES

The experiments were performed on a computer with 1 CPUs at 3.5 GHz, 1 NVIDIA Quadro k6000 GPU and 64-Gb memory. All the proposed models are run over highly efficient GPU using the Keras deep learning framework [56]. In implementation, the proposed algorithm was tested on each data class from the CPSC database. For cross-validation, the training set was divided randomly into five subsets of which one for validation and the others for training, taking turns in 5 times.

The F_1 measure, an averaged F_1 values for all arrhythmia classes, is used to score the performance of the model based on the CPSC dataset. To calculate F_1, the following equations are used.

For each of all the types, F_1 is defined in Equation (1),

$$F_{1x} = \frac{2N_{xx}}{N_{Xx} + N_{xX}}, \quad x = 1, 2...9 \tag{1}$$

where F_{1x} indicates F_1 score of the xth class, N_{xX}, N_{Xx} and N_{xx} indicate the numbers of real, predicted and true positive instances of the xth class, respectively.

The final score is defined in Equation (2),

$$F_1 = \frac{F_{11}+F_{12}+F_{13}+F_{14}+F_{15}+F_{16}+F_{17}+F_{18}+F_{19}}{9} \tag{2}$$

In addition, the F_1 scores for each of the four sub-abnormal types are computed as follows:

$$F_{AF} = \frac{2N_{22}}{N_{2X} + N_{X2}} \tag{3}$$

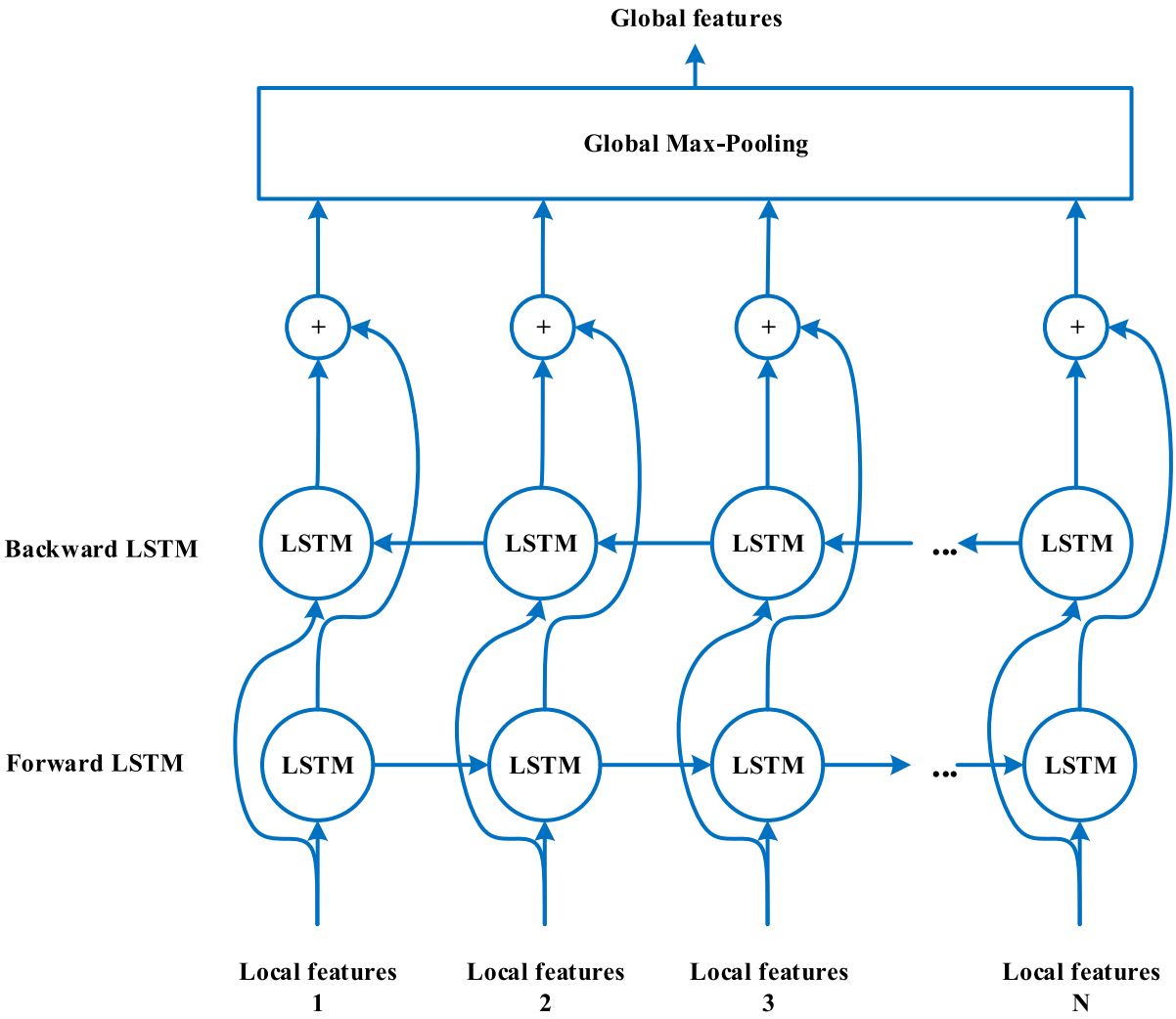


FIGURE 3. The structure unfolded along the time sequence of global features learning part.

$$F_{Block} = \frac{2(N_{33} + N_{44} + N_{55})}{N_{3X} + N_{X3} + N_{4X} + N_{X4} + N_{5X} + N_{X5}} \quad (4)$$

$$F_{PC} = \frac{2(N_{66} + N_{77})}{N_{6X} + N_{X6} + N_{7X} + N_{X7}} \quad (5)$$

$$F_{ST} = \frac{2(N_{88} + N_{99})}{N_{8X} + N_{X8} + N_{9X} + N_{X9}} \quad (6)$$

Using classification results of the validation set, we calculated overall F_1 score for all cases to evaluate the classification results. The results show that the overall F_1 score of the proposed DNNs classifier with 450 test samples reached 0.836, which is the highest among all competitors. The detailed classification results are shown in Table 3 and 4 (http://2018.icbeb.org/Challenge.html).

Results shown in Table 3 illustrate that the classifier achieves a good performance by using relatively small datasets, demonstrating that the classifier can cope with different conditions. The scores of all arrhythmia types are close to or greater than 0.78, especially the score of AF is the highest one, 0.92. However, the identification of normal is less good, only reaching about 0.748.

TABLE 3. Results of the arrhythmia classification based on 450 test samples.

Rank	Team	Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE
1	**He et al.**	**0.748**	**0.920**	**0.882**	**0.889**	**0.883**	**0.787**	**0.851**	**0.780**	**0.780**
2	Cai et al.	0.765	0.927	0.887	0.886	0.880	0.812	0.800	0.784	0.753
3	Chen et al.	0.752	0.930	0.871	0.915	0.839	0.832	0.833	0.800	0.667
4	Mao et al.	0.692	0.940	0.852	1.000	0.899	0.706	0.875	0.762	0.622
5	Yu et al.	0.709	0.907	0.863	0.918	0.838	0.736	0.783	0.714	0.723
6	Zhang et al.	0.757	0.904	0.839	0.887	0.787	0.735	0.755	0.742	0.638

The F_1 scores of four sub-abnormal types were summarized in Table 4. It shows that the proposed model also achieves better recognition in specific arrhythmia classification scenario. The achieved F_1 scores of AF, Block, Premature contraction and ST-segment change are 0.92, 0.884, 0.821 and 0.78, respectively. From the results, the proposed model shows a high-level capability of identifying AF, block and Premature contraction. The score of ST-segment change

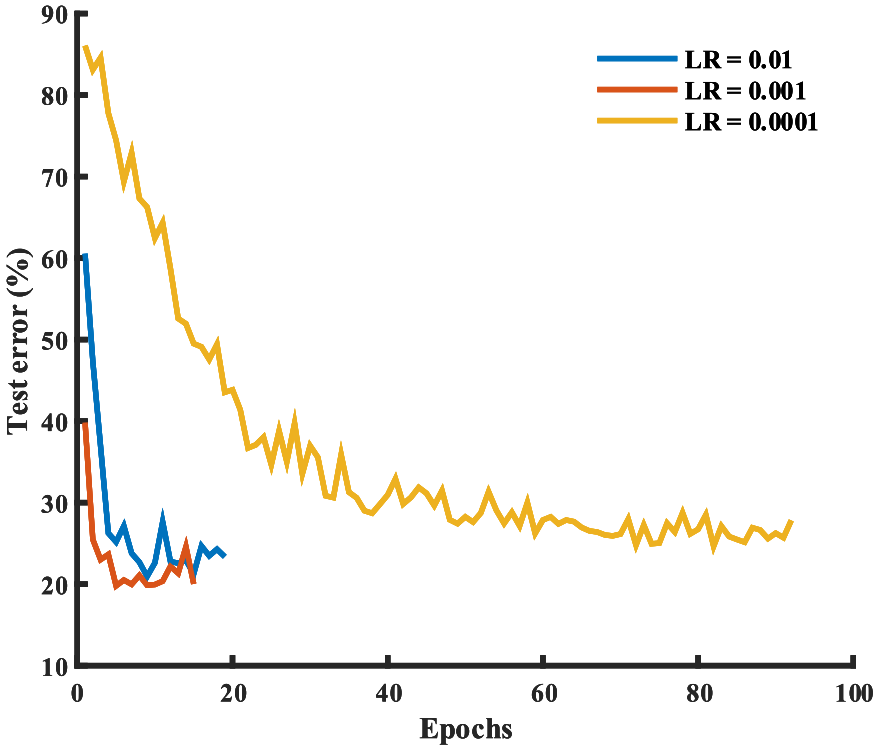


FIGURE 4. Test samples classification errors with different learning rates (LR).

TABLE 4. The F_1 scores of four sub-abnormal types based on 450 test samples.

Rank	Team	F_{af}	F_{block}	F_{pc}	F_{st}	F_1
1	**He et al.**	**0.920**	**0.884**	**0.821**	**0.780**	**0.836**
2	Cai et al.	0.927	0.884	0.806	0.770	0.833
3	Chen et al.	0.930	0.868	0.832	0.738	0.827
4	Mao et al.	0.940	0.898	0.800	0.704	0.816
5	Yu et al.	0.907	0.871	0.758	0.719	0.799
6	Zhang et al.	0.904	0.833	0.745	0.699	0.783

TABLE 5. The F_1 scores of four sub-abnormal types based on the entire test set.

Rank	Team	F_{af}	F_{block}	F_{pc}	F_{st}	F_1
1	Chen et al.	0.933	0.899	0.847	0.779	0.837
2	Cai et al.	0.931	0.912	0.817	0.761	0.830
3	**He et al.**	**0.914**	**0.879**	**0.801**	**0.742**	**0.806**
4	Yu et al.	0.918	0.890	0.789	0.7018	0.802
5	Yan et al.	0.924	0.882	0.779	0.709	0.791
6	Chen et al.	0.905	0.902	0.722	0.708	0.783

is a little lower, possibly resulting from its undistinguishable features with many normal conditions requiring improvement in the future.

For the entire test set (about 3,000), the F_1 scores of four sub-abnormal types were summarized in Table 5 (http://2018.icbeb.org/Challenge.html). By comparing results shown in Table 4 and 5, we can see that the performance of the proposed model on the entire test set declines slightly. This may be attributable to the number and replication of recordings implemented in our model. The use of simple copying of recordings may introduce differences in training and test data distribution, resulting in bias. However, it is necessary for solving the impact of data imbalance. Furthermore, the overall performance is still better than most of competitors, which achieved the third place in the CPSC2018 challenge.

B. FINE TUNING OF HYPERPARAMETERS

To obtain an optimal DNNs structure to classify arrhythmias, the impacts of diverse training and structural hyperparameters are quantitatively analyzed. By doing so, the final error values of test samples are assessed under various comparative conditions. As hyperparameters and their combinations are tremendous, we chose some of them empirically in our experiments for balancing the model performance and training time. The chosen hyperparameters include learning rate, dropout rate, convolutional kernel size, LSTM units number, LSTM dropout rate, LSTM recurrent dropout rate and dense layer cells number. Learning rate controls the speed at which the model updates its weights during the training process. The LSTM dropout rate is fraction of the units to drop for the linear transformation of the inputs, while the LSTM recurrent dropout rate is that of the units to drop for the linear transformation of the recurrent state. After the experiments with selected combinations of these hyperparameters, the final DNN hyperparameters which obtain the minimum test error are shown in Table 6. The proposed model was trained for about 20 epochs with one cross-validation. The total training time of our network was about 5 hours,

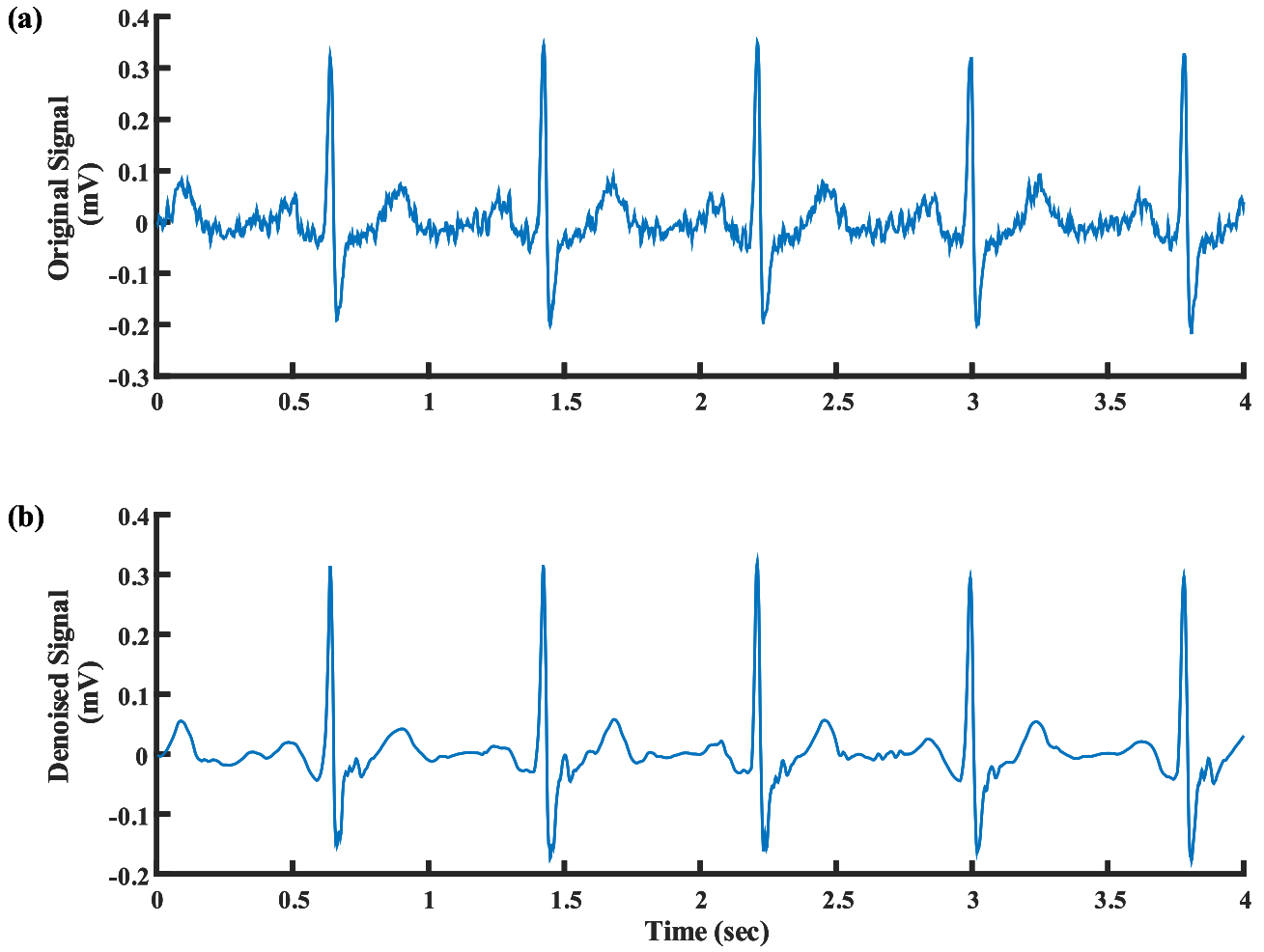


FIGURE 5. The effect of denoising method used in our control experiments. (a) The original ECG signal. (b) The denoised signal.

TABLE 6. The optimal DNNs parameters for arrhythmias detection.

Hyperparameters	Optimal values
Learning rate	0.001
Dropout rate	0.5
Convolutional layer kernel size	16
No. of LSTM units	64
LSTM dropout	0.5
LSTM recurrent dropout	0.1
No. of dense layer cells	64
Computational complexity	10

with a computational complexity (regarded as the decision time for each recording in this study) was 10ms.

For a clear vision of the impacts of hyperparameters, the training and test errors at different training and structural hyperparameters are discussed in the following analysis.

1) Learning Rate: In this case, different learning rates (0.01, 0.001 and 0.0001) are used to update the model weights. As shown in Figure 4, the curves of test errors at different learning rates exhibit significant differences. Among these settings, the test error has the fastest convergence rate and minimum convergence value when the learning rate is 0.001. Therefore, we chose 0.001 as the optimum learning rate for the training of our model.

2) Hyperparameters of Local Features Learning Part: The local features learning part contains dozens of hyperparameters making the tuning a long and expensive process. Therefore, we chose a few key hyperparameters to tune in the experiments, including convolutional layer kernel size and dropout rate. The convergent training and test errors at different combination of hyperparameters are presented in Table 7. As the results indicated, the model achieves the minimum test error among all the combinations when convolutional layer kernel size and dropout rate are 16 and 0.5 respectively. And this combination is reused in the follow-up experiments for tuning other parts of the neural network.

3) Hyperparameters of Global Features Learning Part: After the local features learning part was fine-tuned, the global features learning part are tuned by a series of experiments. Three hyperparameters, namely number of LSTM units, LSTM dropout and LSTM recurrent dropout, are selected to generate various hyperparameter combinations

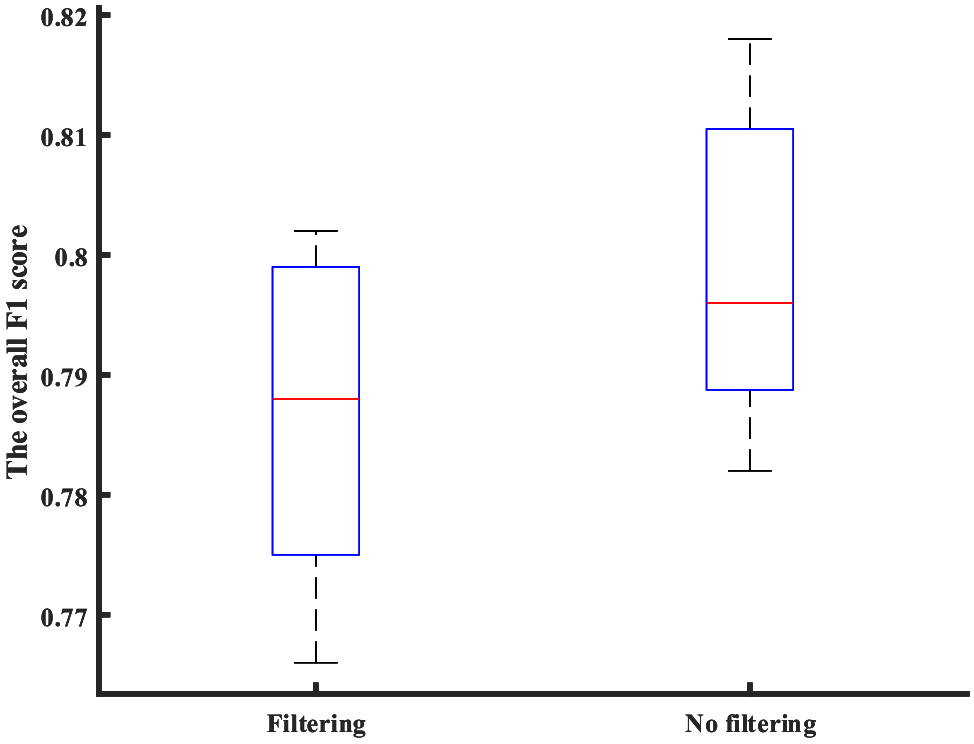


FIGURE 6. The overall F_1 scores of our models with and without filtering preprocessing in a 5-fold cross-validation experiments.

TABLE 7. Training and test error at different hyperparameter settings for local features learning.

Convolutional layer kernel size	Dropout rate	Training Error (%)	Test Error (%)
8	0.5	7.92	21.75
8	0.5	9.88	20.36
8	0.5	11.36	19.16
16	**0.5**	**11.02**	**18.22**
16	0.4	10.03	20.53
16	0.6	14.60	21.11

TABLE 8. Training and test error at different hyperparameter settings for global features learning.

No. of LSTM units	LSTM dropout	LSTM recurrent dropout	Training Error (%)	Test Error (%)
64	0.1	0.1	11.56	18.64
64	0.1	0.3	14.41	19.94
64	0.1	0.5	18.90	18.69
64	0.3	0.1	9.51	19.50
64	0.3	0.5	18.75	20.14
64	**0.5**	**0.1**	**16.06**	**17.94**
64	0.5	0.5	16.88	18.78
32	0.5	0.1	12.83	20.22
128	0.5	0.1	12.93	21.28

during this process. Table 8 shows the corresponding classification errors of different combinations. The minimum classification error of the test samples is achieved when the number of LSTM units is 64, the LSTM dropout is 0.5 and the LSTM recurrent dropout is 0.1.

C. VALIDATION OF ANTI-NOISE PERFORMANCE

To assess the anti-noise performance of the algorithm, we treat the denoising preprocessing in our pipeline as an option and compare the F_1 scores achieved with and without this operation. The denoising preprocessing used in our experiments contains two main steps: baseline wander removal and wavelet denoising. For baseline wander removal, we apply a median filtering operation with window size of 1 second to the ECG signal to estimate the baseline, then subtract the baseline from the original signal. In wavelet denoising, we use the 'db4' wavelet to make a 5-level discrete wavelet transform (DWT) on the ECG signal. The coefficients obtained from DWT are filtered by soft thresholding method and then processed by inverse wavelet transform to reconstruct the target signal [57]. Figure 5 shows an example of the denoising effect in our experiments. We can see that, after the denoising operations, the components of ECG are all preserved while high-frequency noises are significantly suppressed. Therefore, the denoising method is effective and adequate to be applied in the control group for the evaluation of anti-noise performance of our DNN model. The signal in this example is from lead I of recording 'A0002'.

We assess the performance of the models with and without denoising preprocessing both in a 5-fold cross-validation setting. The F_1 scores achieved by the two groups of models are shown in Figure 6. It seems that the model without denoising preprocessing performed slight better than that with denoising. We further evaluate the significance of this observation by doing a paired-sample t-test on the two groups of F_1 scores. The resulting p-value is 0.087 (> 0.05), suggesting the difference between the two groups of values is not significant. Nonetheless, it indicates that the denoising preprocessing

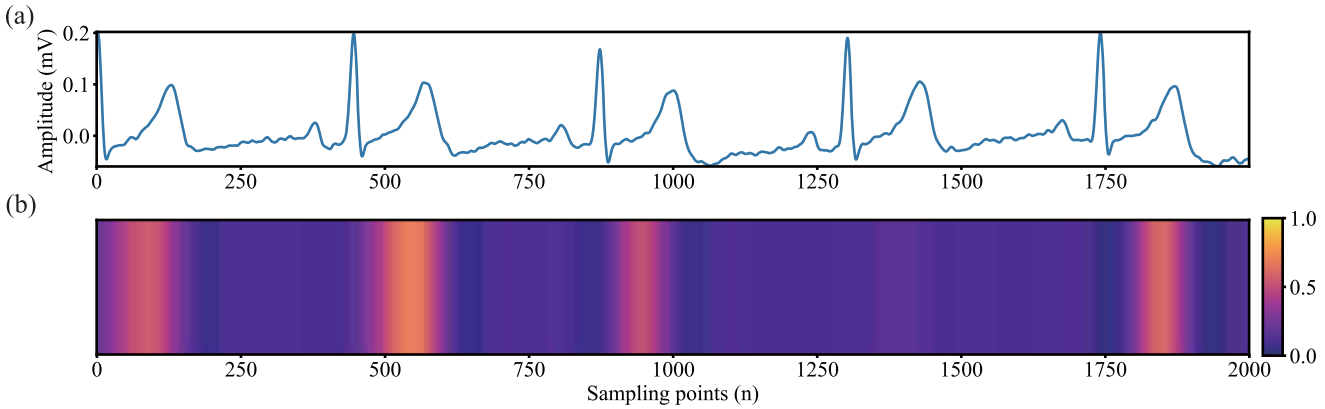


FIGURE 7. (a) Original normal signal. (b) Right attention mapping.

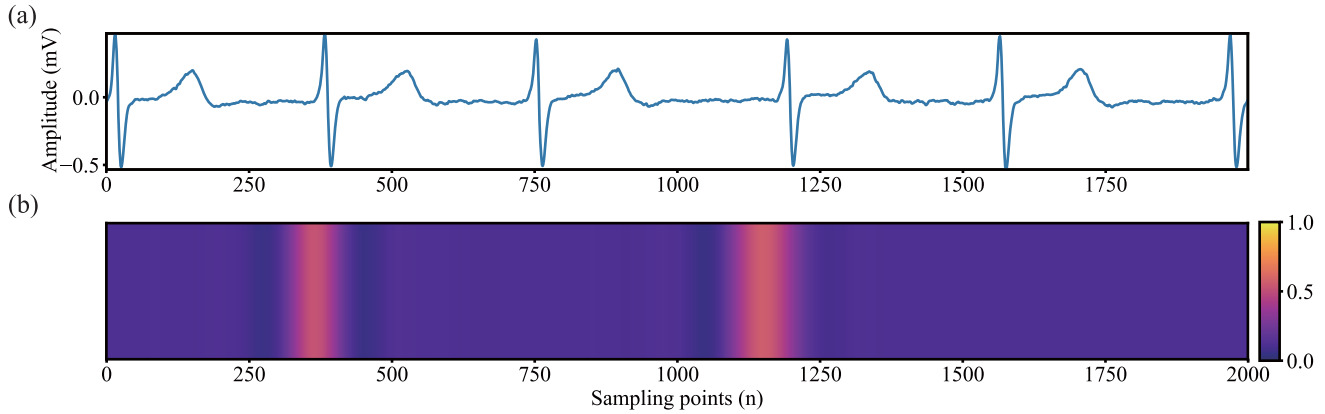


FIGURE 8. (a) Original AF signal. (b) Right attention mapping.

is not necessary in our pipeline. Furthermore, in view of the effectiveness of this denoising method, we can validate that our DNN model is robust to the noise disturbance in the ECG signal. This may result from the denoising ability of the convolutional [58] and dropout operations [59]. The red solid line in the bar chart represents the median of the corresponding group.

D. COMPARISION WITH MODELS OF DIFFERENT STRUCTURES

In order to evaluate the effectiveness of our proposed model structure, we also compared the performance measures of our model with that of several different mainstream structures. For convenience of expression, we name each model structure by its local features learning part's name followed by its global features learning part' name, with a '+' connecting these two parts. For example, we name our proposed model "ResNet+BiLSTM_GMP", which indicates the local features learning part is a 1D residual convolutional network and the global features learning part is a bidirectional LSTM layer stacked with a global maximum pooling (GMP) layer. We compared the results of our model with that of 5 different structures, which are presented as follows:

1) ResNet+Flatten: The local features learning part is in the same structure of ResNet+BiLSTM_GMP, while the global features learning part is just a flatten operation on the local feature maps.

2) ResNet+GAP: With the local features learning part invariant, the global features learning part is a global average pooling (GAP) layer which outputs the average of each feature.

3) ResNet+GMP: With the local features learning part invariant, the global features learning part is a global maximum pooling layer which outputs the maximum of each feature.

4) ResNet+LSTM: With the local features learning part the same, the global features learning part is a LSTM layer whose outputs at the last step will be used as the global features.

5) CNN+BiLSTM_GMP: The local features learning part is composed by 12 convolutional layers each followed by a BN layer and a ReLU activation layer. There are also 6 MaxPooling layers with pool size of 2 interspersed evenly in the CNN part, which means that its output feature map is in the same size of the ResNets in other models. As its name implies, the structure of the global features learning part is same as that of ResNet+BiLSTM_GMP.

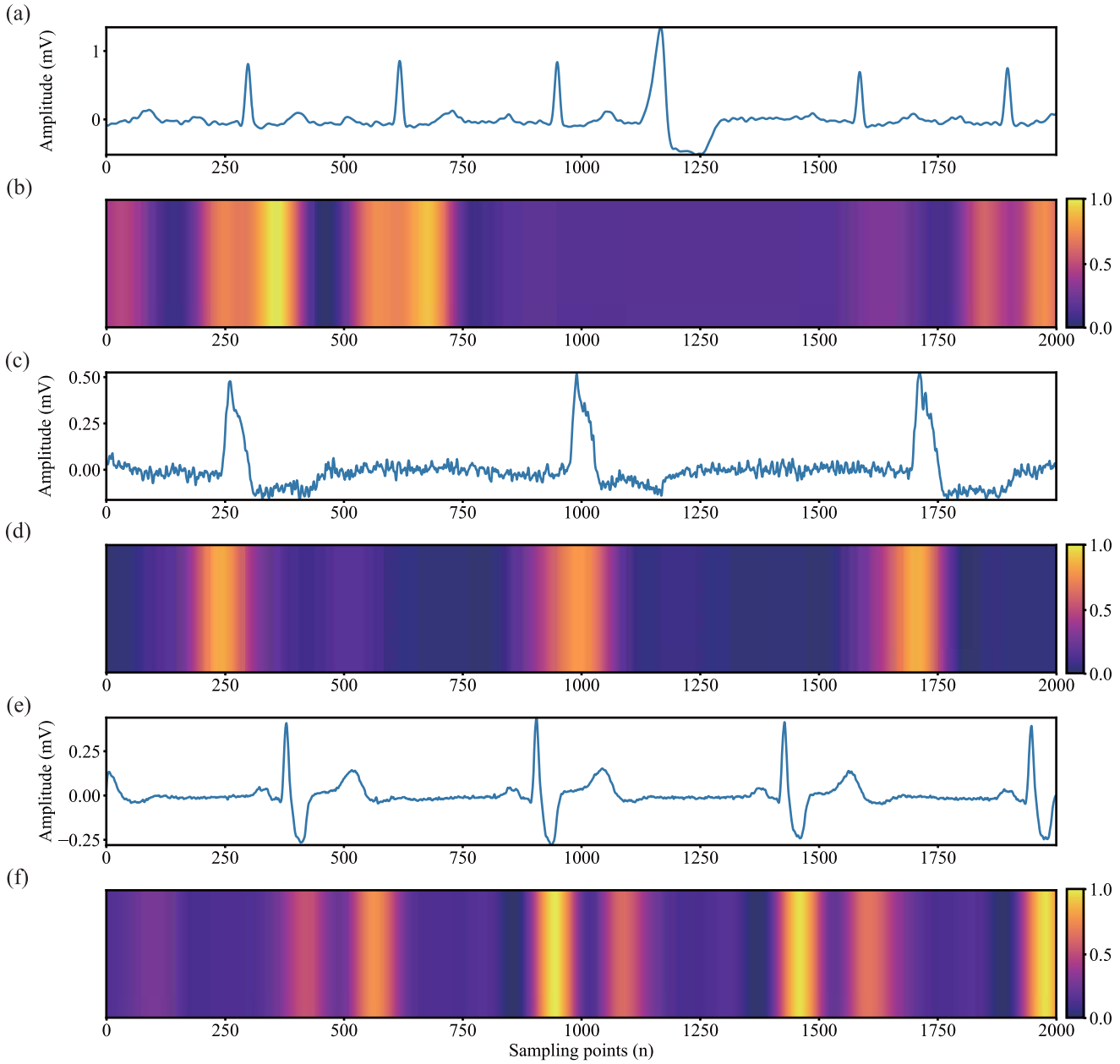


FIGURE 9. (a) Original I-AVB signal. (b) Right attention mapping. (c) Original LBBB signal. (d) Right attention mapping. (e) Original RBBB signal. (f) Right attention mapping.

TABLE 9. F_1 scores achieved by models with different structures in the 5-fold cross-validation experiments. The score in each cell is in the form of "Mean ± STD" of the corresponding F_1 scores got from the cross-validations. The score with maximal mean value in each column is written in bold font.

Model structure	Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE	F_1
ResNet+Flatten	0.604±0.031	0.749±0.029	0.788±0.036	0.811±0.026	0.725±0.035	0.347±0.068	0.619±0.077	0.658±0.045	0.546±0.099	0.65±0.029
ResNet+GAP	0.73±0.026	0.832±0.017	0.854±0.021	0.854±0.02	0.772±0.012	0.627±0.054	0.79±0.028	0.753±0.012	0.682±0.049	0.766±0.006
ResNet+GMP	0.735±0.034	0.834±0.027	0.849±0.027	0.859±0.024	0.765±0.017	0.697±0.043	0.793±0.044	0.783±0.024	0.682±0.009	0.777±0.011
ResNet+LSTM	0.743±0.017	0.836±0.016	0.859±0.014	0.87±0.025	0.772±0.028	0.739±0.02	0.821±0.022	0.782±0.038	0.688±0.048	0.79±0.007
CNN+BiLSTM_GMP	0.72±0.033	0.834±0.009	0.859±0.015	**0.878±0.031**	0.766±0.028	0.747±0.024	0.823±0.033	0.77±0.029	0.665±0.045	0.785±0.011
ResNet+BiLSTM_GMP	**0.755±0.049**	**0.846±0.024**	**0.87±0.021**	0.869±0.028	**0.78±0.028**	**0.751±0.03**	**0.829±0.019**	**0.791±0.01**	**0.704±0.049**	**0.799±0.014**

The results of each model can be found in Table 9. It shows that our proposed model has obvious advantages over others not only in the total score but also in all sub-scores except LBBB whose best result is achieved by the CNN+BiLSTM_GMP model. From the comparison between ResNet+BiLSTM-GMP and

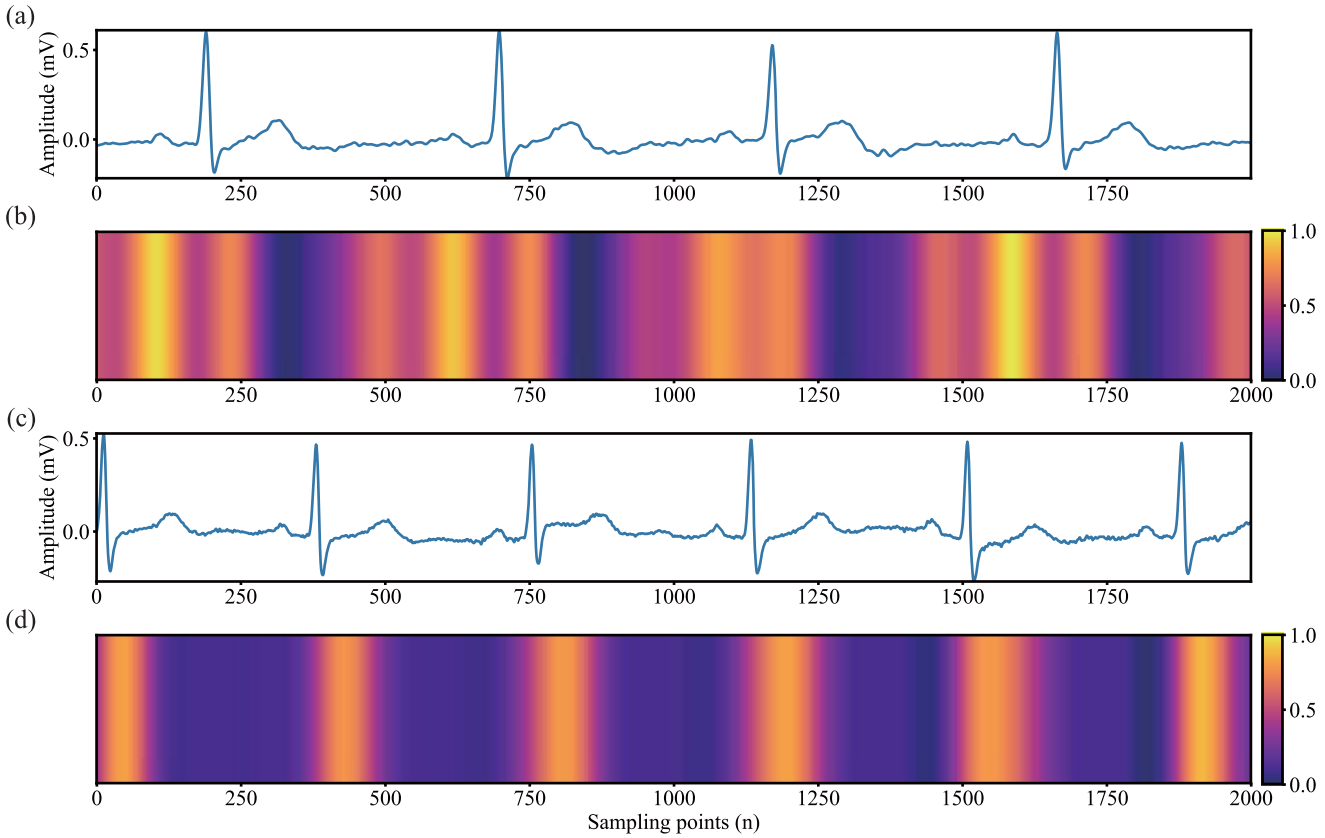


FIGURE 10. (a) Original PAC signal. (b) Right attention mapping. (c) Original PVC signal. (d) Right attention mapping.

CNN+BiLSTM_GMP, we can see the ResNet is better than CNN in the learning of local features. Besides, the comparison between ResNet+BiLSTM_GMP and other models with ResNet as the local features learning part indicates that bidirectional LSTM and global max pooling has a better performance than other structures to summarize the local temporal features into the global features for ECG classification. Therefore, the advantages of each part contribute to the success of the overall model. Furthermore, for some popular public datasets such as the MIT-BIH Arrhythmia Database (MITDB) [60], it is commonly used for the classification of heartbeats, however, the proposed model is suitable for ECG signals with a long-time course. Therefore, we have not implemented the MITDB to test our model.

E. ATTENTION MAPPING ANALYSIS FOR EACH CLASS

Grad-CAM has been used to visually explain the performance of the proposed DNNs [60]. For each class, according to the output of features, the right attention mapping for the ECG signal in nine different conditions are shown in Figures 7-11, with the brighter color representing a higher concentration of the corresponding ECG signal component. As shown in Figure 7 and 8, attention is mainly focused on the T-wave in the normal condition, whilst on abnormal P waves in AF conditions. For three block conditions (I-AVB, LBBB and RBBB), attention concentrates on abnormal QRS complex waves as shown in Figure 9. For two premature contraction conditions (PAC and PVC), the feature components are focused on the distorted P and QRS complex waves for PAC, meanwhile on QRS complex and T waves for PVC as shown in Figure 10. Apparently, for the two ST-segment changed conditions (STD and STE), attention is paid to the ST-segment as shown in Figure 11. As seen from the right attention mapping, the extracted features by the model are in line with the clinical judgment, demonstrating the proposed DNNs is potentially effective for identifying most of arrhythmias.

In some cases, wrong attention mapping for the ECG signal can be obtained as shown in Figure 12 for three typically selected cases. As shown in Figure 12 (a) and 12 (b), a normal case is identified as STD due to the focused ST-segment; while a STE case is classified as a normal condition as the major focused components are not in ST-segment (Figure 12 (c) and 12 (d)). This explains the lower scores achieved for normal and changed ST-segment conditions because of the wrongly extracted features. Finally, the PAC condition is classified as RBBB condition as shown in Figure 12 (e) and 12 (f), due to the focus is on QRS complex and T waves, rather than P waves.

F. OVERALL ANALYSIS OF FEATURE LEARNING

Extracting sufficient and relevant features is crucial to the success of a machine learning task. In many cases, features learned by a DNN can generate a better result than that extracted by elaborated and tedious handwork. As traditional

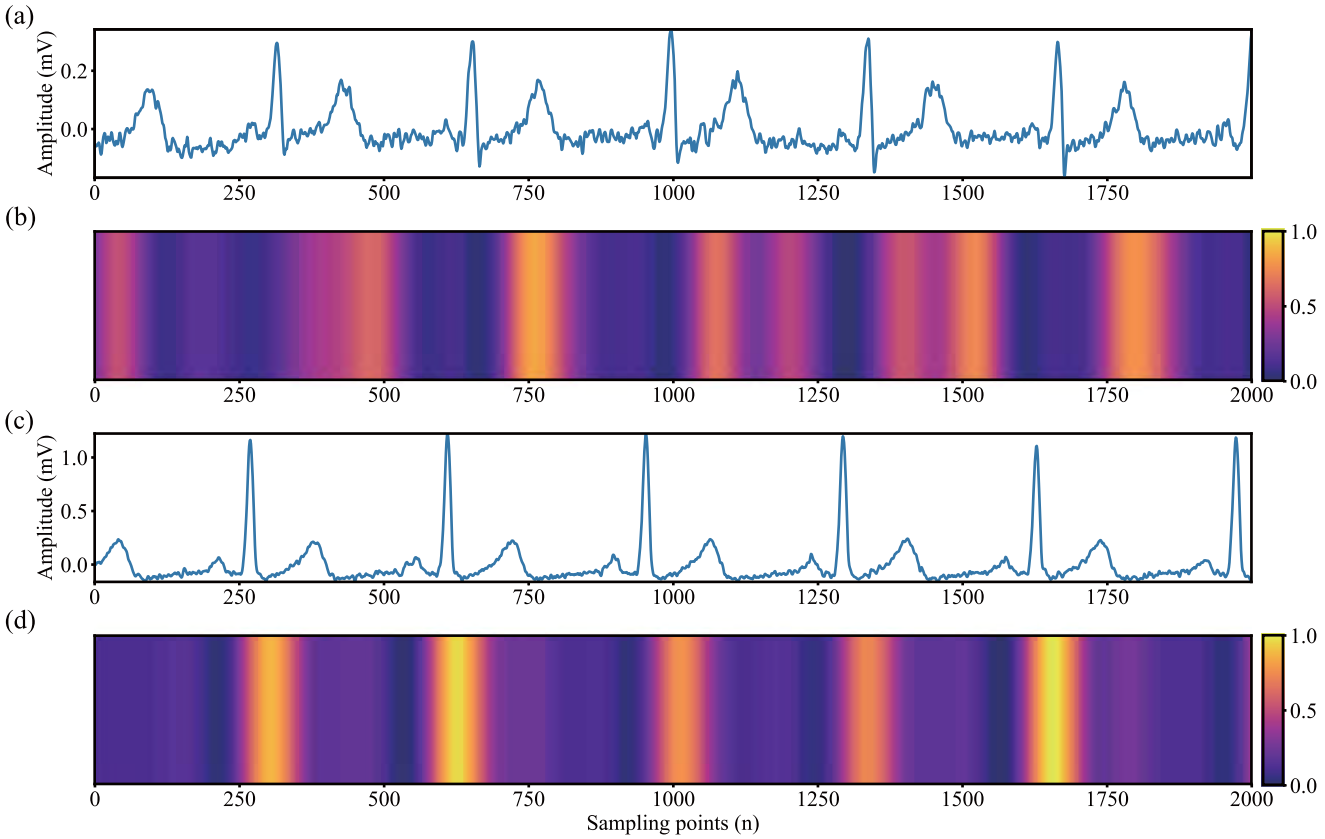


FIGURE 11. (a) Original STD signal. (b) Right attention mapping. (c) Original ECG STE signal. (d) Right attention mapping.

methods struggle to reach a satisfactory result for cardiac arrhythmias classification, deep learning has been considered as a promising alternative approach to solve the problem. The pathology-related features in ECG signals are mostly reflected in the waveform morphology and rhythm changes. Generally, the waveform morphology changes occur in a local time period, such as the ST-segment, while the rhythm changes exist in the global period of the recording. Therefore, the feature learning of DNNs must have the ability to effectively extract both local and global features from ECG recordings.

Previous studies that utilized DNNs to classify ECG are mostly in beat level [33], [34], [36]–[39]. The input of these networks typically consists of hundreds of samples in each channel (i.e., ECG lead). Therefore, these networks only focus on a local time period in an ECG recording, which are unable to depict the rhythm changes in a longer time period. The previous study [32] that applied deep residual networks for multiple rhythms classification was based on a sequence-to-sequence learning manner, where the ECG recordings were labeled at one-second intervals. In this study, however, we focus on the long-term ECG classification and only global-scope labels are available as sources of supervision.

Our DNN models for feature learning are composed of two parts, namely the local features learning part and global features learning part. The structure of the local features learning part is very similar to that in [32], where a stack of residual convolutional modules are utilized to learn the local waveform features and compress the length of feature maps. The shortcut connections in the residual convolutional modules help the gradients propagate in such a deep network. The residual convolutional network structure is effective to learn the local morphological features from raw ECG signal, which is supported by some experimental results mentioned above. For example, the ST-segment changes, including STD and STE, are very subtle changes of the ECG morphology. In the experiments for comparison of different structures, models using residual convolutional network to learn local features (e.g., ResNet+GMP, ResNet+LSTM, ResNet+BiLSTM_GMP) achieved better results on detection of STD and STE than that using CNN (i.e., CNN+BiLSTM_GMP), see Table 9. However, the local features are too dispersed to make a final classification. Therefore, we introduce the global features learning part, where the sequence of local feature vectors is input into a bidirectional LSTM layer.

The bidirectional LSTM is good at characterizing temporal behavior but is hard to deal with very long sequences, such as a sequence with hundreds of steps. After the process of local features learning part, the local morphology features of the input ECG recordings can be extracted,

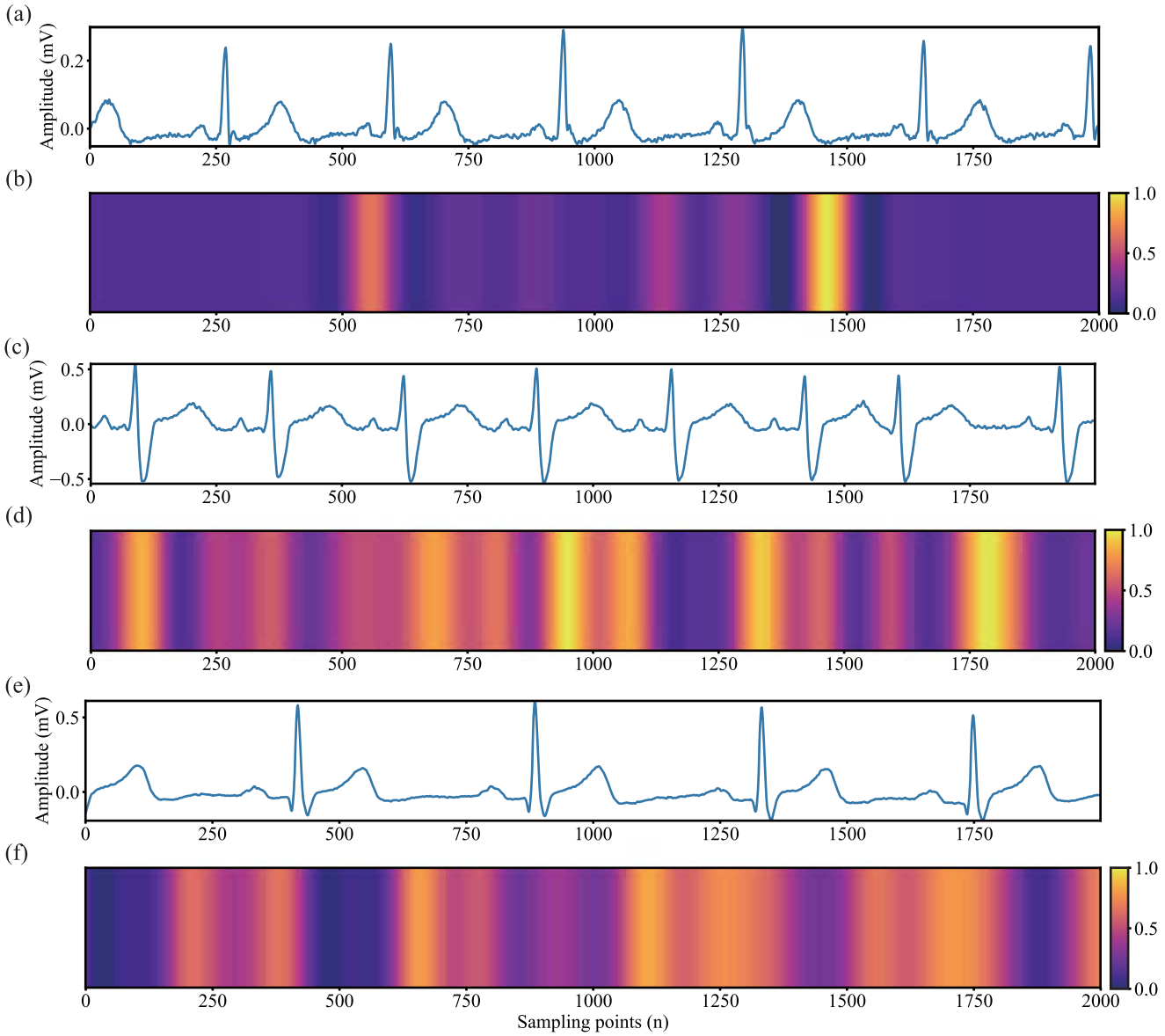


FIGURE 12. (a) Original normal signal. (b) Regared as STD. (c) Original STE signal. (d) Regared as normal. (e) Original PAC signal. (f) Regared as RBBB.

and the sequence length can be compressed, which will enhance the effectiveness and efficiency of RNN-based structures to learn global features. Our experimental results also indicate that LSTM-based structures (both LSTM and bidirectional LSTM) perform better in extraction of long-term features than other structures. For example, the models with LSTM (i.e., ResNet+LSTM, ResNet+BiLSTM_GMP and CNN+BiLSTM_GMP) achieved higher scores in detection of AF, PAC and PVC than others. All these three diseases exhibit obvious rhythm changes over a longer period, which are not easily characterized by the local features learning part.

The output of the bidirectional LSTM layer at a certain step manifests a set of local-focused global features which are affected greatly by the nearby inputs around the step. To get a more representative global feature vector, a global max-pooling layer is applied to summarize these local-focused global feature vector into a single global vector. As indicated by the experimental results, the network can automatically learn more significant features for the classification task.

The ECG signals for features learning in this study contain data recorded from all the 12 leads. However, the relevant features of some arrhythmias merely present in a few leads. Therefore, in the future works, new methods based on feature learning from the related leads of certain pathological condition are expected, which can reduce the disturbances of irrelevant data.

IV. CONCLUSION

In this paper, a novel framework based on DNNs has been proposed for automated classification of arrhythmias. The proposed DNNs has an end-to-end classification structure

composed of three parts, namely local features learning part, global features leaning part and classification part. In summary, the contributions of this research are the following. (1) Residual convolutional modules are the main components of the local features learning part, while the global features learning part consists of a bidirectional LSTM layer and a global max-pooling layer, which can automatically focus on the most of major features. (2) The model uses raw ECG signals without filtering, showing its capability to handle with signals corrupted by noise. In this work, the CPSC database has been used for validating the performance of the proposed method which obtains a high overall F_1 score of 0.806 in the presence of the different noises in the 12 leads ECG recordings. Classification results showed that DNNs significantly increases the classification accuracy for various arrhythmia types. Furthermore, the proposed method demonstrated good performance in anti-noise interference, automatic feature extraction and low classification error, making it a potential choice for clinical automatic arrhythmia classification in the future.

ACKNOWLEDGMENT

RH and YL wrote the first draft of the text and performed the design and implementation of the algorithm. HZ and QL contributed to write the manuscript and helped with the algorithm. KW and YY commented on and approved the manuscript. NZ participated the research in the classification of arrhythmias. All authors read and approved the final manuscript. *(Runnan He and Yang Liu are co-first authors.)*

REFERENCES

[1] R. E. Kass, and C. E. Clancy, Eds., *Basis and Treatment of Cardiac Arrhythmias*, vol. 171. Berlin, Germany: Springer, 2005.

[2] V. X. Afonso and W. J. Tompkins, "Detecting ventricular fibrillation," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 2, pp. 152–159, Mar. 1995.

[3] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: A diagnostic system," *J. Biomed. Eng.*, vol. 11, no. 4, pp. 320–328, Jul. 1989.

[4] K. I. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 179–185, Feb. 1999.

[5] J. A. Kastor, Eds., *Arrhythmias*. Philadelphia, PA, USA: WB Saunders Company, 2000.

[6] M. Milanesi, N. Martini, N. Vanello, V. Positano, M. F. Santarelli, and L. Landini, "Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals," *Med. Biol. Eng. Comput.*, vol. 46, no. 3, pp. 251–261, Mar. 2008.

[7] O. Sayadi and M. B. Shamsollahi, "Model-based fiducial points extraction for baseline wandered electrocardiograms," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 347–351, Jan. 2008.

[8] S. Ari, M. K. Das, and A. Chacko, "ECG signal enhancement using S-Transform," *Comput. Biol. Med.*, vol. 43, no. 6, pp. 649–660, Jul. 2013.

[9] M. Zivanovic and M. González-Izal, "Simultaneous powerline interference and baseline wander removal from ECG and EMG signals by sinusoidal modeling," *Med. Eng. Phys.*, vol. 35, no. 10, pp. 1431–1441, Oct. 2013.

[10] A. Fasano and V. Villani, "Baseline wander removal for bioelectrical signals by quadratic variation reduction," *Signal Process.*, vol. 99, pp. 48–57, Jun. 2014.

[11] M. Merino, I. M. Gómez, and A. J. Molina, "Envelopment filter and *K*-means for the detection of QRS waveforms in electrocardiogram," *Med. Eng. Phys.*, vol. 37, no. 6, pp. 605–609, Jun. 2015.

[12] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[13] C. A. Bustamante, S. I. Duque, A. Orozco-Duque, and J. Bustamante, "ECG delineation and ischemic ST-segment detection based in wavelet transform and support vector machines," in *Proc. Pan Amer. Health Care Exchanges (PAHCE)*, Medellin, Colombia, Apr. 2013, pp. 1–7.

[14] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "Feature selection algorithm for ECG signals using range-overlaps method," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3499–3512, Apr. 2010.

[15] M. Llamedo and J. P. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 616–625, Mar. 2011.

[16] J. Oster, J. Behar, O. Sayadi, S. Nemati, A. E. W. Johnson, G. D. Clifford, "Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2125–2134, Sep. 2015.

[17] L. S. C. de Oliveira, R. V. Andreão, M. Sarcinelli-Filho, "Premature ventricular beat classification using a dynamic Bayesian network," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Boston, MA, USA, Aug./Sep. 2011, pp. 4984–4987.

[18] G. De Lannoy, D. Francois, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 241–247, Jan. 2012.

[19] X. Jiang, L. Zhang, Q. Zhao, and S. Albayrak, "ECG arrhythmias recognition system based on independent component analysis feature extraction," in *Proc. TENCON IEEE Region Conf.*, Hong Kong, Nov. 2006, pp. 1–4.

[20] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[21] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.

[22] V. S. R. Kumari and P. R. Kumar, "Cardiac arrhythmia prediction using improved multilayer perceptron neural network," *Int. J. Electron., Commun. Instrum. Eng. Res. Develop.*, vol. 3, no. 4, pp. 73–80, Oct. 2013.

[23] C. V. Banupriya and S. Karpagavalli, "Electrocardiogram beat classification using probabilistic neural network," *Int. J. Comput. Appl.*, vol. 1, no. 7, pp. 31–37, Apr. 2014.

[24] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1750–1761, Nov. 2007.

[25] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004.

[26] X. D. Zeng, S. Chao, and F. Wong, "Ensemble learning on heartbeat type classification," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, Macau, China, Jun. 2011, pp. 320–325.

[27] P. D. Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[28] S. Yang and H. Shen, "Heartbeat classification using discrete wavelet transform and kernel principal component analysis," in *Proc. IEEE Tencon-Spring*, Sydney, NSW, Australia, Apr. 2013, pp. 34–38.

[29] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Jun. 2016.

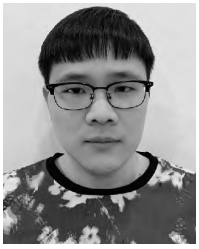
[30] R. He, K. Wang, N. Zhao, Y. Liu, Y. Yuan, Q. Li, and H. Zhang, "Automatic detection of atrial fibrillation based on continuous wavelet transform and 2D convolutional neural networks," *Frontiers Physiol.*, vol. 9, p. 1206, Aug. 2018.

[31] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[32] Y. H. Awni, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, Jan. 2019.

[33] M. Zubair, J. Kim, and C. Yoon, "An automated ECG beat classification system using convolutional neural networks," in *Proc. 6th Int. Conf. Converg. Secur. (ICITCS)*, Prague, Czech Republic, Sep. 2016, pp. 1–5.

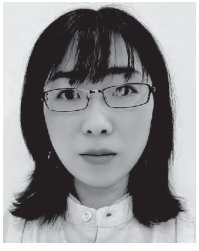
[34] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018.
[35] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
[36] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan, and L. Wang, "A Restricted Boltzmann Machine Based Two-Lead Electrocardiography Classification," in *Proc. IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Cambridge, MA, USA, Jun. 2015, pp. 1–9.
[37] S. S. Xu, M.-W. Mak, C.-C. Cheung, "Towards end-to-end ECG classification with raw signal extraction and deep neural networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1574–1584, Jul. 2019.
[38] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Comput. Biol. Med.*, vol. 99, pp. 53–62, Aug. 2018.
[39] K. Luo, J. Li, Z. Wang, and A. Cuschieri, "Patient-specific deep architectural model for ECG classification," *J. Healthcare Eng.*, vol. 2017, May 2017, Art. no. 4108720.
[40] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM based auto-encoder model for ECG arrhythmias classification," *IEEE Trans. Instrum. Meas.*, to be published.
[41] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput. Biol. Med.*, vol. 102, no. 1, pp. 278–287, Nov. 2018.
[42] M.-G. Kim and S. B. Pan, "Deep learning based on 1-D ensemble networks using ECG for real-time user recognition," *IEEE Trans. Ind. Informat.*, to be published.
[43] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ECG pattern analysis and classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 326–333, Feb. 2014.
[44] S. Dilmac and M. Korurek, "ECG heart beat classification method based on modified ABC algorithm," *Apply Soft Comput.*, vol. 36, pp. 641–655, Nov. 2015.
[45] S. Shadmand and B. Mashoufi, "A new personalized ecg signal classification algorithm using block-based neural network and particle swarm optimization," *Biomed. Signal Process. Control*, vol. 25, pp. 12–23, Mar. 2016.
[46] J. Mateo, A. M. Torres, A. Aparicio, and J. L. Santos, "An efficient method for ECG beat classification and correction of ectopic beats," *Comput. Electr. Eng.*, vol. 53, pp. 219–229, Jul. 2016.
[47] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, and J. Li, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Jul. 2018.
[48] R. He, K. Wang, Q. Li, Y. Yuan, N. Zhao, Y. Liu, and H. Zhang, "A novel method for the detection of R-peaks in ECG based on K-nearest neighbors and particle swarm optimization," *EURASIP J. Adv. Signal Process.*, vol. 82, no. 1, pp. 1–14, Dec. 2017.
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[50] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 630–645.
[51] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[52] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, Haifa, Israel, Jun. 2010, pp. 807–814.
[53] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," Jul. 2012, *arXiv:1207.0580*. [Online]. Available: https://arxiv.org/abs/1207.0580
[54] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Proc. IJCAI*, Barcelona, Spain, Jul. 2011, pp. 1237–1242.
[55] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[56] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[57] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 613–627, Aug. 1995.
[58] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[59] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, Helsinki, Finland, Jun. 2008, pp. 1096–1103.
[60] G. B. Moody and R. G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," in *Proc. Comput. Cardiol.*, Sep. 1990, pp. 185–188.
[61] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 618–626.



RUNNAN HE received the B.S. degree in biomedical engineering from the Shenyang University of Technology, China, in 2012, and the M.S. degree in biomedical engineering from Northeastern University, China, in 2014. He is currently pursuing the Ph.D. degree with the Perception Computing Center, School of Computer Science and Technology, Harbin Institute of Technology, China. His research fields include medical signal processing, machine learning, and deep learning.

YANG LIU received the B.S. and M.S. degrees from the Harbin Institute of Technology, China, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree in computer science and technology. His research interests include biomedical signal processing, pattern recognition in time-series signals, artificial neural networks, and computer-aided diagnosis.

KUANQUAN WANG (M'01–SM'07) was the Associate Dean of the School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, and also the Dean of the School of Computer Science and Technology, HIT, Weihai, from 2011 to 2014. He is currently a Full Professor and a Ph.D. Supervisor with the School of Computer Science and Technology, and the Deputy Director of the Research Center of Perception and Computing, HIT. His main research areas include image processing and pattern recognition, biometrics, biocomputing, modeling and simulation, virtual reality, and visualization. He has published over 300 papers and six books, and holds more than 10 patents. He is a Senior Member of the China Computer Federation (CCF) and ACM, and a Senior Member of the Chinese Society of Biomedical Engineering. He was a recipient of the second prize of the National Teaching Achievement.



NA ZHAO received the B.S. degree in biomedical engineering from the Shenyang University of Technology, China, in 2012, and the M.S. degree in biomedical engineering from Northeastern University, China, in 2014. She is currently pursuing the Ph.D. degree with the Perception Computing Center, School of Computer Science and Technology, Harbin Institute of Technology, China. Her research fields include modeling, simulation, and machine learning.

YONGFENG YUAN received the B.S. and Ph.D. degrees in computer science and technology from the Harbin Institute of Technology, China, in 2002 and 2010, respectively, where he is currently an Associate Professor with the School of Computer Science and Technology. His research fields include computational cardiology, image processing, and virtual reality.

QINCE LI received the bachelor's degree in physics with astrophysics, and the Ph.D. degree in biophysics from The University of Manchester, in 2007 and 2012, respectively. From 2013 to 2015, he was a Postdoctoral Fellow with The University of Alabama at Birmingham, USA. He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology, China. His research fields include computational modeling in the cardiovascular systems, intracellular Ca2+ and ROS dynamics, mitochondria energetics, and cardiac optogenetics. He has published over 30 papers in peer-review journals and international conferences. He has been a member of the China Computer Federation, since 2016.

HENGGUI ZHANG received the Ph.D. degree in mathematical cardiology from the University of Leeds, in 1994. Then, he worked as a Postdoctoral Research Fellow with The Johns Hopkins University School of Medicine (1994–1995) and the University of Leeds (1996–2000), and then as a Senior Research Fellow with the University of Leeds (2000–2001). In October 2001, he moved to UMIST to take up a lectureship. Since then, he worked as a Lecturer with UMIST (2001–2004), and as a Senior Lecturer (2004–2006) and Reader (2006–2009) with The University of Manchester. He currently holds the Chair of the Biological Physics Group, School of Physics and Astronomy, The University of Manchester. He is also a Professor of biological physics. He has published more than 400 scientific papers, among them over 200 papers were published in prestigious peer-reviewed journals in his field. Related works have attracted wide public interests, and have been covered by many prestigious media such as BBC. He has been elected as a Fellow of the world-renowned societies as the recognition of distinctions.