IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH

Received November 2, 2019, accepted November 20, 2019, date of publication November 26, 2019,
date of current version December 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956050

A Novel Method to Detect Multiple Arrhythmias Based on Time-Frequency Analysis and Convolutional Neural Networks

ZIQIAN WU[1], (Student Member, IEEE), TIANJIE LAN[1], CUIWEI YANG[1,2], (Member, IEEE), AND ZHENNING NIE[3]

[1]Department of Electronic Engineering, Fudan University, Shanghai 200433, China
[2]Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai Engineering Research Center of Assistive Devices, Fudan University, Shanghai 200093, China
[3]Department of Cardiology, Shanghai Institute of Cardiovascular Diseases, Zhongshan Hospital, Fudan University, Shanghai 200032, China

Corresponding authors: Cuiwei Yang (yangcw@fudan.edu.cn) and Zhenning Nie (nie.zhenning@zs-hospital.sh.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61071004 and Grant 81671777, in part by the Shanghai Science and Technology Support Project under Grant 18441900900, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2017SHZDZX01, and in part by the Project of Shanghai Engineering Research Center under Grant 15DZ2251700.

ABSTRACT Electrocardiogram (ECG) is an efficient and commonly used tool for detecting arrhythmias. With the development of dynamic ECG monitoring, an effective and simple algorithm is needed to deal with large quantities of ECG data. In this study, we proposed a method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks. For a short-time (10 s) single-lead ECG signal, the time-frequency distribution matrix of the signal was first obtained using a time-frequency transform method, and then a convolutional neural network was used to discriminate the rhythm of the signal. ECG data in multiple databases were used and were divided into 12 classes. Finally, the performance of three kinds of time-frequency transform methods are evaluated, including short-time Fourier transform (STFT), continuous wavelet transform (CWT), and pseudo Wigner-Ville distribution (PWVD). The best result was obtained by STFT, with an accuracy of 96.65%, an average sensitivity of 96.47%, an average specificity of 99.68%, and an average F_1 score of 96.27%, respectively. Especially, the area under curve (AUC) value is 0.9987. The proposed method in this work may be efficient and valuable to detect multiple arrhythmias for dynamic ECG monitoring.

INDEX TERMS Arrhythmia detection, convolutional neural networks, ECG, time-frequency analysis.

I. INTRODUCTION

Cardiovascular disease is now the leading cause of mortality. According to statistics, about 17 million people die of cardiovascular disease each year, accounting for 37% of the global death toll. In developing countries, the mortality of cardiovascular disease is accounting for 3/4 of the total number of deaths [1]. The incidence of cardiovascular disease will further increase and this will become a serious public health problem as the population ages and the causes of cardiovascular disease increase, such as obesity, stress [2]. Arrhythmia, which is mainly caused by the abnormal electrical activity of the heart, is one of the main manifestations of cardiovascular disease [3]. As a direct reflect of cardiac electrical activity, body surface electrocardiogram (ECG) is an efficient and low-cost way to detect arrhythmias [4]. With the development of sensor technology and wearable devices, long-term dynamic ECG is widely used in the detection of arrhythmia. To deal with large amount of ECG data, it is necessary to develop an automatic detection algorithm to classify different arrhythmias.

In the past few decades, researchers have proposed lots of methods in the field of automatic arrhythmia detection. Early studies have mainly focused on the extraction of ECG morphology features [5], [6] and heart rate variability [7]–[9]. In these studies, it is necessary to extract ECG features before arrhythmia detection, including QRS complex [10], [11], P wave and T wave [12]. In case of false or miss detection, the performance of the algorithm will decrease significantly. Other researchers used non-morphological features to detect

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

arrhythmias, such as features in frequency domain [13] or time-frequency domain [14], [15]. However, these features cannot express all the characteristics of the signal and the selection of features also depends heavily on the experience of the researcher. Besides, a large number of feature calculations will slow down the processing speed and reduce the efficiency of the algorithm.

In recent years, deep learning method has made great progress in different fields, such as image classification [16], speech processing [17] and many other research fields. This method extracts more abstract and advanced features simply through deep networks and performs better than traditional feature extraction process [18]. With the growth of medical information, the deep learning method began to be applied in biological and medical data analysis [19], [20], and some researchers began to explore the application of deep learning method in the field of ECG analysis. For example, A. Y. Hannun *et al*. [21], X. Fan *et al*. [22] and S. L. Oh *et al*. [23] used end-to-end one-dimensional deep neural networks for automatic diagnosis of arrhythmia. M. M. Al Rahhal *et al*. [24], Y. Xia *et al*. [25] and R. He *et al*. [26] proposed some methods similar to image recognition. They converted the signal into a time-frequency distribution and then classified it by a two-dimensional convolutional neural network. These methods have achieved better performance than traditional arrhythmia detection algorithms. Nevertheless, the research of deep learning methods in the field of ECG is still in its infancy due to the difficult annotation of ECG data and so far there are few studies on classifications of multiple arrhythmias.

Inspired by some studies using time-frequency analysis and CNNs to detect atrial fibrillation (AF) [25], [26] or to classify different ECG beats [24], this study proposed a method to detect multiple arrhythmias based on time-frequency analysis and CNNs, and has achieved excellent performance on public databases. We used three kinds of time-frequency transform methods to compare the performance of different time-frequency transforms, including: short-time Fourier transform (STFT), continuous wavelet transform (CWT), and pseudo Wigner-Ville distribution (PWVD). Then a convolutional neural network (CNN) with a fixed architecture was designed to classify twelve heart rhythms including normal sinus rhythm, pacing rhythm, noise and nine kinds of arrhythmia. To summarize, this study classified more kind of ECG rhythm classes than many previous studies and was the first work to compare performances between different time-frequency method in the research field where time-frequency analysis and deep learning were used to classify ECG signals. Compared to those methods using many manually proposed features, the proposed method with simple CNN structure may be an effective method to detect multiple arrhythmias in clinical practice.

The rest of this paper is organized as follows: Section II describes the detailed methods, including the overview of the proposed method, the selection and preprocessing of data used in this study, the time-frequency transform of ECG signal, the CNN architecture and the training method. Section III shows the results of the experiment and evaluate the performance of the proposed method. Section IV presents the discussion on this study, and Section V concludes this study.

II. METHODS AND MATERIALS

A. OVERVIEW

Fig. 1 illustrates the schematic of the proposed method. Length of each signal segment used in this study was fixed to 10 s. Each signal segment was resampled to 125 Hz due to different sampling frequency in different databases. A second order high-pass with a cutoff frequency of 0.5 Hz was used to remove the baseline of each signal segment. Then a time-frequency transform was used to obtain the time-frequency distribution matrix. The matrix was normalized and used as input of the designed CNN. Finally, the network extracted the deep features of the input matrix and output the classification result. We conducted three trials, each using a different time-frequency transform method and a same network architecture.

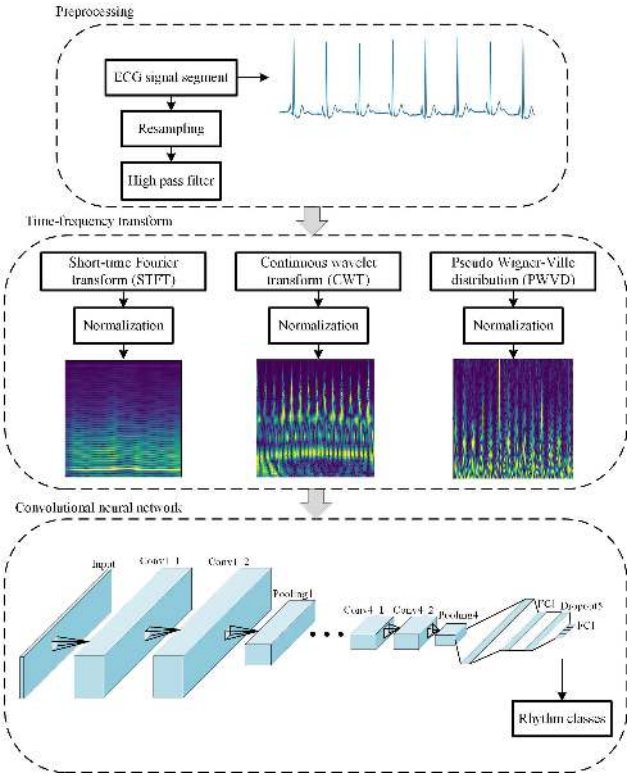

FIGURE 1. Schematic of the proposed method.

In order to overcome the limitations of insufficient data, we used a five-fold cross-validation method. Due to the unbalanced data distribution of different heart rhythms, we had adopted oversampling method for the training set in each training process. A variety of metrics including accuracy, sensitivities, specificities, F_1 scores and the area under curve (AUC) values were used to evaluate the performance

of the proposed method. The final results showed that the method has excellent performance on detecting arrhythmias.

The detailed experimental process of each step is described below.

B. DATA SELECTION AND PREPROCESSING

To obtain ECG data of different heart rhythm, we used ECG data in multiple databases in physionet.org [27]. These databases include the MIT-BIH Arrhythmia Database (DB1) [28], the MIT-BIH Malignant Ventricular Arrhythmia Database (DB2) [29], the MIT-BIH Atrial Fibrillation Database (DB3) [30], the Long-Term AF Database (DB4) [31], the MIT-BIH Normal Sinus Rhythm Database (DB5) [27], the MIT-BIH Noise Stress Test Database (DB6) [32]. Data in DB1 to DB5 contain ECG records of two channels, and we selected data of the first channel. We split the data in DB1 to DB5 into segments with a fixed length of 10 s due to the rhythm annotations which were stored along with the data in the online database. Each segment contains only one kind of rhythm, and there is no overlap between different segments.

In practical applications, the signals severely contaminated by noise have no diagnostic significance, but these signals are quite common during ECG measurement process due to the presence of human respiration, myoelectric signals and noise from environment. Since the labeled noise-contaminated ECG signals were rare in public databases, we used the DB6 to generate the simulated noise-contaminated ECG signals. The DB6 was commonly used in the research field of ECG signal quality [33], and it includes 3 half-hour recordings of noise typical in ambulatory ECG recordings including baseline wander (BW), muscle artifact (MA) and electrode motion artifact (EM). These noise recordings were made using physically active volunteers and standard ECG recorders, leads, and electrodes. These three types of noise data were split into segments with a fixed length of 10 s. All these segments made up the raw noise data. We generated noise-contaminated ECG signals by adding these noise signals to the raw ECG signals in DB1-DB5. The data of BW was not selected to generate noise-contaminated ECG signals, as the baseline affects little on the morphological feature of ECG signals and the high-pass filter in the preprocessing can remove some of its influence.

One or two kind of noise signals including MA and EM were added to the raw ECG signals in a way which is defined as follows:

$$y = x + a * n \tag{1}$$

where y is the signal contaminated by noise, x is the raw ECG signal, a is the gain of noise signal, n is the noise signal. The value of a is calculated by:

$$a = \sqrt{\exp(\frac{-\ln(10) * SNR}{10}) * \frac{S}{N}} \tag{2}$$

where S is the power of raw ECG signal, N is the power of noise signal, SNR is the signal-to-noise ratio and it is set to -6 db or -12 db.

Fig. 2 shows the process of generating an ECG signal contaminated by MA when SNR is set to -6 db. The raw noise signal and the generated noise-contaminated data were both labelled as noise.

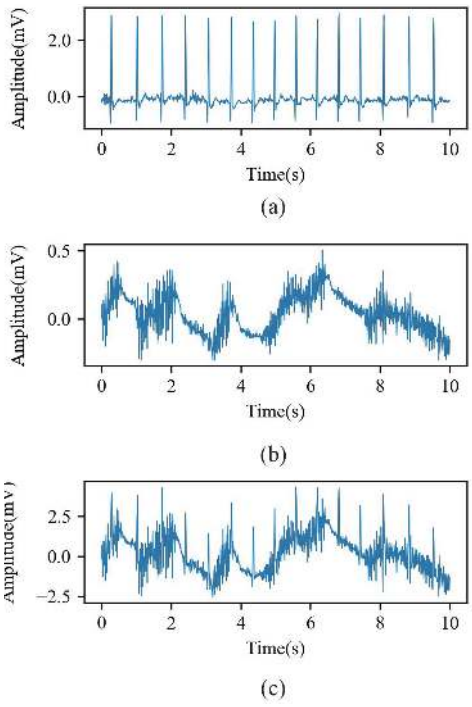

FIGURE 2. Process of adding noise to a raw ECG signal segment. (a) the raw ECG signal, (b) the noise signal, (c) the ECG signal contaminated by noise.

The number of some rhythm segments in the datasets were too small to be of experimental significance such as pre-excitation (PREX), asystole (ASYS), so these rhythms were neglected in this study. We just selected those heart rhythms with a sample size larger than 200, and there are finally 12 kind of heart rhythms in our selected dataset. For heart rhythms which sample size were larger than many other rhythms, we randomly selected 1500 samples for each rhythm. For rhythms which size was smaller than 1500, all the samples are selected. Meanwhile, 1500 segments of ECG data labelled as NOISE were added to our dataset including 1000 noise-contaminated ECG data segments and 500 segments of raw noise data. The 1000 noise-contaminated ECG data segments contains 200 segments of ECG data of normal sinus rhythm (NSR) in DB5, 100 segments of ECG data of NSR and 100 segments of ECG data of arrhythmias each in DB1 to DB4. Table 1 shows the distribution of the heart rhythm and Fig. 3 illustrates the examples of each rhythm.

The sampling rate in each database is different. Since the most of the energy of QRS complex distributes between 3 Hz and 40 Hz [34], we resampled each segment to 125 Hz which was enough for dynamic ECG monitoring. The baseline drift contains most of the energy of signal, and the low-frequency coefficients of time-frequency matrix will be much larger than the high-frequency components without removing the

TABLE 1. Distribution of each heart rhythm.

Symbol	Meaning	Amount	Distribution
AB	Atrial bigeminy	1074	DB1: 7, DB2: 0, DB3: 0, DB4: 1067, DB5: 0, DB6: 0.
AF	Atrial fibrillation	1500	DB1: 500, DB2: 0, DB3: 500, DB4: 500, DB5: 0, DB6: 0.
AFL	Atrial flutter	637	DB1: 56, DB2: 0, DB3: 581, DB4: 0, DB5: 0, DB6: 0.
BI	First degree heart block	411	DB1: 0, DB2: 411, DB3: 0, DB4: 0, DB5: 0, DB6: 0.
NOISE	Noise and signal contaminated by noise	1500	DB1: 0, DB2: 0, DB3: 0, DB4: 0, DB5: 0, DB6: 500, Generated signal: 1000.
NSR	Normal sinus rhythm	1500	DB1: 300, DB2: 300, DB3: 300, DB4: 300, DB5: 300, DB6: 0.
P	Paced rhythm	459	DB1: 315, DB2: 144, DB3: 0, DB4: 0, DB5: 0, DB6: 0.
PVC	Premature ventricular contractions including ventricular bigeminy and ventricular trigeminy	737	DB1: 210, DB2: 44, DB3: 0, DB4: 483, DB5: 0, DB6: 0.
SBR	Sinus bradycardia	1500	DB1: 150, DB2: 0, DB3: 0, DB4: 1350, DB5: 0, DB6: 0.
SVTA	Supraventricular tachyarrhythmia	481	DB1: 12, DB2: 148, DB3: 0, DB4: 321, DB5: 0, DB6: 0.
VF	Ventricular flutter	225	DB1: 0, DB2: 225, DB3: 0, DB4: 0, DB5: 0, DB6: 0.
VT	Ventricular tachycardia	495	DB1: 8, DB2: 487, DB3: 0, DB4: 0, DB5: 0, DB6: 0.

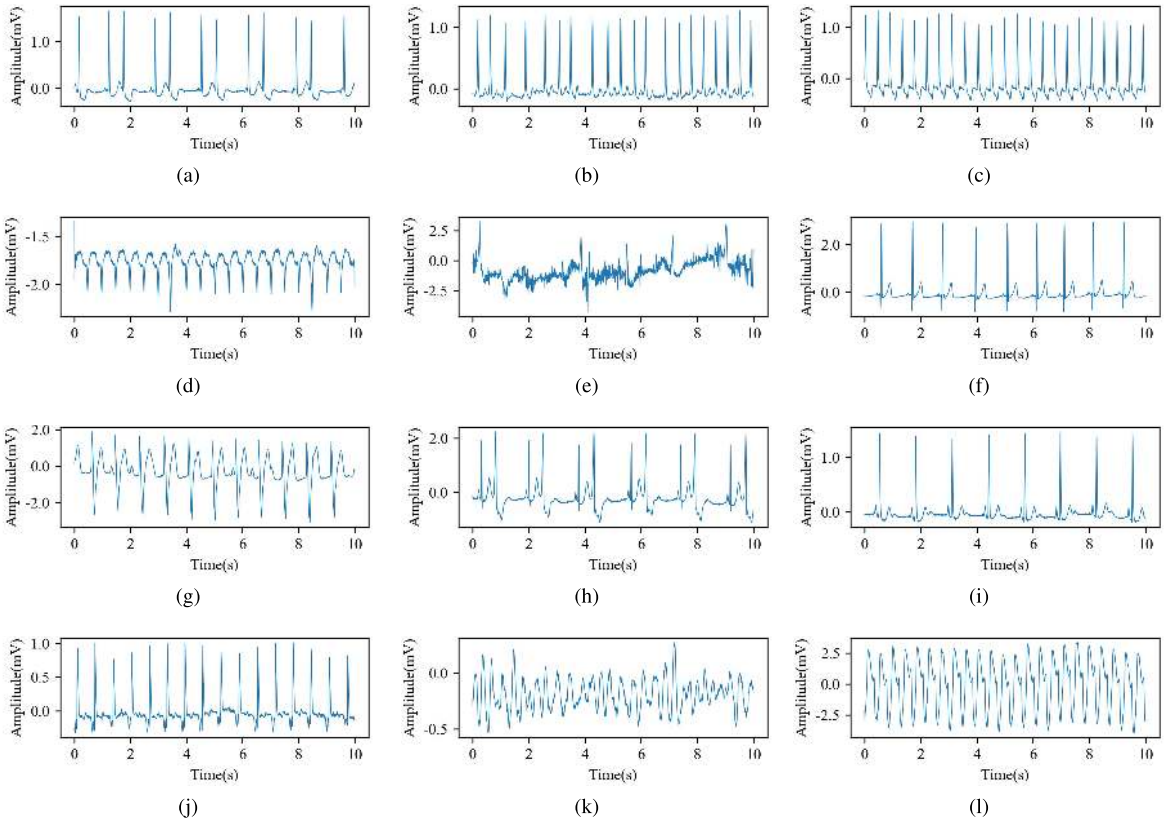

FIGURE 3. Examples of each heart rhythm: (a) atrial bigeminy (AB), (b) atrial fibrillation (AF), (c) atrial flutter (AFL), (d) first degree heart block (BI), (e) Noise and signal contaminated by noise (NOISE), (f) normal sinus rhythm (NSR), (g) paced rhythm (P), (h) premature ventricular contractions (PVC), (i) sinus bradycardia (SBR), (j) supraventricular tachyarrhythmia (SVTA), (k) ventricular flutter (VF), (l) ventricular tachycardia (VT).

baseline drift, resulting unstableness of experiments such as a vanishing gradient, so a second-order high-pass Butterworth filter with a cutoff frequency of 0.5 Hz was used to reduce the baseline drift. After the above preprocessing was completed, the time-frequency transform was applied to each segment.

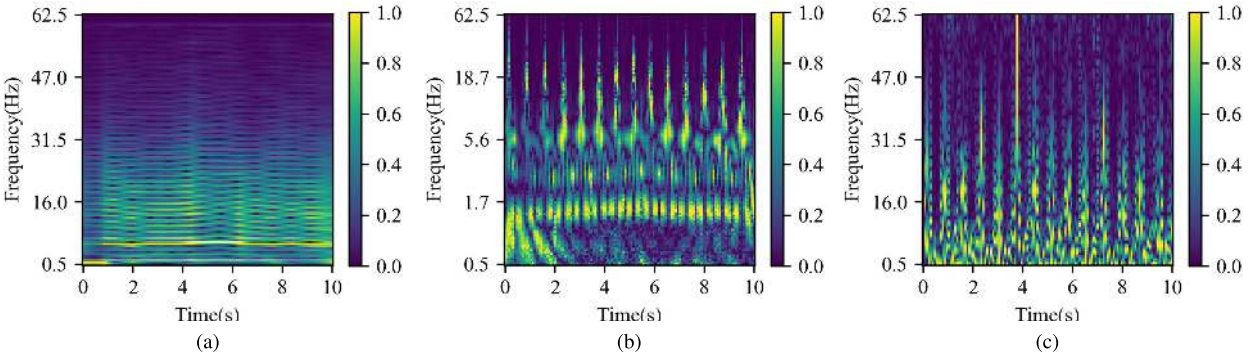

FIGURE 4. Time-frequency distribution matrices of an ECG signal segments of 10 s using different transform method: (a) STFT, (b) CWT, (c) PWVD.

C. TIME-FREQUENCY ANALYSIS

Time-frequency analysis is one of the important methods to process non-stationary signals and provides joint distribution information of time domain and frequency domain [35]. It can clearly describe the change of signal frequency with time. Conventional time-frequency analysis methods can be divided into two major categories: linear methods and quadratic transform methods. The linear method mainly uses the linear correlation between the signal and the kernel function to calculate the spectrum. The commonly used linear method includes STFT [36], CWT [37], s-transform [38]. The main limitation of the linear transform method is that the window function may cause the problem of spectrum leakage and the time resolution and the frequency resolution cannot be optimal at the same time, so the window function should be selected carefully. A typical quadratic time-frequency analysis method is Wigner-Ville distribution (WVD) [39]. The WVD method does not require the use of a window function, so the time-frequency resolution is high. However, when the signal contains multiple time components, cross-terms interference will occur, causing blurring on the time-frequency distribution matrix [40]. In order to reduce the interference of cross-terms, some researchers proposed improved quadratic time-frequency transform method, such as the pseudo WVD (PWVD) method using window function [41] and the Cohen's class distribution using kernel functions [42]. In this study, we used three time-frequency analysis methods: STFT, CWT and PWVD.

The STFT of the sequence $x(t)$ is defined as:

$$STFT(t, w) = \int_{-\infty}^{+\infty} x(\tau) w(\tau - t) e^{-jw\tau} d\tau \tag{3}$$

where $w(t)$ is the window function. In this study, we used a hamming window with a window width of 2 s and the step length was set to 0.08 s. Length of each fast Fourier transform (FFT) was 250, so the frequency resolution is 0.5 Hz. We used the frequency range from 0.5 Hz to 62.5 Hz (the Nyquist frequency). The CWT of the sequence $x(t)$ is defined as:

$$CWT(a, b) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}) dt \tag{4}$$

where a is the scale factor and b is the time shift factor, $\psi(t)$ is the function of wavelet basis. The scale can be converted to frequency by:

$$F = \frac{F_C * f_s}{a} \tag{5}$$

where F_C is the center frequency of the wavelet basis, f_s is the sampling frequency of signal. In this study, we used a Morlet wavelet basis, and it is defined as:

$$\psi(t) = \exp(-\frac{t^2}{2}) \cos(5t) \tag{6}$$

We used a specific set of scales, and it is a geometric sequence, which makes the scales evenly distributed in the logarithmic domain. The corresponding frequencies of the scales range from 0.5 Hz and 62.5 Hz which are also evenly distributed in the logarithmic domain. The time shift factor increases evenly with a length of 0.08 s.

At last, the PWVD is defined as:

$$PWVD(t, w) = \int_{-\infty}^{+\infty} w(\tau) x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-jw\tau} d\tau \tag{7}$$

where $w(t)$ is the window function. In this study, we used a hamming window, and the step length was set to 0.08 s.

After the time-frequency transform is done, the time-frequency distribution amplitude matrix can be calculated as follows:

$$|Y(a, b)| = \sqrt{Y^2(a, b)} \tag{8}$$

where $Y(a, b)$ is the time-frequency distribution matrix. Then the matrix was scaled to a range between 0 and 1, the results of time-frequency transform after normalization were showed in Fig. 4.

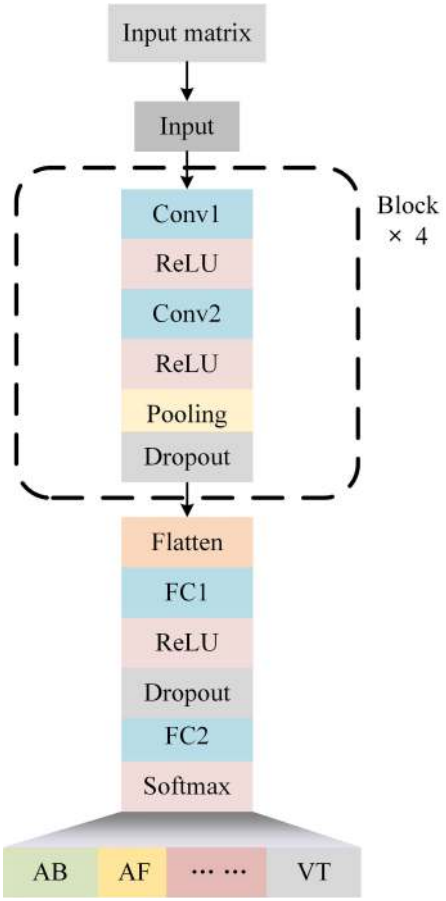

FIGURE 5. The basic architecture of the designed network.

D. THE ARCHITECTURE OF DESIGNED CNN

CNN is a widely used deep learning method. In this study, we designed a relatively common but very effective convolutional neural network. The network contains several common structures in CNNs: convolutional layer, pooling layer, fully connected layer, and dropout layer.

The basic architecture of the designed network is shown in Fig. 5. It consists of three parts: an input layer to input the matrix, the convolutional blocks to extract the deep features of the input tensor, and the fully connected layer to connect all the features output by the convolutional block and to finally output the classification result. Table 2 lists the configurations of each layer of the network. Each part of the network is described as follows:

1) INPUT LAYER

The network input is a time-frequency distribution matrix, which is a single-channel input, similar to the gray-scale image in image processing.

2) CONVOLUTIONAL BLOCK

Each convolutional block consists of two convolution layers, a pooling layer, and a dropout layer. The parameters of the two convolution kernels are the same. Each convolutional layer has 64 convolutional kernels with a kernel size of 5×5 and a step size of 1, and the activation function is Rectified Linear Unit (ReLU). After the input tensor is processed by the two convolutional layers, a maximum pooling of 2×2 is performed to reduce the number of features. A dropout layer with a dropout rate of 0.3 was added after the pooling layer. It is used to reduce overfitting of the network and speed up the training process by randomly removing some of neurons in each training iteration [43]. In this study, we used four convolutional blocks with the same parameters.

3) FULLY CONNECTED LAYER

The input matrix is processed by four convolutional blocks to obtain deep features of multiple channels. These features are combined into a one-dimensional feature through a flatten layer, then the features are processed by the first fully connected layer, which has 128 cells, and an activation function of ReLU. A dropout layer with a dropout rate of 0.5 was added after the first fully connected layer. The last fully connected layer outputs the classification result. The output number is 12, and the activation function is the softmax function which outputs the probability that the input matrix belongs to each category.

For the sake of simplicity, the network whose time-frequency distribution matrix was obtained by STFT is called Net1, by CWT is called Net2 and by PWVD is called Net3.

E. THE TRAINING METHOD

In the training process, the loss function is defined as cross entropy:

$$loss = -\sum_{i=1}^{n} y_i \log(y_i') \tag{9}$$

TABLE 2. Summaries of the network.

Layer	Output number	Kernel size/Pool size	Stride	Activation function	Padding	Dropout rate
Input layer	1	-	-	-	-	-
Conv1_1	64	5x5	1x1	ReLU	same	-
Conv1_2	64	5x5	1x1	ReLU	same	-
MaxPool1	64	5x5	2x2	-	valid	-
Dropout1	64	-		-	-	0.3
Flatten	-	-		-	-	-
FC1	128	-		ReLU	-	-
Dropout5	128	-		-	-	0.5
FC2	12	-		Softmax	-	-

where y_i is the actual label and y_i' is the output of the network. The root mean square prop (RMSprop) method was applied to update the weights of network. The update functions for weight W and offset b are defined as:

$$W = W - \alpha \frac{dW}{\sqrt{s_{dw}} + \varepsilon} \tag{10}$$

$$b = b - \alpha \frac{db}{\sqrt{s_{db}} + \varepsilon} \tag{11}$$

where s_{dw}, s_{db} is the accumulation of momentum and bias in previous iteration, α is the learning rate, ε is a small number to prevent the denominator from being zero. In this study, the initial learning rate, decay of learning rate, number of epochs and batch size were set to 0.0001, 0.0000001, 60 and 64.

As showed in Table 1, the dataset has some certain limitations. For example, some rhythm classes have a small number of samples, and the dataset is unbalanced. In order to obtain a stable and reliable model, we used a 5-fold cross-validation and adopted an oversampling method for the unbalanced training set to balance the number of samples in each training set. Fig. 6 shows the process of cross-validation and oversampling. We divided the dataset into five equal sized parts. In each training process, one of them was used as the validation set and the other four were used as the training set. In order to equalize the amount of samples for each rhythm class in the training set, we randomly copied some samples of the same rhythm class and added them to the training set. Then we evaluated the performance of the algorithm using the entire dataset.

FIGURE 6. The cross-validation and oversampling method.

III. RESULTS

A. TRAINING PROCESS

The experiments were conducted on a computer with 1 Intel Core i7-8700 CPU at 4.2 GHz, 1 NVIDIA GTX-1060 GPU and 32-Gb RAM. The models were trained efficiently in GPU using the Keras [44] deep learning framework with a tensorflow backend [45]. The average time for 1 iteration was 4 millisecond, and 63 s for 1 epoch.

B. COMPARISONS BETWEEN TRADITIONAL FEATURES AND DEEP FEATURES

In a large number of previous studies, the traditional method of classifying ECG signals often relied on the selection and calculation of various characteristics, which are dependent on the extraction of ECG features, such as the location of R wave and P wave, etc. When there are false or miss detections, the performance of the algorithm will be significantly reduced. Besides, the selection of characteristics also depends on the experience of researchers, and these characteristics are often not comprehensive.

In this paper, the higher-level and more abstract time-frequency features extracted by the CNNs contain more information and are more stable than the traditional features. Fig. 7 shows the 16 of 64 feature maps output by the second convolutional block of Net1. The feature maps output by the same convolutional layer represent features of the input data on different bases (convolution kernels) at the same level, which contain multifaceted information of the input data. As the network goes from shallow to deep, feature maps output by different layers change from specific to abstract, which is a process changing from local to global.

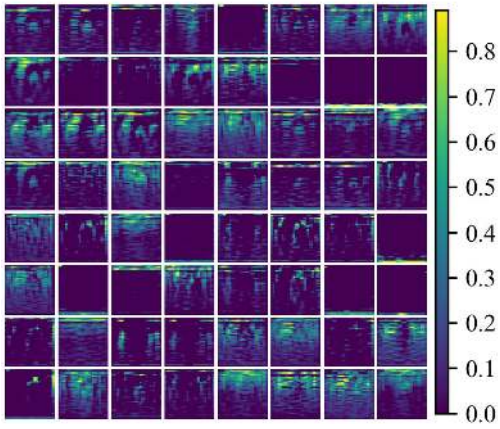

FIGURE 7. Feature maps output by the second convolutional block of Net1.

C. RESULTS AND PERFORMANCE EVALUATION

A five-fold cross-validation is introduced to train and test CNNs. All the 10519 segments of ECG data were used to evaluate the performance of the proposed method. The confusion matrices for each classifier is drawn in Fig. 8. The performance was evaluated using metrics including the accuracy, sensitivity, specificity, and F_1 score which are defined as follows:

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

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

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

$$F_1 = \frac{2TP}{2TP + FN + FP} \tag{15}$$

where TP, FN, FP, TN are true positive, false negative, false positive and true negative for each rhythm. The definition of these indexes are listed in Table 3 using NSR rhythm as an example.

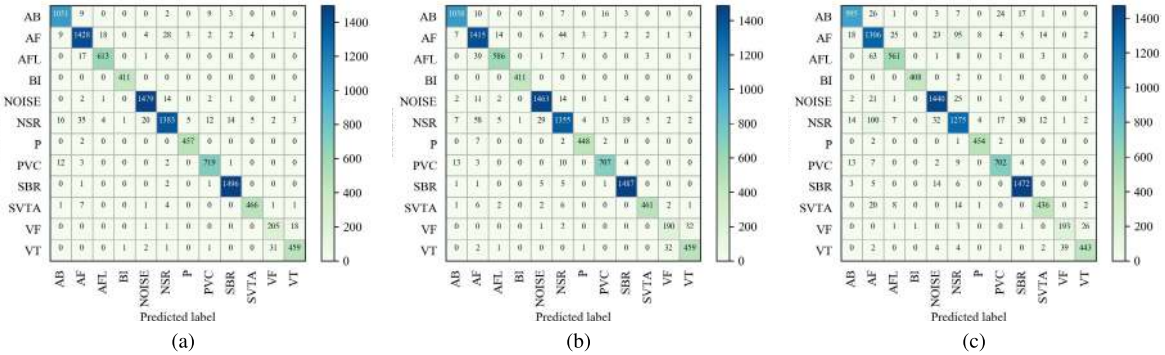

FIGURE 8. Confusion matrices for each classifier: (a) Net1, (b) Net2, (c) Net3.

TABLE 3. Definition of indexes related to NSR rhythm.

	Label = NSR	Label = other rhythms
Predict = NSR	TP(NSR)	FP(NSR)
Predict = other rhythms	FN(NSR)	TN(NSR)

The sensitivity, specificity, and F_1 score for each of the 12 rhythms are listed in Table 4 for each network. Meanwhile, Table 5 shows the accuracy, the average sensitivity, the average specificity, and the average F_1 score.

We also use the receiver operating characteristic (ROC) curve to measure the performance of the classifier. The ROC curves of Net1, Net2 and Net3 are shown in Fig. 9, and their area under curve (AUC) values are 0.9987, 0.9983 and 0.9963, respectively.

Considering all the evaluation metrics, Net1 performed best in the dataset we selected, followed by Net2 and the worst Net3.

IV. DISCUSSION

The performance of ECG classifiers is often affected by many factors, such as insufficient samples, the signal differences caused by different sensor types and measurement environments, and the physical conditions of different subjects. To overcome such problems and to get a more stable and robust model, we used multiple databases. The experiment results show that the performance is excellent in our selected dataset. This indicates that the proposed method is suitable for a variety of situations and may be applied in clinical practice.

According to the literature, we found that in the same kind of research field using the time-frequency distribution combined with deep learning method, the classification of heart rhythm types and the analysis of time-frequency method in this paper is the most comprehensive. In other words, this study was the first work to classify a wide range of ECG rhythms using time-frequency analysis and CNNs, and the performances of different methods of time-frequency transform were evaluated for the first time. With multiple databases used, the proposed method had good stability and robustness. The results preliminarily imply that this method is an excellent method to identify abnormal heart rhythm.

For the results of Net3, where PWVD and CNN were used and get the worst effect, we can speculate that although the window function is used to limit the influence of cross terms, the influence of cross-terms in PWVD cannot be ignored. The erroneous energy distribution and ambiguity caused by the cross-terms will affect the information of time-frequency domain, thus affecting the classification effect [40]. For ECG signals which frequency domain components are more complex, the impact of cross-terms is significant. And in this study, we skipped some points between each time-frequency transformation and resulted a loss of partial information of an ECG signal when using the quadratic time-frequency method, causing the decreasing of classification performance. For the Net2 where the CWT was applied, we used scales evenly distributed in the logarithmic domain, and it made the resolutions of high frequency bands lower than the low frequency parts, thus resulting a less information of high frequency bands and a decreasing of classification performance.

Compared with the traditional methods, the proposed method eliminates the extraction or calculation of manually selected features of ECG signals, and extracts the abstract and deep features in the time-frequency domain directly through the deep network. The features obtained by this method are more comprehensive and efficient, and in terms of the actual effect, it also shows excellent classification performance. Some traditional methods based on HRV or RR interval analysis require a long duration signal segment, while this method only requires a short length of 10 s. Once the network is trained, the classification result can be obtained quickly by inputting the time-frequency distribution matrix into the network. It is more efficient than traditional methods.

For now, some end-to-end neural networks had been used to classify ECG signals [21], [46], [47]. For one-dimensional (1-D) CNNs, the depths of networks were too deep, and it required a higher computing costs. The time-frequency matrix was already an advanced feature, and the network architecture was rather simple. The research of image classification using two-dimensional (2-D) CNNs was better developed and maturer than time sequence classification using 1-D CNNs. Although end-to-end networks had many advantages,

TABLE 4. Classification performance for each rhythm class using the proposed method.

Classes	Sensitivity(%)			Specificity(%)			F_1 score(%)		
	STFT	CWT	PWVD	STFT	CWT	PWVD	STFT	CWT	PWVD
AB	97.86	96.65	92.64	99.60	99.67	99.47	97.18	96.87	93.91
AF	95.20	94.33	87.07	99.16	98.48	97.27	95.07	92.73	85.58
AFL	96.23	91.99	88.07	99.77	99.76	99.56	96.31	93.99	90.41
BI	100.00	100.00	99.27	99.98	99.99	99.93	99.76	99.88	98.79
NOISE	98.60	97.53	96.00	99.68	99.51	99.12	98.34	97.31	95.40
NSR	92.20	90.33	85.00	99.33	98.92	98.07	93.99	91.80	86.47
P	99.56	97.60	98.91	99.92	99.92	99.87	98.92	97.92	98.06
PVC	97.56	95.93	95.25	99.72	99.63	99.48	96.97	95.54	94.23
SBR	99.73	99.13	98.13	99.76	99.65	99.27	99.17	98.51	96.91
SVTA	96.88	95.84	90.64	99.91	99.90	99.68	97.49	96.85	91.89
VF	91.11	84.44	85.78	99.66	99.63	99.61	88.17	83.89	84.28
VT	92.73	92.73	89.49	99.76	99.59	99.67	93.83	92.26	91.25

TABLE 5. Classification performance for all rhythm classes.

Method	Accuracy(%)	Sensitivity(%)	Specificity(%)	F1score(%)
STFT	96.65	96.47	99.68	96.27
CWT	95.26	94.71	99.55	94.80
PWVD	92.07	92.19	99.25	92.26

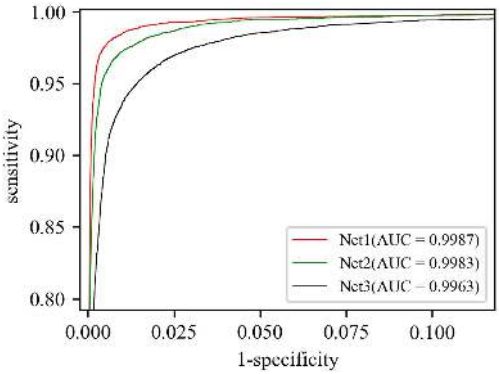

FIGURE 9. ROC curves for each classifier.

considering the computational efficiency, the method of using time-frequency transform and 2-D CNNs is still of significance.

For the classification of multiple arrhythmias, the problem of insufficient sample size and unbalanced sample distribution is inevitable. We met the same problems even though multiple authoritative databases were contained in this study. To get a more stable and robust model, we used a cross-validation method to compensate for the insufficient sample size and an oversampling method to process the training data to deal with the unbalanced data distribution. The results of cross-validation showed that the proposed method has stable performance on the whole dataset. From the results, we can speculate that the effect of unbalanced data on the results still exists. The two heart rhythms with the fewest samples are VF and VT, whose performances are not as good as others. However, it should be noticed that for some heart rhythms such as BI we have achieved a better classification performance than other rhythms. We can infer that the reason lies in the small sample size of patients with BI heart rhythm in the database, which leads to the lack of individual differences in the interception of BI segments. In fact, the ECG signals of different patients with the same arrhythmia may vary from each other due to many factors, such as the way the electrodes are attached, the physical conditions of the human body, and the measuring environment. Another point to be aware of is the classification accuracy of NSR heart rhythm is a little lower, and the confusion matrix shows that a lot of samples of NSR heart rhythm are misclassified as AF or NOISE. We suspect that these signals in the database carry a degree of noise, so they are easily misidentified. In future studies, we expect that the proposed method will perform well in a larger dataset of more patients and be evaluated by clinical data.

V. CONCLUSION

In this study, based on multiple authoritative databases, we constructed a more balanced ECG dataset including the noise-contaminated ECG signals. We developed a method to detect 12 rhythm classes using the time-frequency distribution and CNNs. This work was the first to evaluate the classification performance over a wide range of arrhythmias with different time-frequency analysis methods and CNNs. This method does not rely on extraction and calculation of morphological features such as R wave, P wave, or any other manually selected features. Under the same network structure, the method using STFT and a CNN achieved the best classification performance, with an average accuracy of 96.65%, an average sensitivity of 96.47%, an average specificity is 99.68%, an average F_1 score of 96.27% and a value of AUC of 0.9987. The results demonstrate that the proposed approach can classify a broad range of distinct arrhythmias from single-lead ECGs with high diagnostic performance. We are going to evaluate this method in clinical data and apply this method to clinical practice. Through subsequent studies and validations by more data, this method will improve the efficiency of dynamic ECG interpretation.

ACKNOWLEDGMENT

The authors would like to thank other students in the Electrophysiology and Pacing Laboratory of the Department

of Electronic Engineering, Fudan University. *(Ziqian Wu and Tianjie Lan are co-first authors.)*

REFERENCES

[1] World Health Organization, *Global Status Report on Noncommunicable Diseases 2014*. Geneva, Switzerland: World Health Organization, 2014.

[2] G. V. Naccarelli, H. Varker, J. Lin, and K. L. Schulman, "Increasing prevalence of atrial fibrillation and flutter in the united states," *Amer. J. Cardiol.*, vol. 104, no. 11, pp. 1534–1539, 2009.

[3] E. Sandøe and B. Sigurd, *Arrhythmia: A Guide to Clinical Electrocardiology*. Bingen, Germany: Publishing Partners, 1991.

[4] D. M. Mirvis and A. L. Goldberger, "Electrocardiography," *Heart Disease*, vol. 1, pp. 82–128, 2001.

[5] P. E. Dilaveris, E. J. Gialafos, S. K. Sideris, A. M. Theopistou, G. K. Andrikopoulos, M. Kyriakidis, J. E. Gialafos, and P. K. Toutouzas, "Simple electrocardiographic markers for the prediction of paroxysmal idiopathic atrial fibrillation," *Amer. Heart J.*, vol. 135, no. 5, pp. 733–738, 1998.

[6] P. De 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.

[7] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in poincare plot," *Biomed. Eng. Online*, vol. 8, no. 1, p. 38, 2009.

[8] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals," *Med. Biol. Eng. Comput.*, vol. 39, no. 6, pp. 664–671, Nov. 2001.

[9] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris, "Arrhythmia classification using the RR-interval duration signal," in *Proc. Comput. Cardiol.*, Sep. 2002, pp. 485–488.

[10] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[11] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986.

[12] F. Gritzali, G. Frangakis, and G. Papakonstantinou, "Detection of the *p* and *t* waves in an ECG," *Comput. Biomed. Res.*, vol. 22, no. 1, pp. 83–91, 1989.

[13] C.-H. Lin, "Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier," *Comput. Math. Appl.*, vol. 55, no. 4, pp. 680–690, 2008.

[14] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, 2013.

[15] M. Thomas, M. K. Das, and S. Ari, "Automatic ECG arrhythmia classification using dual tree complex wavelet based features," *Int. J. Electron. Commun.*, vol. 69, no. 4, pp. 715–721, Apr. 2015.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[17] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech signal Process.*, May 2013, pp. 6645–6649.

[18] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 25. New York, NJ, USA: Determination Press, 2015.

[19] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, "Convolutional neural network with shape prior applied to cardiac MRI segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1119–1128, May 2019.

[20] J. Shi, Z. Li, S. Ying, C. Wang, Q. Liu, Q. Zhang, and P. Yan, "MR image super-resolution via wide residual networks with fixed skip connection," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1129–1140, May 2019.

[21] A. Y. Hannun, 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, no. 1, pp. 65–69, 2019.

[22] 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.

[23] 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.

[24] M. M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, and B. BenJdira, "Convolutional neural networks for electrocardiogram classification," *J. Med. Biol. Eng.*, vol. 38, no. 6, pp. 1014–1025, 2018.

[25] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Comput. Biol. Med.*, vol. 93, pp. 84–92, Feb. 2018.

[26] 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.

[27] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[28] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.

[29] S. D. Greenwald, "The development and analysis of a ventricular fibrillation detector," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 1986.

[30] G. Moody, "A new method for detecting atrial fibrillation using R-R intervals," in *Proc. Comput. Cardiol.*, 1983, pp. 227–230.

[31] S. Petrutiu, A. V. Sahakian, and S. Swiryn, "Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans," *EP Europace*, vol. 9, no. 7, pp. 466–470, 2007.

[32] G. B. Moody, W. K. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, vol. 11, no. 3, pp. 381–384, 1984.

[33] Q. Li, C. Rajagopalan, and G. D. Clifford, "A machine learning approach to multi-level ECG signal quality classification," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 435–447, Dec. 2014.

[34] N. V. Thakor, J. G. Webster, and W. J. Tompkins, "Estimation of QRS complex power spectra for design of a QRS filter," *IEEE Trans. Biomed. Eng.*, vol. BME-31, no. 11, pp. 702–706, Nov. 1984.

[35] J. B. Tary, R. H. Herrera, J. Han, and M. van der Baan, "Spectral estimation—What is new? What is next?" *Rev. Geophys.*, vol. 52, no. 4, pp. 723–749, 2014.

[36] J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-25, no. 3, pp. 235–238, Jun. 1977.

[37] J. Morlet, G. Arens, E. Fourgeau, and D. Glard, "Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media," *Geophysics*, vol. 47, no. 2, pp. 203–221, 1982.

[38] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.

[39] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," in *Part I: Physical Chemistry. Part II: Solid State Physics*. Berlin, Germany: Springer, 1997, pp. 110–120.

[40] T. Claasen and W. Mecklenbrauker, "The Wigner distribution—A tool for time-frequency signal analysis," *Philips J. Res.*, vol. 35, no. 3, pp. 217–250, 1980.

[41] P. Goncalves and R. G. Baraniuk, "Pseudo affine Wigner distributions: Definition and kernel formulation," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1505–1516, Jun. 1998.

[42] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[44] F. Chollet *et al. Keras*. Accessed: 2015. [online] Available: https://github.com/fchollet/keras

[45] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.

[46] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, no. 1, pp. 81–90, Sep. 2017.

[47] D. Lai, X. Zhang, Y. Bu, Y. Su, and C.-S. Ma, "An automatic system for real-time identifying atrial fibrillation by using a lightweight convolutional neural network," *IEEE Access*, vol. 7, pp. 130074–130084, 2019.

ZIQIAN WU was born in Yiyang, China, in 1996. He received the B.Sc. degree in electronic engineering from Fudan University, Shanghai, China, in 2017, where he is currently pursuing the M.E. degree. His research interests are cardiac mapping and cardiac electrophysiology.

TIANJIE LAN was born in Hubei, China, in 1996. She received the B.S. degree in electronic engineering from Fudan University, Shanghai, China, in 2018, where she is currently pursuing the master's degree. Her research interests are cardiac electrophysiology and deep learning.

CUIWEI YANG received the B.Sc. degree from Zhejiang University, in 1992, the M.Sc. degree from the Huazhong University of Science and Technology, in 1995, and the Ph.D. degree in biomedical engineering from Fudan University, China, in 2008. She is currently a Professor with the School of Information Science and Technology, Fudan University. Her research interests fall in medical signal detection and processing, medical instrumentation, and cardiac electrophysiology, especially in cardiac mapping.

ZHENNING NIE received the M.D. degree from Fudan University, in 2000. He is currently an Attending Physician with the Cardiology Department, Zhongshan Hospital affiliated to Fudan University. He specializes in clinical diagnosis and treatment of cardiovascular diseases. His research works focus on the mechanism analysis and catheter ablation of arrhythmias.