Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS. 2017. Doi Number*

# A deep biometric recognition and diagnosis network with residual learning for arrhythmia screening using electrocardiogram recordings

**HAO DANG[1,2], YARU YUE[1,2], DANQUN XIONG[3], XIAOGUANG ZHOU[1,2], XIANGDONG XU[3], and XINGXIANG TAO[1,2]**

[1] School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] Engineering Research Center of Information Network, Ministry of Education, Beijing 100876, China
[3] Department of Cardiology, Jiading District Central Hospital Affiliated Shanghai University of Medical and Health Sciences, Shanghai 201800, China

Corresponding author: Xiaoguang Zhou (zxg@126.com), Xiangdong Xu (xuxiangdong8416@163.com), and Xingxiang Tao (taoxx@bupt.edu.cn)

This work was supported in part by the Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) under Grant SKLNST-2018-1-18, and in part by the Foundation of Jiading Health Science under Grant 2018-KY-04.

**ABSTRACT** Arrhythmia is one of the most persistent chronic heart diseases in the elderly and is associated with high morbidity and mortality such as stroke, cardiac failure, and coronary artery diseases. It is significant for patients with arrhythmias to automatically detect and classify arrhythmia heartbeats using electrocardiogram (ECG) signals. In this paper, we develop three robust deep convolutional neural network (DCNN) models, including a plain-CNN network and two MSF(multi-scale fusion)-CNN architectures (A and B), to aid in better feature extraction for the detection of arrhythmia and thus significantly improve the performance metrics. The proposed models are trained and tested with a public MIT-BIH arrhythmia database on five types of signals. Six groups of ablation experiments are conducted to analyze the performance of the models. The accuracy, sensitivity, and specificity obtained from MSF-CNN architecture A are higher than those from the plain-CNN model, demonstrating that the different parallel group convolution blocks (1×3, 1×5, and 1×7) dramatically improve a model's performance. Additionally, the best model MSF-CNN architecture B achieves an average accuracy, sensitivity, and specificity of 98.00%, 96.17%, and 96.38%, respectively. This illustrates the method with residual learning and concatenation group convolution blocks has a profound effect on the feature learning of the model. The results of ablation experiments show that our proposed biometric recognition and diagnosis network with residual learning (MSF-CNN B) achieves a rapid and reliable diagnosis approach on ECG signal classification, which has the potential for introduction into clinical practice as an excellent tool for aiding cardiologists in reading ECG heartbeat signals.

**INDEX TERMS** Heartbeat, Arrhythmia, Deep Learning, Convolutional Neural Network, Electrocardiogram Signal

## I. INTRODUCTION

Arrhythmias are an important group of cardiovascular diseases that are characterized by slow, fast, or irregular heartbeats [1,2]. They may occur alone or in conjunction with other cardiovascular diseases. Some serious arrhythmias also may occur suddenly and lead to sudden death, stroke, cardiac failure, and coronary artery diseases [3].

Electrocardiogram (ECG), a noninvasive, inexpensive, and reliable diagnostic tool, which reflects the specific changes in electrical signal activity over time. It is an important standard in the diagnosis of arrhythmias [4]. ECG signals include important morphological information, which are usually obtained by ECG inspection equipment, such as electrocardiograph, 24-hour Holter, and wireless wearable devices [5]. And they are widely used in the analysis of cardiac function. Cardiac arrhythmias are currently diagnosed by manual interpretation of the ECG signal. To automatically diagnose arrhythmias through ECG records, monitoring equipment must be able to analyze the morphological characteristics of ECG signals [6] as well as the correlation between heartbeats, and finally detect abnormal heartbeats and determine types.

According to the standard from the Association for the Advancement of Medical Instrumentation (AAMI) [7], ECG signals can be divided into five categories: normal beat (N), supraventricular ectopic (S), ventricular ectopic (V), fusion beat (F), and unknown beat (Q). The AAMI standard focuses on the detection of ventricular ectopic beats (VEBs) and non-VEBs, and each category includes several types of heartbeats. The specific classification is shown in Table 1. In Table 1, each heartbeat represents different cardiac activity patterns. Under different cardiac activity states, each ECG signal has a different implication and requires different targeted treatments [8]. At present, visual evaluation based on cardiologists is an important standard of diagnosis. It requires numerous well-trained specialists to correctly identify the type of signal, which not only leads to the deviation between subjective judgment and the actual situation [9], but also consumes considerable time and energy. Therefore, it is of utmost importance for cardiologists to automatically identify abnormal heart rhythms before clinical treatment.

TABLE 1 MAPPING OF THE MIT-BIT ARRHYTHMIA DATABASE HEARTBEAT TYPES TO THE AAMI STANDARD

| AAMI heartbeat types | N | S | V | F | Q |
|---|---|---|---|---|---|
| MIT-BIH heartbeat types | NOR<br>LBBB<br>AE<br>RBBB<br>Nodal(junctional) | AP<br>aAP<br>NP<br>SP | PVC<br>VE | fVN | P<br>fPN<br>U |

**Abbreviations:**
**Heartbeat types:** N: Any heartbeat not in the S, V, F, Q classes; S: Supraventricular Ectopic beat; V: Ventricular ectopic beat; F: Fusion beat; Q: Unknown beat; NOR: Normal beat; LBBB: Left bundle branch block beat; AE: Atrial escape beats; RBBB: Right bundle branch block beat; AP: Atrial premature beat; AAP: Aberrated atrial premature beat; PAC: Premature atrial contraction beat; NP: Nodal(junctional) premature beat; SP: Supraventricular premature beat; PVC: Premature ventricular contraction; VE: Ventricular escape beat; fVN: Fusion of ventricular and normal beat; P: Paced beat; FPN: Fusion of paced and normal beat; U: Unclassified beat.

Over the past decades, ECG signal recognition and classification have become an established technique that can effectively assist physicians in clinical diagnosis [4]. The relevant automatic recognition models mainly rely on traditional pattern matching methods. These methods have achieved great progresses, but the complex feature extraction process consumes considerable computing resources [4]. In recent years, deep learning has become a mainstream pattern recognition method. It is an end-to-end learning approach that does not require complex process of hand-crafted extracted features. Moreover, great achievements have been obtained in the fields of image classification [10-14], object detection [15-17], and image segmentation [18-21]. Therefore, in this paper, we introduce a deep learning technology into the study of one-dimensional signals and propose a more accurate, rapid, and robust discriminant model to analyze the classification of ECG signals.

This paper is organized as follows: Section II introduces literature related to the classification of ECG signals, including data pre-processing, machining learning methods, and deep learning methods. Then the database is described in section III. We propose a plain-CNN network and two MSF-CNN architectures (A and B) and deeply analyze the configuration parameters of three network architectures in section IV. In section V, the experimental results are shown in detail, and the performance evaluation is also compared with recent popular algorithms. Finally, we conclude our work and propose future research directions in section VI.

## II. RELATED WORK

In this section, we survey related literature on traditional machine learning approaches and recent popular deep learning methods based on the detection and classification of ECG signals. In general, traditional machine learning methods mainly consist of three steps for the classification of arrhythmias: data preprocessing, feature extraction and selection, and feature classification. However, the deep learning approach is an end-to-end model, which shows the capacity to self-learn from the input ECG signal segmentation.

### A. DATA PRE-PROCESSING

The pre-processing of ECG signals mainly includes denoising and segmentation. Firstly, the ECG signals are contaminated by various noise and artefacts [22]. In arrhythmias, as the ECG signals belong to low-amplitude and low-frequency signals, diverse noises lead physicians to perform an incorrect assessment and reduce the accuracy of diagnosis. Therefore, the denoising of ECG signals is a significant baseline [23] of data pre-processing. The goal is to reduce noises and artefacts and determine the point of interest, which is beneficial to extract effective waveform features from ECG signals. Many scholars have proposed different preprocessing methods. In general, they can be divided into four categories: filtering methods, transformation filtering methods, statistical methods, and a combination of these methods [24-28]. Additionally, the ECG signals segmentation is also necessary, which mainly divides the whole signal record into a large number of heartbeats or RR intervals, and the heartbeats or RR intervals belonging to same classification are grouped together according to the annotations of the expert.

### *B. MACHINING LEARNING METHODS*

In recent decades, traditional machine learning algorithms have been widely used in the classification of arrhythmia signals and have made remarkable achievements. The machining learning methods include the complex processing of feature extraction, feature selection and feature learning.

#### 1) FEATURE EXTRACTION AND SELECTION

Feature extraction and selection are a pivotal part of the classification of ECG signals in traditional machine learning methods, which is conducive to obtaining the most essential features of signals and providing an accurate feature for the final classification. The main features of ECG signals include time-domain features (also known as waveform features), frequency domain features, and statistical features [22].

Time-domain features mainly refer to physical parameters reflecting the activity regularity of the ECG signal, including the frequency and amplitude of each waveform, such as P-wave, Q-wave, R-wave, S-wave, T-wave, and intervals information, such as PR-interval, QT-interval, and RR-interval. The QRS-complex and RR-interval features from ECG signals are significant in the time-domain, which mainly reflect the position, duration, amplitude, and shape of a specific waveform or deflection in signals [29-30]. Otherwise, digital filters [31], neural networks [32], high-order moments [33], and phasor transforms [34] have also been used for detecting of the QRS-complex.

Frequency-based approaches are one of the most popular feature extraction techniques for representing ECG signals [22]. Many researchers claim the wavelet transform is the best approach for feature extraction and selection from the ECG signals [35]. Within the wavelet transform, the discrete wavelet transforms (DWTs) is the most widely used in ECG signal classification. In addition to DWT, continuous wavelet transforms (CWTs) are also used to extract features from ECG signals, which overcomes the disadvantages of representation coarseness and instability from DWT [36].

The main statistical features are the expectation, variance, maximum, minimum, standard deviation, and high-order moment of ECG signal [24]. In general, these features provide an effective method for analyzing the complexity and distribution of waves on any time series. Therefore, in the case of ECG recording, these functions are conducive for distinguishing the variation process of particular patients and diseases [22].

In general, the above feature extraction and selection methods are implemented in machine learning classification algorithms. In this work, we introduce the deep learning approach into 1-D ECG signal classification. It is an end-to-end model with self-learning. The features are automatically extracted from the ECG signals by the convolutional neural network. The hand-crafted feature extraction and selection process is unnecessary.

#### 2) FEATURE LEARNING METHODS

These methods are summarized according to different types of classifiers, including statistical methods [37], decision tree classification models [38-39], neural network methods [39-40], and support vector machines (SVMs) [40].

For example, Li and Zhou [38] presented an approach to classify ECG signals using wavelet packet entropy (WPE) and random forests (RF) following the recommendations from AAMI. The experimental results have shown that the WPE and RF methods are superior to several state-of-the-art competitive methods. A. M. Alqudah [37] introduced a novel method to model cardiac-related biological signals (ECG and PPG) based on Gaussian mixture waves. The proposed method has been applied to the MICIC and MIT-BIH arrhythmia databases.

Moreover, A. M. Alqudah et al. [39] utilized two classifier techniques, the probabilistic neural network (PNN) algorithm and random forest (RF) algorithm to extract gaussian mixture and wavelets features, which were applied to classify the ECG beat into six classes, normal beat (N), left bundle branch block beat (LBBBB), right bundle branch block beat (RBBBB), premature ventricular contraction (PVC), atrial premature beat (APB), and aberrated atrial premature (AAP).

Hammad et al. [40] employed four support vector machines (SVM), two Neural Networks (NNs), and a k-nearest neighbor (KNN) classifier to classify the ECG signals. These algorithms extracted 13 features from each ECG segmentation and set them as an input of the proposed classifier. All the records of the MIT-BIH arrhythmia database were used to validate these algorithms.

In general, although these above methods have shown favorable classification performances, they also have numerous shortcomings. First, these automatic ECG signal classification models mainly depend on machine learning and pattern recognition. In the process, ECG signal segmentations are regarded as a sequence of stochastic patterns. The hand-crafted extracted feature process requires burdensome computational resource and time. Second, in terms of classification algorithms and training datasets, the robustness of classification models is still limited because they fail to handle large intra-class variations. In addition, the above algorithms often subject to overfitting and show poor performance during validating the different datasets. Furthermore, the classifier algorithms don't perform well in practical applications under the condition of the various ECG signals from different patients, which shows a common disadvantage of inconsistent performance results when classifying a new ECG record. This makes them less reliable clinically or in practice. Finally, the recent ECG monitoring models require well-established cardiologists for diagnosis, which also consumes a lot of time and energy.

### *C. DEEP LEARNING METHODS*

Deep learning is a new technology that has become the mainstream in computer vision and pattern recognition. In the past few years, deep learning has been widely used in the fields of image classification [10-14], object detection [15-17], and image segmentation [18-21]. In recent years, deep learning-based methods have been successfully applied to analyze ECG signal so that overcome the challenges from traditional machine learning-based methods.

For example, Kiranyaz et al. [41] presented a fast and accurate patient-specific ECG classification system for recognizing the two types of signals of supraventricular ectopic beats (S) and ventricular ectopic beats (V). The model designed three convolutional layers and two multi-layer perceptron to obtain the experimental result.

In additional, Jun et al. [42] proposed a deep neural network for the classification of premature ventricular contraction (PVC) beats. Acharya et al [43] developed a 9-layer CNN model to automatically classify five classes of heartbeats. Murugesan et al. [6] also implemented three robust deep neural networks (DNNs) (CNNs, LSTM, and CNN-LSTM) to detect the two types of Premature Ventricular Contraction (PVC) and premature atrial contraction (PAC). The results showcased the potential of the network as a feature extractor for ECG signal classification.

Moreover, in [44], the CNN was transferred in this study to carry out automatic ECG arrhythmia diagnostics after employing the higher-order spectral algorithms. Transfer learning strategies were applied on a pre-trained convolutional neural network, namely AlexNet and GoogleNet, to carry out the final classification.

Compared with traditional machine learning methods, the most critical feature of deep learning is that it does not require the processes of feature extraction and feature selection. The deep learning approaches have the ability to self-learning from input signals. In other words, the previous processes of feature extraction and selection in machine learning are embedded in the deep learning model, which can continuously learn features from input data. However, the above deep learning methods also showcased some imperfections. The research directions of [41], [42] and [6] were a two-class problem. It was a simple research point compared to the five-class problem in this work. Otherwise, [37] and [40] presented a plain CNNs model to extract features from ECG signals. The structure of the plain model was not conducive to the extraction of features from deep layers. Moreover, [43] proposed 9-layer models, which is enough to features extraction. But the model didn't fully consider the imbalance between data classes, which may lead to the overfitting of model. Additionally, the influence of different lengths of input signal and the problem of unbalanced original data classification on model's performance has not been fully considered.

Broadly speaking, the fundamental disadvantages and challenges of existing machine learning methods for ECG signal detection and classification are that hand-crafted extracted feature, which not only greatly affects the accuracy of the algorithm, but also consumes a lot of calculation time and cost. The deep convolutional neural network is essentially realized by stacking automatic encoders. Considerable feature representational power effectively reveals unknown abstract features of input signals. It can achieve self-learning through end-to-end model design. Meanwhile, the radical problem of both methods is that they only focus on how to propose a better model, but do not pay attention to data processing issues: such as data denoising, data augmentation, and multi-scale data training and testing. The data preprocessing of signals should be focus on because signals and images are different data types.

Hence, in this work, inspired by these previous efforts, a more accurate, comprehensive, and robust method based on deep learning is proposed to identify five different types of arrhythmia signals. The proposed model not only pays attention to the superiority of model design but also presents the importance of data processing in this paper. The final results also prove that the application of ECG signal classification using the convolutional neural network is reliable. The deep learning architecture outperforms the hand-crafted feature extractors assembled by machine learning models in terms of classification accuracy, sensitivity, specificity, and confusion matrix.

The contributions of this work are as follows:

(1) We propose an end-to-end plain-CNN architecture and two MSF-CNN architectures (A and B) to replace additional hand-crafted feature extraction, selection, and classification using machine learning methods. The plain-CNN is a baseline model, the MSF-CNN A and B are implemented based on this baseline network. Thus, it significantly enhances the performance against recent state-of-the-art studies.

(2) Moreover, the signal processing problems are fully considered. We first design multi-scale input signals, including 251 samples (named set A) and 361 samples (named set B). This design can improve the generalization ability of the model by extracting multi-scale signal features. Then, the signal denoising and data augmentation also are implemented in this paper. The data augmentation strategy is a major innovation in this paper. This problem has not been paid much attention in most ECG signal research papers before.

(3) In particular, we present six sets of detailed ablation experiments on ECG signal classification and achieve excellent performance metrics. And we also compare the results from our model to recent state-of-the-art methods. Additionally, detailed analysis and comparison are presented in this paper.

## III. ECG DATABASE DESCRIPTION AND PRE-PROCESSING

It is crucial to acquire and process the research data in our work. In this section, we first introduce the MIT-BIH Arrhythmia Database in detail, and then we fully illustrate the data pre-processing, including denoising, data segmentation, and data augmentation.

### *A. THE DESCRIPTION OF DATABASE*

The MIT-BIH Arrhythmia Database (MITDB) [45] is an open-source PhysioBank database that is widely used to research the detection and classification of ECG signals. The

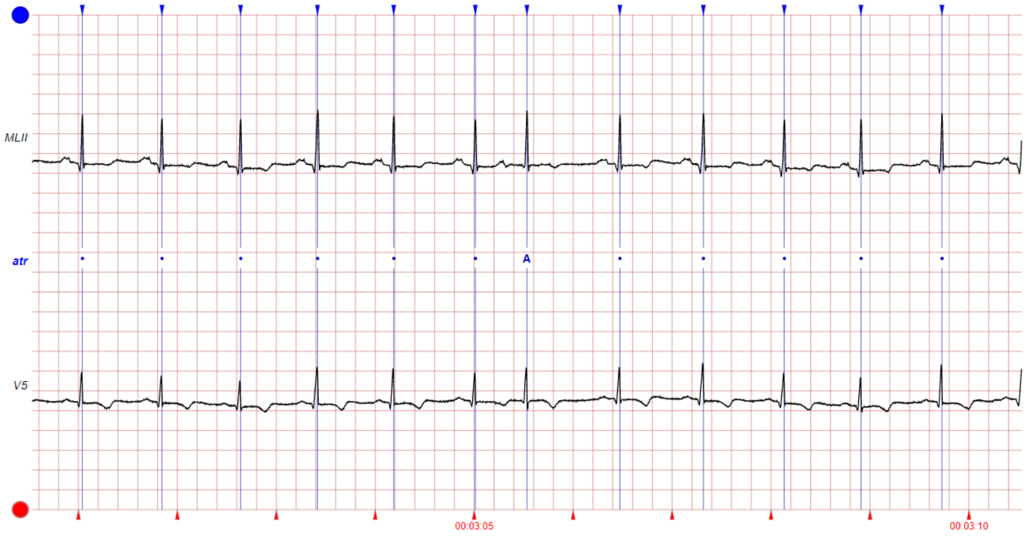


Figure 1. **A 10 s signal example of MLII and V5 from MITDB. Each ECG record is approximately 30 minutes, which includes two leads. The MLII of Figure 1 denotes the signal of lead II, and V5 describes the lead V.**

database consists of 48 half-hour ECG records obtained from 47 subjects, and each ECG record contains two leads (lead II and lead V) originating from different electrodes. Figure 1 shows an example of signals from the MITDB. Each ECG record duration is approximately 30 minutes, and the signal sampling frequency is 360Hz. These subjects comprise 25 males aged range from 32 to 89 and 22 females aged 23 to 89. The Arrhythmia database is divided into 25 subjects of normal ECG recordings and 23 subjects with abnormal ECG recordings.

In this paper, two-lead signals (lead II or MLII) are used to train, validate and test the algorithm. In addition, all the signal records are independently annotated by at least two cardiologists. A total of 109,454 heartbeats are extracted in this work (shown in Table 2). The data directory contains the entire MIT-BIH arrhythmia data, which uses a custom format to save file length and storage space. An ECG record consists of three parts: a header file (.hea), a data file (.dat), and an annotation file (.atr).

## *B. DATA PRE-PROCESSING*

We process the original raw data from the MIT-BIH arrhythmia database through a series of approaches such as denoising, data segmentation, and data augmentation to form the new data sets, and finally train a network with stronger robustness and better generalization ability. The specific processes are as follows:

### 1）DENOISING

The main function is to eliminate power-line interferences and baseline wanderings caused by patient respiration or movement, which will lead to several problems in detecting heart diseases. Baseline wandering is a low-frequency noise signal. For baseline wandering, the median filtering method is adopted to remove this kind of noise. Power-line interference is an interfering voltage with an integer multiple of 50 Hz that completely masks the ECG waveform [4]. Power-line interference and high-frequency noise are usually removed by a low pass filter. Considering the feature, first, the wavelet transform multi-resolution theory is leveraged to decompose the noisy signal. Then, we take advantage of the different distribution of signal and noise on the spectrum to remove the detail component on the scale of wavelet decomposition directly corresponding to the noise. Finally, wavelet inverse transformation is used to reconstruct signals, which can effectively remove the noise in the signal component.

### 2）DATA SEGMENTATION

The denoised ECG signals are classified into 5 classifications: normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q) according to the annotation from cardiologists, and these signals will be fed into the classification network. A complete normal heartbeat is shown in Figure 2, including an integrated rhythm from P-wave onset to T-wave offset (or U-wave onset). Considering the different lengths of ECG signals contain different amounts of feature information, data segmentation follows two strategies: 251 samples and 361 samples. The original raw ECG signals with denoising are segmented into a mass of heartbeats centered around the R-peak without the inclusion of the first and last heartbeats. Each heartbeat consists of 251 samples (60 samples before the R-peak and 190 samples after R-peak), including an integrated P-, Q-, R-, S-, and T-peak. We regard these signals included 251 samples as set A. Likewise, these original raw signals with denoising also are segmented into 361 samples of a heartbeat (120 samples before the R-peak and 240 samples after the R-peak). We regard these signals included 361 samples as set B.

### 3）DATA AUGMENTATION

It is an important part of this work, mainly to balance the number of five classifications (N, S, V, F, Q), which is more conducive to feature learning in deep neural networks. A total

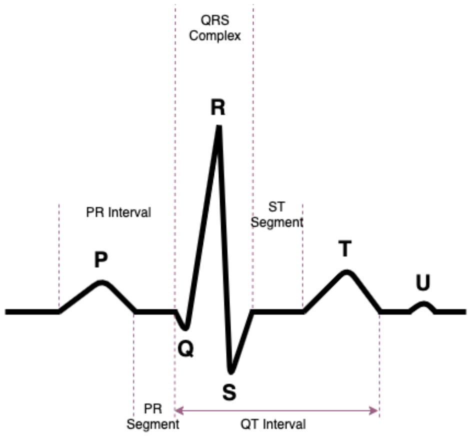


**Figure 2. A complete normal heartbeat. A complete heartbeat is a section of rhythm ranging from P onset to T offset (or U onset), consisting of P-wave, PR-interval, Q-wave, R-wave, S-wave, T-wave, QT-interval and U wave. Each waveform corresponds to the physiological process of cardiac excitement. The total duration of a heartbeat is approximately 0.8 s.**

of five types of ECG signals are considered in this work. As seen in Table 2, the number of samples in each category is different. The number of F signals is the lowest before data augmentation. Although unbalanced data distribution is more common in practical applications, the large difference in the number of categories is not beneficial to train the network model.

Therefore, the data augmentation approaches are leveraged to balance the types of signals. Additionally, the unbalanced data distribution is modestly maintained in this paper. Specifically, the number of segmentations in the N class remains invariable because they are the most adequate. The number of remaining classes (S, V, F, Q) is augmented to match the number in the N class. In this paper, three methods are leveraged to implement the data augmentation strategy. The first method is time shift augmentation, which randomly shifts the signal by rolling it along the time sequence. The second method is noise augmentation. We add random white noise with a damping coefficient of 0.4 to the original signal. We also combine two signals proportionally to obtain the new signals in the same category.

TABLE 2 THE DATA DISTRIBUTION OF HEARTBEATS IN THE MIT-BIH ARRHYTHMIA DATABASE

| Classification | Number of instances (without augmentation) | Number of instances (with augmentation) |
|---|---|---|
| N (Normal) | 90,595 | 90,595 |
| S (Supraventricular ectopic beat) | 2,781 | 55,620 |
| V (Ventricular ectopic beat) | 7,235 | 72,350 |
| F (Fusion beat) | 802 | 32,080 |
| Q (Unknow beat) | 8,041 | 80,410 |
| Total | 109,454 | 331,055 |

It should be noted that data augmentation is a process that generates new samples as a supplement to real data, which is applied only to the training processes. In testing, we leverage the original data without augmentation.

## IV. NETWORK ARCHITECTURE

In this section, we first introduce the model structure of the most popular convolutional neural networks. Then, three different architectures, a plain-CNN, and two MSF-CNN models (A and B), are proposed. The primary idea of the network is to build a robust MSF-CNN-based feature extraction to derive features from ECG signals. The network would also be easily adaptable to multiple datasets by transfer learning.

### *A. CONVOLUTIONAL NEURAL NETWORK*

Convolutional neural networks (CNNs) are one of the most frequently used in the field of artificial neural networks [46]. Since AlexNet [47] won first place in the ImageNet competition in 2012 by using a 7-layer CNN, CNN has been widely used in the fields of image classification, semantic segmentation, video recognition, and speech recognition and has also achieved great success. The standard architecture of CNNs includes six parts: the convolutional layer, pooling layer, rectified linear activation function, batch normalization, fully connected layer, and softmax function.

#### 1) CONVOLUTIONAL LAYER

Each convolutional layer is composed of several convolutional units, and all the parameters are optimized by the back-propagation algorithm. The main function of the convolution operation is to map the input to the hidden layer feature space so that extract different features from the input signal. The shallow layers can only extract some low-level local features such as edges, lines, and angles, while the deep layers iteratively extract corresponding detail features from high layers. The convolution operation is computed by the following equation (1).

$$y_n = \sum_{k=0}^{N-1} x_k f_{n-k} \tag{1}$$

where $x$ denotes the input signals, $f$ represents the convolution kernel, and $N$ is the number of elements in the input signal $x$. The output vector is denoted by $y$.

#### 2) POOLING LAYER

The pooling layer, namely down-samples, aims to reduce the number of feature maps so that it decreases the calculation cost by lessening the network parameters. The common pooling operations mainly include max-pooling and average-pooling. The max-pooling only outputs the maximum number in each kernel, thus reducing the size of the feature maps and retaining the local features. The average-pooling outputs the mean value in each kernel, thus aggregating the global feature information. It follows equation (2).

$$x_i = \max_{r \in R}[x_{i-1}(n \times s + r)]$$

$$or\ \ x_i = \underset{r \in R}{\text{mean}}[x_{i-1}(n \times s + r)] \qquad (2)$$

where max and mean denote the max-pooling and average-pooling, respectively. $s$ describes the stride. $n$ is the element index of a feature map. In this study, max-pooling is implemented in shallow layers, and mean-pooling is leveraged in deep layers. Thus, this configuration retains both global and local features.

#### 3) RECTIFIED LINEAR ACTIVATION FUNCTION

The rectified linear activation function implements nonlinear mapping from the output of the convolutional layer, realizing the nonlinear transformation between the input and output of the neuron. Nair et al. [48] has reported that faster convergence and higher accuracy can be obtained using ReLU. Hence, the activation function of ReLU is utilized in this paper. Its characteristic is fast convergence and reducing the disappearing gradient. The ReLU is computed by the following equation (3).

$$\text{ReLU(x)} = \begin{cases} x, x > 0 \\ 0, x \leq 0 \end{cases} \qquad (3)$$

#### 4) BATCH NORMALIZATION

It is complicated that training a CNN by the fact that distribution of each layer's inputs changes during training, because the parameters of previous layers usually change with the update of gradient. This makes it very difficult to train models, which requires lower learning rates and perfect parameter initialization to solve the problem. This phenomenon is called internal covariate shift. In order to overcome the problem, Loff et al. [49] proposed a method called Batch Normalization (BN), which demonstrates that the network training converges faster if its inputs are whitened (linearly transforming the input to have zero means and unit variances).

#### 5) FULLY CONNECTED LAYER

The fully connected layer plays the role of a classifier in the deep neural network. It implements a weighted sum of the feature from previous layers. The feature space is mapped to the sample marker space by a linear transformation.

#### (6) SOFTMAX FUNCTION

Softmax functions are often used in the last layer of the convolutional neural network, which is an output layer for multi-classification. Softmax function maps multiple scalars to a probability distribution with each value range of (0,1), which follows equation (4).

$$\partial(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \qquad for\ j = 1, \cdots, K \qquad (4)$$

The output of the softmax function is an $X$ dimensional vector, and $X$ is the number of classes. In this work, there are five classifications (N, S, V, F, and U).

### B. RESIDUAL LEARNING NETWORK

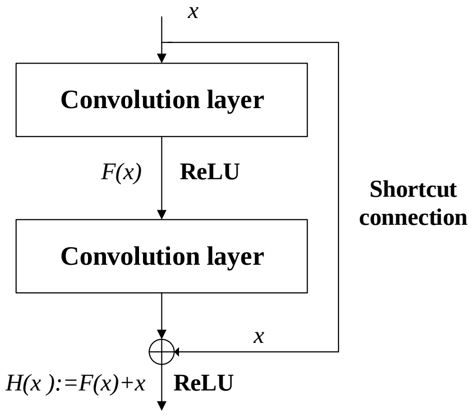


**Figure 3.** A building block of the residual learning network.

A residual learning network was first proposed in [10] about image classification, which resolves the degradation problem of deep networks. The degradation problem appears with the deepening of the network layer. The specific phenomenon is that the accuracy saturates and then decreases rapidly with increasing network depth. The residual learning network is implemented by identity shortcut connections. As shown in Figure 3, it directly skips one or more convolutional layers, so that the output from the first several layers is introduced into the input of the following layers. And it is also a vital innovation of this paper to introduce the residual learning block into the one-dimensional signal analysis.

The main reason that the residual network addresses the degradation problem is that the identity shortcut connections make every layer fit a residual mapping instead of requiring each few stacked layer to directly fit a desired underlying mapping. Formally, the desired underlying mapping is represented as $H(x)$, and we hope that each nonlinear layer will map $F(x) := H(x) - x$. The original mapping is recast into $F(x) + x$, which is implemented by a feedforward neural network with shortcut connections (Figure 3). Thus, the residual network optimizes the residual function $F(x) := H(x) - x$ instead of $H(x)$. Although both forms of the objective function can approximate the required function in principle, the difficulty of optimization is different. A large number of experiments also have confirmed this conclusion. If the optimal function is closer to the identity mapping than the zero mapping, it is much easier for the solver to optimize the residual function to zero than to fit identity mapping by nonlinear layers.

In detail, the residual learning block is divided into two parts: identity mapping and residual mapping. As shown in Figure 4, the shortcut connection of the right curve is identity mapping, and $F(x)$ is the residual learning block, which is composed of two convolutional layers in our work. In the network model, the number of feature maps from the input and output may be different, and there are two representations of the residual learning block following equations (5) and (6).

$$H(x) = F(x) + x \qquad (5)$$

Equation (5) is the representation of residual learning when the number of feature maps from the input and output is the same. If the number of feature maps from the input and output is different, the convolution of $1 \times 1$ will be leveraged to increase the dimension or decrease dimension.

$$H(x) = F(x) + h(x) \tag{6}$$

where $h(x)$ is a convolution operation of $1 \times 1$ added in the shortcut connection.

In addition to solving the degradation problem by optimizing the residual function, residual learning can also effectively reduce gradient dispersion.

When the layer of network becomes deep, the gradient back propagation is as follows.

$$\frac{\partial Loss}{\partial x_1} = \frac{\partial F_N(X_{L_N},W_{L_N},b_{L_N})}{\partial X_L} * \cdots * \frac{\partial F_2(X_{L_2},W_{L_2},b_{L_2})}{\partial X_1} \tag{7}$$

During the backpropagation of this gradient value, if $N$ is large, the gradient value will decrease as it propagates to the first few layers, and the gradient may disappear when it is deeper in the deep neural network. However, residual learning solves this problem at the level of the neural network structure. The gradient back propagation is as follows when the residual learning is utilized in the model.

$$\frac{\partial Loss}{\partial x_1} = \frac{\partial X_L + F(X_L,W_L,b_L)}{\partial X_L} = 1 + \frac{\partial F_2(X_L,W_L,b_L)}{\partial X_L} \tag{8}$$

Hence, even with deep network layers, gradient dispersion will be effectively contained.

### C. THE PROPOSED NETWORK ARCHITECTURE

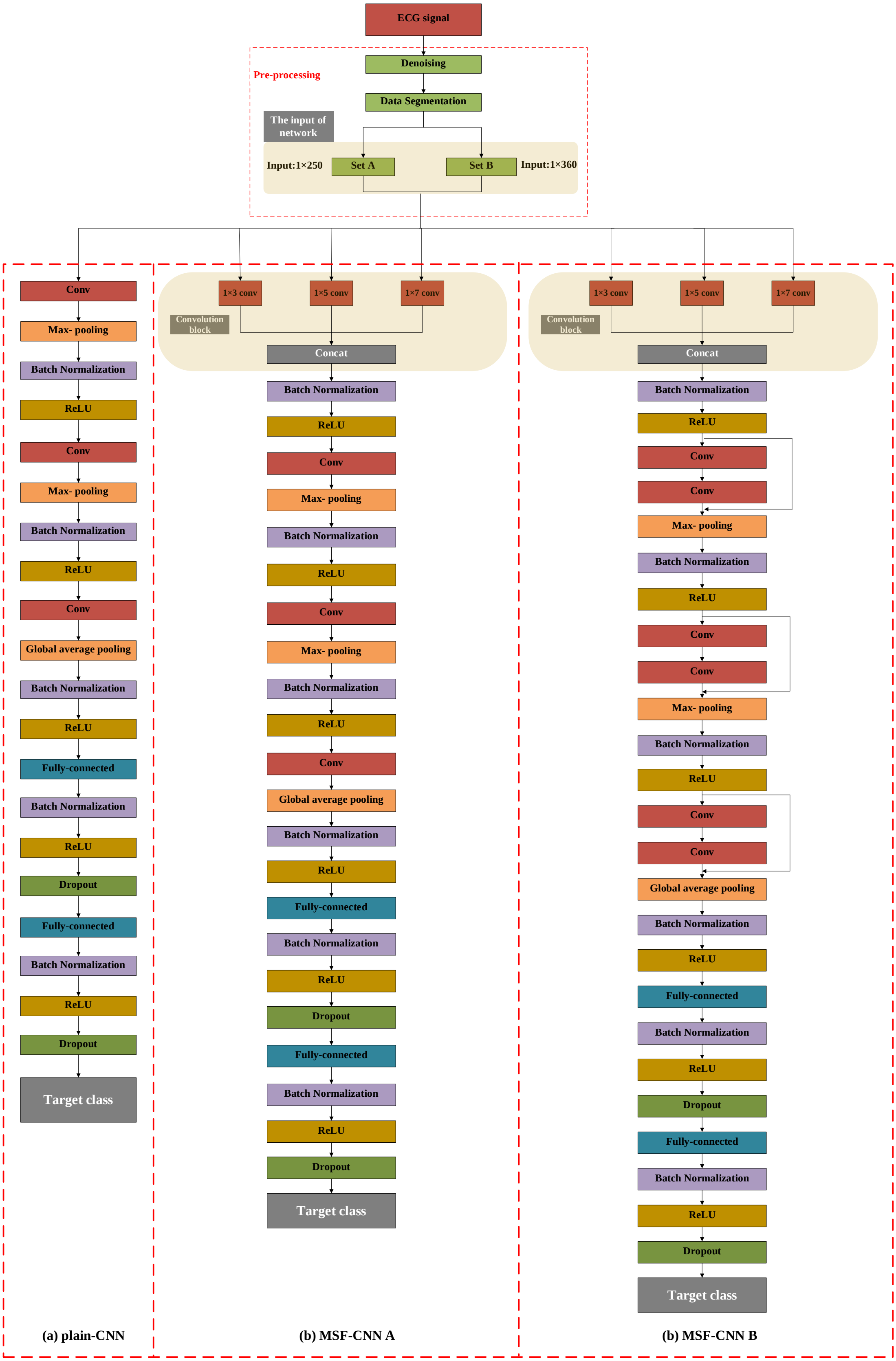
The design of the network mainly relies on the six parts computing units mentioned above. In this work, we design three network architectures (plain-CNN, MSF-CNN A, and MSF-CNN B.) with a highly modularized block, which are inspired by the idea of VGG published as a conference paper at ICLR 2015[50]. VGG is a mature deep neural network that has been proven to effectively solve various problems in the field of computer vision.

As shown in Figure 4 (a), the plain-CNN network, a baseline network, is a simple CNN architecture to verify the processing ability of 1-D CNN for ECG signals. It includes three convolution layers, two fully connected layers, and corresponding nonparametric layers (pooling layer, batch normalization layer, ReLU layer, and softmax layer). The input signals of set A and set B are directly fed into the convolution layer. The first two convolution layers are followed by a max-pooling layer, a batch normalization (BN) layer, and a ReLU layer, respectively. The last convolution layer is followed by global average pooling. The fully connected layer is followed by a BN layer, a ReLU layer, and a dropout layer. The plain-CNN is an ordinary multi-layer convolution network.

In addition, we propose a multi-scale fusion CNN architecture A (MSF-CNN A, in Figure 4 (b)) that integrates different spatial features by using one parallel group convolutional block (1×7,1×5, and 1×3). The MFS-CNN A is upgraded network based on the plain-CNN to verify the processing ability of three parallel convolution kernels for ECG signals. As shown in Figure4 (b), the network mainly includes one parallel group convolutional block, three convolution layers, two max-pooling layers, one global average-pooling layer, two full convolutional layers, and the corresponding BN, ReLU, and dropout. The datasets are first divided into two subsets (set A and set B) according to the different length of ECG signals and fed into three different parallel convolution kernels (1×7, 1×5, 1×3). The three outputs are then concatenated. This strategy can enable the network model to learn the hierarchical feature information from different spaces, and finally obtain more continuous and better representation. Then it is followed by the BN and ReLU layers. The trick of BN relieves overfitting, and ReLU increases nonlinear expression. The first two convolutional blocks contain a convolutional layer, max-pooling, BN and ReLU, and the last convolutional blocks are connected to a global max-pooling layer. The two fully connected layers are followed by BN, ReLU, and dropout operations. The MSF-CNN A is mainly introduced three parallel convolution kernels to fully extract the feature from set A and set B.

Finally, we design another multi-scale fusion CNN architecture B (MSF-CNN B, in Figure 4 (c)) based on the MSF-CNN A, which is inspired by VGGNets [50] and ResNet [10]. The MFS-CNN B is upgraded network based on the MFS-CNN A to verify processing ability of the concatenation group convolution blocks and residual learning blocks for ECG signals. The architecture includes one parallel group convolutional block (1×7, 1×5, and 1×3) as the MSF-CNN A, 7 convolution layers, two residual learning blocks, two max-pooling layers, one global average pooling, and two fully connected layers. The parallel group convolution block is the same as the MSF-CNN A. The difference between network A and B is that two or three convolutional layers (named the concatenation group convolution block) are grouped together in the deep layer of MSF-CNN B, sharing the same number of filters, and the concatenation group convolution blocks are separated by the max-pooling layer. Therefore, one parallel group convolutional block and two concatenation group convolutional blocks constitute the entire convolution MSF-CNN B, and the global average pooling layer is behind the third concatenation group convolutional blocks. Most importantly, we implement the residual learning block to avoid the degradation problem described above. The concatenation group convolution blocks and residual learning blocks are a vital innovation of this model.

In training, the operation of the fully connected layer is replaced by a full convolutional layer in the network. Since the output of the convolutional layer maintains the spatial locality between the feature signals, and the input size of ECG signals is not limited. Additionally, this conversion greatly reduces the number of parameters that need to be trained, and it can also provide a better effect. The corresponding function is shown in equation (9).



(a) plain-CNN

(b) MSF-CNN A

(b) MSF-CNN B

**Figure 4. Example network architecture. (a): the plain network as a reference. (b): the MSF-CNN architecture A. (c): the MSF-CNN architecture B. Table 2 shows more details and other variants.**

$$y_j = f\left(\sum_{i \in M} k_{ij} * x_i + b_j\right) \qquad (9)$$

Where $x$ and $y$ are the input and output of the network, respectively. $M$ is the convolution kernel size, $j$ denotes the index of convolution kernels, and $i$ denotes the index of input feature maps. $k_{ij}$ describes the convolution kernel for the $i-th$ input map and $j-th$ output map.

In the plain-CNN, the number of convolution kernels is 64 in the first convolutional layer and then increases by a factor of two after each max-pooling layer until it reaches 256. In the MSF-CNN A, the number of convolution kernels is also 64 in the parallel group convolutional block as the plain-CNN. However, it then increases by a factor of two after each max-pooling layer until it reaches 512. In the MSF-CNN B, the configuration of convolution kernels is the same as MSF-CNN A, and the number of convolution kernels is 64 in the parallel group convolutional block and then increases by a factor of two after each max-pooling layer until it reaches 512 in the concatenation group convolution blocks. The detailed configuration of the three network architectures evaluated in this paper is described in Table 3.

## V. ABLATION EXPERIMENTS

In this section, we first briefly describe the implementation details of the experiment and then introduce our performance metrics of the three models in our experiment. Finally, we carry out detailed experiments and performance comparison. Additionally, we also discuss the advantages and limitations of the proposed model.

### A. IMPLEMENTATION DETAILS

The network is designed with a fixed input of 251 (set A) and 361 (set B) samples, and the output is the probability of five categories. The outline of model is presented in Algorithm 1. Taking set B as an example, first, the original data is called set B after pre-processing, and set B is divided into trainSet and testSet. Then, trainSet is divided into 10 equal parts for cross-validation. Compared with the results $r_t$ of 10 cross-validation, the model *m* with the best performance is obtained through the validation and comparison of the training process. Finally, the testSet is loaded to evaluate the model.

The network model optimizes the cross-entropy function with the Adam optimizer, which is optimized by using a mini-batch size of 128 tensors on the 4 NVIDIA TITAN Xp GPUs. The Adam optimization is leveraged in this paper to update the parameters of the proposed network structure. It has been observed that it allows the network to converge at a fast rate, thus improving the efficiency of the training process. The mini-batch size is chosen as 128 to trade off two considerations. The size results in a short convergence time by reducing the variance of training and brings more power for Adam optimizer to jump out of shallow minima in training. According to the experiments, the learning rate starts from 0.001 and is divided by 10 when the error plateaus. The decay rate is also set to 0.0001. The initialization momentum is 0.5, and it is annealed to 0.9 after a multiple epoch gradually.

**Algorithm 1** MSF-CNN B

```
Input:
        SetA/SetB is the dataset;
        10 is cross-validation times;
        T is test data;
        optim Algorithm is Adam;
        D is pre-trained model;
        N is heartbeat classes
Output:
            The predicted probability p (·);
(trainSet; testSet) ← split (SetA/SetB)
S ← (split trainSet in equal parts of 10)
for each round t=1, 2, ... ,10 do
    {verify; train} ← {$S_t$; $S – S_t$}
    (tf; vf) ← (generate spam feature of train and verify)
    $m_t$ ← modelFit(Adam; tf)
    $r_t$ ← modelEvaluate(mt; vf)
end for
m ← bestModel(($m_t$; $r_t$)|t = 1, 2, , 10)
test ← (generate spam feature of testSet)
res ← modelEvaluate (m; test)
```

In the fully connected layer, dropout operation is adopted to reduce overfitting and improve generalization ability. Considering one-dimensional signals and the number of neurons, the dropout parameter is set to 0.3. According to equation (10), the cross-entropy loss function of five classification problems can be obtained.

$$\mathrm{L(X, y)} = \frac{1}{n}\sum_{i=1}^{n} \log p(y|x) \qquad (10)$$

where *X* is the input ECG signal, *y* is the ground truth of each input ECG signal, and *p* (·) is the predicted probability.

In addition, 10-fold cross-validation is leveraged to evaluate model performance. The original dataset is randomly divided into 10 equal-sized subsets. The 9 subsets are used for training, and the remaining subset is used to test the proposed model. The process is repeated according to iterations. The performance metrics (specificity, sensitivity, and accuracy) are evaluated in each epoch. Finally, the classification results of each validation are obtained and averaged to estimate the performance of the model on the whole dataset.

We find that gradient explosion and overfitting may exist in the comparative experiments. Therefore, to avoid these problems, regularization is introduced to our proposed model. In the experiment, the L2 norm of the model parameters (equation (11)) is implemented to relieve these problems. Specifically, the threshold is set to 0.5 to stabilize the training process.

$$l(\mathrm{x}) = \mathrm{L(X)} + \sigma \sum_{i=1}^{3} ||w_i||^2 \qquad (11)$$

where $l(x)$ is the loss function with L2 regularization and $L(x)$ is the cross-entropy loss function from equation (9). σ denotes a penalty factor, which is to balance the goal of achieving better training results and keeping smaller parameter values. Thus, the regularization can avoid overfitting effectively by narrowing down all the parameters.

TABLE 3 THE DETAILED CONFIGURATION OF THE PROPOSED NETWORKS

| Layers | Plain-CNN network | MSF-CNN architecture A | MSF-CNN architecture B |
|---|---|---|---|
| 1 | Conv (kernel size1×3, feature map 64) | Convolution block (kernel size 1×3, 1×5, 1×7, feature map 64) | Convolution block (kernel size1×3, 1×5, 1× 7, feature map 64) |
| 2 | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU | Concatenation<br>Batch Normalization<br>ReLU | Concatenation<br>Batch Normalization<br>ReLU |
| 3 | Conv (kernel size1×3, feature map 128) | Conv (kernel size1×3, feature map 128) | Conv (kernel size1×3, feature map 128) |
| 4 | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU | Conv (kernel size1×3, feature map 128) |
| 5 | Conv (kernel size1×5, feature map 256) | Conv (kernel size1×3, feature map 256) | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU |
| 6 | Global average pooling (stride 2)<br>Batch Normalization<br>ReLU | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU | Conv (kernel size1×3, feature map 256) |
| 7 | Fully connected layer (512) | Conv (kernel size1×5, feature map 512) | Conv (kernel size1×3, feature map 256) |
| 8 | Batch Normalization<br>ReLU<br>Dropout (0.3) | Global average pooling (stride 2)<br>Batch Normalization<br>ReLU | Max-pooling (stride 1)<br>Batch Normalization<br>ReLU |
| 9 | Fully connected layer (1024) | Fully connected layer (1024) | Conv (kernel size1×3, feature map 512) |
| 10 | Batch Normalization<br>ReLU<br>Dropout (0.3) | Batch Normalization<br>ReLU<br>Dropout (0.3) | Conv (kernel size1×3, feature map 512) |
| 11 | Softmax (5 classes) | Fully connected layer (1024) | Global average pooling (stride 2)<br>Batch Normalization<br>ReLU |
| 12 | ———— | Batch Normalization<br>ReLU<br>Dropout (0.3) | Fully connected layer (1024) |
| 13 | ———— | Softmax (5 classes) | Batch Normalization<br>ReLU<br>Dropout (0.3) |
| 14 | ———— | ———— | Fully connected layer (1024) |
| 15 | ———— | ———— | Batch Normalization<br>ReLU<br>Dropout (0.3) |
| 16 | ———— | ———— | Softmax (5 classes) |

$w_i$ describes the weight of $i-th$ layers.

### B. EVALUATION METRICS

For the evaluation, the four-standard metrics of accuracy, sensitivity (also known as recall), specificity (also known as the true negative rate), and confusion matrix are used to evaluate the classification performance of the plain-CNN, MSF-CNN A, and MSF-CNN B, respectively. Accuracy is defined as the ratio of the number of correct predictions (It is means that positive samples are classified into positive and negative samples are classified into negative) to the total number of predictions. Sensitivity describes the proportion of positive cases identified with accounts for all positive cases, which is to judge model's ability of detecting positives accurately. Specificity denotes the proportion of negative cases identified accounts for all negative cases, which is to judge model's ability of detecting negatives accurately. Among them, sensitivity and specificity are two commonly judgment standards in the field of medical classification tasks. These metrics are defined in the following equations (12), (13), and (14):

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

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

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

TP (true positive) refers to the number of samples that are truly identified as positive samples, TN (true negative) refers to the number of samples that are truly identified as negative samples, FP (false positive) refers to the number of samples that are mistaken for positive samples, which actually is negative samples, and FN (false negative) refers to the number of samples that are mistaken for negative samples, which are actually positive samples. Because of the large differences in different categories, sensitivity and specificity are more relevant performance criteria in arrhythmia detection than accuracy.

In addition, the confusion matrix is leveraged to validate the performance of proposed model, which is an important standard to judge the performance of multi-classification model.

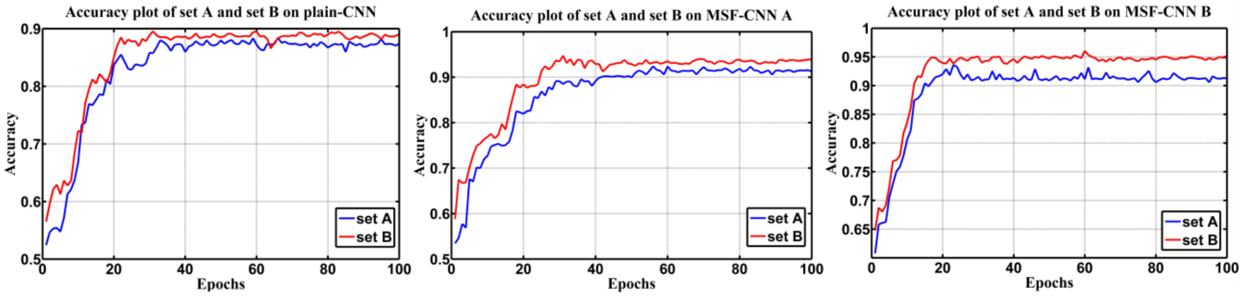
In the confusion matrix, the greater the number of true positive cases and true negative cases are, the better the model's performance is. Likewise, the fewer false positive examples and false negative examples, the better the overall performance of the model is.

### C. PERFORMANCE COMPARISON AND DISCUSSION

In this section, we implement six groups of ablation experiments to analyze the performance of model. First, we carry out a set of experiments to compare the effects of different lengths (set A and set B) of signals on our models' performances. Moreover, we show the change of performances by using the data augmentation method on training process. In addition, we conduct a set of experiments to demonstrate the function of denoising on the pre-processing of data. Meanwhile, we specially designed an experiment to verify the effect of the residual learning network. And the convergence analysis experiment is shown to validate our models' convergence ability in the fifth group experiment. Finally, the confusion matrix also is implemented to analyze each classification signals' performances. The detailed discussion about the six specific groups of experiments is as follows.

#### 1) SET A VS. SET B

We design a set of experiments to verify the effect of set A and set B on three models in the first phase. Every heartbeat includes 251 samples in set A and 361 samples in set B. Figure 5 presents the performances' trends of the two datasets on the three models. According to Figure 5, the changing curves of accuracy from the three models (plain-CNN, MSF-CNN A, MSF-CNN B) indicate that the accuracy of set B is slightly better than set A, mainly because each heartbeat from set B includes more samples than set A, and these models can learn more abundant features information. Otherwise, the overall average classification performances (accuracy, sensitivity, and specificity) for set A and set B in the three models are shown in Table 4. In set A, the average accuracies of the three networks are 83.15%, 86.40%, 89.17%, respectively. The result of MSF-CNN A is 3.25% higher than the performance of the plain-CNN in the set A. Additionally, the result of MSF-CNN B without residual learning is 2.77% higher than the performance of MSF-CNN A in set A. In set B, the performances of the three models also differ by 4.42% and 2.78%, respectively. Otherwise, sensitivity and specificity of 75.90% and 87.64% are also obtained in this experiment from set B. It is lower than the metrics from the plain-CNN network and MSF-CNN A in set B without residual learning. However, they are higher than the metrics from the three models in set A. It is analyzed that data imbalance may lead to this problem. In Table 2, the number of instances of each category without data augmentation is quite different. Overall, the results also suggest that the parallel group convolutional block in MSF-CNN A and B and the concatenation group convolution block in MSF-CNN B without residual learning have an important effect on the performance improvement of the proposed models. In theory, longer ECG records cover more heartbeat rhythm information, which will lead to better classification performance. Thus, in the following experiments, we use the data from set B to implement ablation experiment analysis.



(**a**) (**b**) (**c**)

**Figure 5. The accuracy plot of set A and set B. (a): the result of the plain-CNN. (b): the result of MSF-CNN A. (c): the result of MSF-CNN B.**

TABLE 4. THE AVERAGE CLASSIFICATION RESULTS FOR SET A AND SET B ON THE PROPOSED THREE MODELS

| network | Set A (251 samples) | | | Set B (361 samples) | | |
|---|---|---|---|---|---|---|
| | Acc. (%) | Se. (%) | Sp. (%) | Acc. (%) | Se. (%) | Sp. (%) |
| Plain-CNN | 83.15 | 65.14 | 85.08 | 85.23 | 87.41 | 79.50 |
| MSF-CNN A | 86.40 | 77.69 | 81.54 | 89.65 | 83.96 | 88.67 |
| MSF-CNN B (w/o residual learning) | 89.17 | 68.79 | 74.63 | **92.43** | **75.90** | **87.64** |

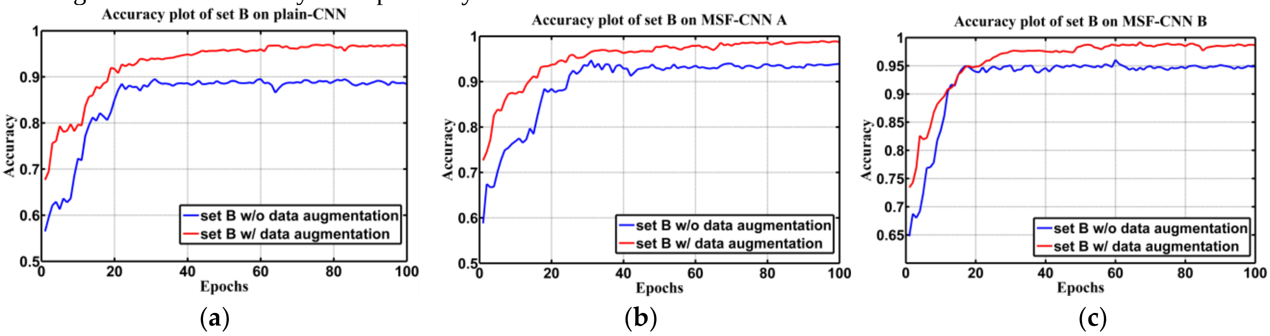
#### 2) DATA AUGMENTATION VS. WITHOUT DATA AUGMENTATION

In the second phase, we set up a set of experiments to analyze the impact of data augmentation on the model. The data used in this experiment are from set B. The strategy of data augmentation is implemented in accordance with the description of section III. B, and the total number of heartbeats increased to 331,055 after data augmentation (shown in Table 2). In Figure 6, we compare the performances of the proposed three networks architectures with data augmentation and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3016938, IEEE Access

without data augmentation in set B. As seen in Figure 6, the models with data augmentation perform dramatically better than these models' performances without data augmentation. Table 5 shows detailed evaluation metrics of the model predictions. The average accuracies of set B are 92.81%, 95.48%, and 95.96% with data augmentation on the three models. The results are 7.58%, 5.83%, and 3.53% higher than those of the three models without data augmentation.

Otherwise, due to data augmentation, the independent performance assessment of MSF-CNN B without residual learning results in sensitivity and specificity of 96.58% and 92.67%, respectively. It is better than the metrics from the plain-CNN and MSF-CNN A with data augmentation. Additionally, the performances are superior to the results of the three models without data augmentation. The experiment confirms that data augmentation dramatically improves the classification performance of ECG signals, which is also beneficial to data balancing in the dataset. Therefore, we adopt set B with data augmentation to perform the following experiments.



(a) (b) (c)

**Figure 6. The accuracy plot of set B. (a): the result of the plain -CNN. (b): the result of MSF-CNN A. (c): the result of MSF-CNN B (w/o residual learning).**

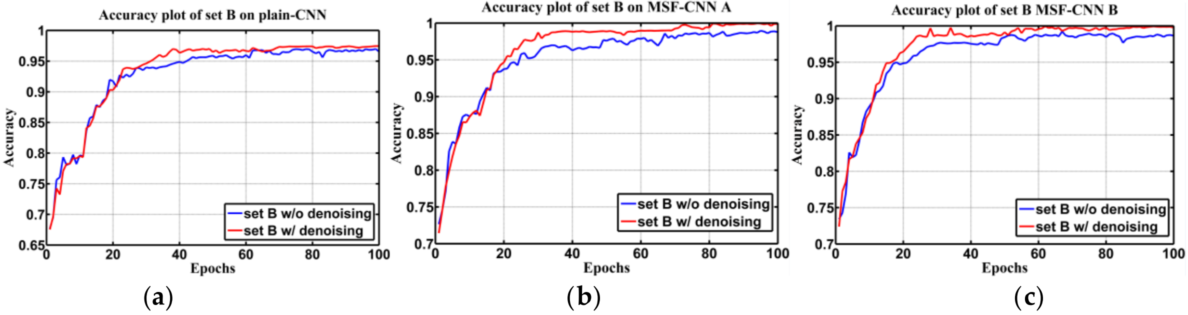
TABLE 5. THE AVERAGE CLASSIFICATION RESULTS FOR SET B WITH DATA AUGMENTATION ON THE PROPOSED THREE MODELS

| network | Set B (361 samples, without data augmentation) | | | Set B (361 samples, with data augmentation) | | |
|---|---|---|---|---|---|---|
| | Acc. (%) | Se. (%) | Sp. (%) | Acc. (%) | Se. (%) | Sp. (%) |
| Plain-CNN | 85.23 | 87.41 | 79.50 | 92.81 | 95.84 | 93.92 |
| MSF-CNN A | 89.65 | 83.96 | 88.67 | 95.48 | 96.53 | 87.74 |
| MSF-CNN B (w/o residual learning) | 92.43 | 75.90 | 87.64 | **95.96** | **96.58** | **92.67** |

### 3) DENOISING VS. WITHOUT DENOISING

In this experiment, we set up a set of experiments to analyze the impact of denoising on the model. The data used in this experiment are from set B with data augmentation. As shown in Figure 7, the performance of denoising performs slightly better than these models' performances without the processing of denoising. The detailed classification measures are reported in Table 6. The average accuracies of set B are 93.41%, 96.38%, and 97.03% with denoising on the three models without residual learning, respectively.

The results are 0.6%, 0.9%, and 1.07% higher than those of the three models without denoising. Moreover, compared with all the other models, very high sensitivity (94.43%) and specificity (96.41%) are obtained in this experiment. It is necessary to emphasize that the data augmentation strategy is



(a) (b) (c)

**Figure 7. The accuracy plot of set B. (a): the result of the plain-CNN. (b): the result of MSF-CNN A. (c): the result of MSF-CNN B (w/o residual learning).**

TABLE 6 THE AVERAGE CLASSIFICATION RESULTS FOR SET B WITH DATA AUGMENTATION AND DENOISING ON THE PROPOSED THREE MODELS

| network | Set B (361 samples, with augmentation, without denoising) | | | Set B (361 samples, with augmentation, with denoising) | | |
|---|---|---|---|---|---|---|
| | Acc. (%) | Se. (%) | Sp. (%) | Acc. (%) | Se. (%) | Sp. (%) |
| Plain-CNN | 92.81 | 95.84 | 93.92 | 93.41 | 87.16 | 89.73 |
| MSF-CNN A | 95.48 | 96.53 | 87.74 | 96.38 | 91.82 | 92.58 |
| MSF-CNN B (w/o residual learning) | 95.96 | 96.58 | 92.67 | **97.03** | **94.43** | **96.41** |
| **MSF-CNN B (w/residual learning)** | ——— | ——— | ——— | **98.00** | **96.17** | **96.38** |

implemented in this experiment. It is clear that the denoising technique has an influence on the performance of the models.

### 4) RESIDUAL LEARNING VS. WITHOUT RESIDUAL LEARNING

Next, we evaluate the effect of the residual learning block on MSF-CNN B with augmentation and denoising on set B. The baseline network is the same as the above MSF-CNN B without the residual learning block. The MSF-CNN B with residual learning adds a shortcut connection to each pair of 1 ×3 as in Figure 4 (c). We make two major observations from Table 6 (**the last row**) and Figure 8. First, the result situation (accuracy) is reversed with residual learning—the MSF-CNN B with residual learning is better than it without residual learning (differ by 0.97%). Most importantly, the performances of sensitivity and specificity also exhibit excellent and stable metrics. This indicates that the residual learning block dramatically enhances the optimization efficiency by providing faster convergence at the early stage.

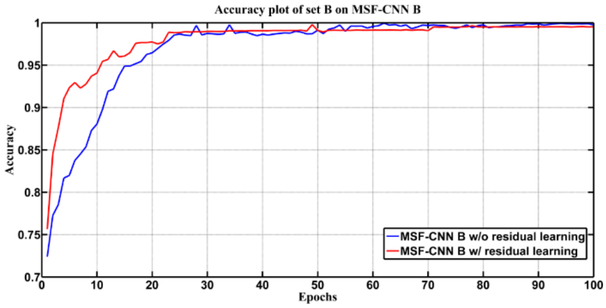


**Figure 8.** **The accuracy plot of set B with denoising. The blue solid denotes MSF-CNN B without residual learning, and the red solid denotes MSF-CNN B with residual learning.**

### 5) CONVERGENCE ANALYSIS

Then, we obtain the loss details during the training and validation processes. Figure 9 illustrates the change curve of loss of set B on MSF-CNN B without residual learning block, and Figure 10 also shows the result of set B on MSF-CNN B with residual learning block. As shown in the figures 9 and 10, the convergence effect of the model with residual learning is better than that of the model without residual learning. In addition, these experiments' results also show that the model converges after between 60 and 100 epochs during training and between 80 and 100 epochs during validation. Hence, 100 epochs are used in this experiment to ensure full convergence of the model and reduce overfitting. Moreover, the speed of convergence from the model with residual learning is faster.

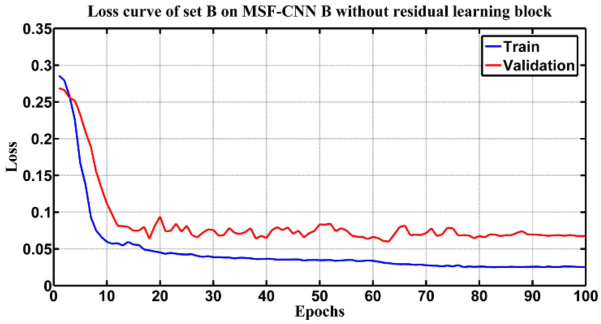


**Figure 9.** **Training and validation loss function of set B on MSF-CNN B without residual learning over the epochs.**

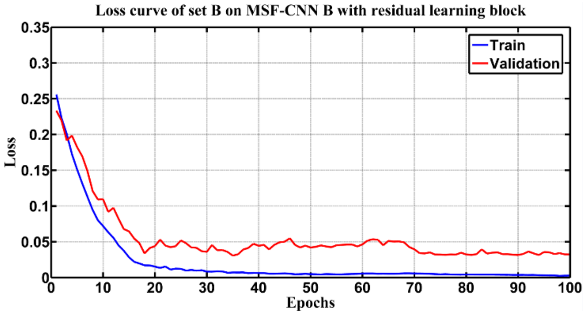


**Figure 10.** **Training and validation loss function of set B on MSF-CNN B with residual learning over the epochs.**

### 6) CONFUSION MATRIX ANALYSIS

Finally, in addition to evaluating each classification signal's performances of the model with residual learning block, we also assessed a confusion matrix of ECG heartbeats (Tables 7 and 8). They show the accuracy, sensitivity, and specificity of each classification. Table 8 shows a confusion matrix from the MSF-CNN B without a residual learning block. Table 9 describes a confusion matrix from the model with residual learning block. According to Table 8, on average less than 1.12% of the ECG heartbeats are wrongly classified across all 10-fold when the model does not utilize a residual learning block. Likewise, for the model with residual learning block, less than 1.00% of the ECG heartbeats are wrongly classified across all 10-folds. The minimal sensitivity recorded for both models are attributed to the detection of class F and are 92.25% and 92.32%, respectively. The minimal specificity for the model without residual learning block is attributed to the detection of class Q and is 95.33%. And the minimal specificity is 96.81%, which is a model with residual learning block attributed to the detection of class V. The results also demonstrate that the residual learning block has a positive impact on the performance of the model.

TABLE 7 A CONFUSION MATRIX OF ECG HEARTBEATS WITHOUT RESIDUAL LEARNING BLOCK ACROSS ALL 10-FOLDS

| Confusion Matrix | | Predicted | | | | | Acc (%) | Sen (%) | Spe (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | N | S | V | F | Q | | | |
| True | N | 87904 | 511 | 1036 | 926 | 218 | 98.49 | 97.03 | 98.05 |
| | S | 36 | 54579 | 427 | 105 | 473 | 99.98 | 98.13 | 97.24 |
| | V | 678 | 236 | 69815 | 1231 | 390 | 98.42 | 96.50 | 96.96 |
| | F | 803 | 176 | 924 | 29617 | 560 | 98.44 | 92.25 | 99.11 |
| | Q | 721 | 93 | 243 | 367 | 78986 | 99.05 | 98.23 | 95.33 |

Acc=accuracy, Sen=sensitivity, Spe=specificity.

TABLE 8 A CONFUSION MATRIX OF ECG HEARTBEATS WITH RESIDUAL LEARNING BLOCK ACROSS ALL 10-FOLDS

| Confusion Matrix | | Predicted | | | | | Acc (%) | Sen (%) | Spe (%) |
|---|---|---|---|---|---|---|---|---|---|
| | | N | S | V | F | Q | | | |
| True | N | 88837 | 267 | 48 | 526 | 917 | 99.46 | 98.06 | 97.68 |
| | S | 43 | 54930 | 26 | 497 | 124 | 97.52 | 98.76 | 99.68 |
| | V | 544 | 103 | 70903 | 267 | 533 | 99.41 | 98.00 | 96.81 |
| | F | 97 | 233 | 307 | 31438 | 5 | 99.32 | 92.32 | 99.46 |
| | Q | 86 | 279 | 108 | 299 | 79638 | 99.28 | 99.04 | 98.37 |

Acc=accuracy, Sen=sensitivity, Spe=specificity.

Recent advances and representative techniques in arrhythmias are summarized in Table 9, which also yield high-performance results. However, compared to recent advances, the benefits of our proposed MSF-CNN B are as follows:

(1) Compared with most literature, the evaluation metrics from our proposed model, including accuracy, sensitivity, specificity, and confusion matrix, is comprehensive and outperform the most of recent advances. And our proposed MSF-CNN is end-to-end based on deep learning, which replaces additional hand-crafted feature extraction using traditional machining learning.

(2) Even though the performance of our model is slightly lower than [61], our proposed model deals with multi-classification problems, rather than the two-classification problem studied in [61].

(3) We implemented the 10-fold cross-validation approach in the proposed models, thus boosting the robustness of the models.

Otherwise, compare with our work, even though the average accuracy from reference [59] is better than our model's performance result, the performance metrics (accuracy, sensitivity, and specificity) of our paper are more comprehensive than the metrics (only accuracy) of [59]. And the deep learning method of STFT-Based Spectrogram [59] also provide a new idea for future work. In additional, the CNN and RNN (Recurrent Neural Network) is two popular deep learning methods to process the time series data. In [62], though the performance is superior to our models' result, compared with the LSTM-based auto-encoder network in [62], our model is more lightweight and less computationally expensive. The LSTM is a replacement of the traditional RNN. And it is a bidirectional model, which is utilized to extract the bidirectional information from the forward model and backward model at the same time. There is no doubt that the advantage will also cost a lot of computational expensive. Most importantly, we think the LSTM-based auto-encoder (AE) network [62] is a positive strategy, which can effectively extract the characteristic information of time series signals. We will fully consider the optimization methods of [62] in our future work.

## VI. CONCLUSION AND FUTURE WORK

In this study, three end-to-end network models, including a plain-CNN and two MSF-CNN architectures (A and B), are presented to automatically identify and classify the five different types of ECG heartbeats. The plain-CNN is a baseline network with multiple convolution layers, which is a simple CNN architecture to verify the processing ability of 1-D CNN for ECG signals. The MSF-CNN A is proposed to improve the learning ability of the plain-CNN. It is an upgraded network based on baseline network to verify the processing ability of three parallel convolution kernels for ECG signals, which increases a parallel group convolution block (including three different convolution kernels with 1×7,1×5, and 1×3). Finally, the MSF-CNN B based on the MSF-CNN A is improved by implementing a residual learning block with three concatenation groups convolution blocks to promote the performance of the model. It is an upgraded network based on the MFS-CNN A to verify processing ability of the concatenation group convolution blocks and residual learning blocks for ECG signals.

The three proposed models are trained and tested with a public MIT-BIH arrhythmia database on five types of signals, N, S, V, F, and Q. Six groups of ablation experiments are also conducted to analyze the performances of these models. The best model MSF-CNN B with residual learning and group convolution blocks (including the parallel and concatenation group convolution blocks) achieves an average accuracy, sensitivity, and specificity of 98.00%, 96.17%, and 96.38% in set B. Otherwise, the strategy of multi-scale data, data augmentation, and denoising also have an important effect on the training of the three models in our experiments.

Therefore, our proposed deep neural network algorithm (MSF-CNN B) shows the potential of deep learning-based approach for feature extraction of the MIT-BIH arrhythmia database. As is evident from these results, the proposed approach is an efficient automatic cardiac arrhythmia classification method and provided a reliable recognition system based on well-established CNN architectures instead of training a deep CNN from scratch. It has the potential to provide accurate ECG signal classification in clinical practice.

In future work, we would like to introduce more clinical diagnosis data to test the proposed model. Additionally, the temporal (heartbeats) and spatial (spectrogram) signal features will be combined to improve the performance metrics of the models in future work. We would also like to determine the severity grades of patients with chronic heart diseases by the detection and classification of ECG signals, which may represent normal, abnormal, and cardiac electrical activity conditions that may be life-threatening.

Specifically, compared with the self-organizing structural size method [63-65], the deep convolutional neural network is complicated to fast determine its optimal structure given specific applications. Hence, we will propose a new method combined the self-organizing maps and convolutional neural network to the ECG signal research in the future work.

Moreover, we will try our best to propose a new method combined the optimization approaches [66-68] and convolutional neural network to the ECG signal research in the future. This new method will focus on the following aspects:

(1) The real-world constraints must be considered in the new model. We will put theory research results into a specific filed or for a specific product.

(2) It’s considerable to design an adaptive parameter system to improve the robustness of optimization model.

(3) We will consider the imbalanced data classification problem and sufficient prior knowledge. The dendritic neuron model [69] and evolutionary cost-sensitive [70] will provide a new idea in future work.

TABLE 9 A SUMMARY OF SELECTED WORKS FOR AUTOMATIC ARRHYTHMIA CLASSIFICATION OF ECG SIGNALS FROM THE DATABASE OF MIT-BIH ARRHYTHMIA

| Literature and Time | Main Work | Database | Approach | Performance (%) | | |
|---|---|---|---|---|---|---|
| | | | | Accuracy (Correct recognition rate) | Sensitivity (Recall) | Specificity (positive predictivity) |
| 2017 [51] | Five classification (N, S, V, F, Q) | MIT-BIH arrhythmia database | ML and DL: MFSWT; DNN | 97.50 | / | / |
| 2018 [52] | Two classification (S, V) | MIT-BIH arrhythmia database | ML: PSO optimized least-square twin SVM | 89.90 | 80.80(S), 82.20(V) | 96.70(S), 99.00(V) |
| 2019 [53] | Ten classification | Chinese Cardiovascular Disease Database | ML: PPNN | 74.16 | 75.23 | 73.92 |
| 2019 [54] | The ventricular ectopic beat detection | MIT-BIH arrhythmia database | DL: 1D-CNN | 95.50 | 85.80 | 64.50 |
| 2019 [55] | Five classification (N, S, V, F, Q) | MIT-BIH arrhythmia database | DL: DRNNs based on BGRU | 98.40 | / | / |
| 2019 [56] | Seven classification | Chinese Cardiovascular Disease Database | DL: Parallel GRU RNN | 95.98 | / | / |
| 2019 [57] | Six classification (Normal, L, R, V, A, P) | MIT-BIH arrhythmia database | ML: KNN | 97.70 | / | / |
| 2019 [58] | Two-classification | MIT-BIH arrhythmia database | DL: CNN | 94.70 | 77.30(S), 93.70(V) | 97.70(S); 98.80(V) |
| 2019 [59] | Five classification | MIT-BIH arrhythmia database | DL: CNN of STFT-Based Spectrogram | 99.00 | / | / |
| 2020 [60] | Multi-classification | Chinese Cardiovascular Disease Database | DL: MTGBi-LSTM | 88.86 | 94.19 | / |
| 2020 [61] | Two-classification | Personal Wearable Devices | DL: LSTM-RNN | 99.20(VEB)<br>98.30(SVEB) | 93.00(VEB)<br>66.90(SVEB) | 99.80(VEB)<br>99.80(SVEB) |
| 2020 [62] | Five classification | MIT-BIH arrhythmia database | ML and DL: LSTM, SVM | 99.45 | 98.63 | 99.66 |
| **This paper** | **Five classification (N, S, V, F, Q)** | MIT-BIH arrhythmia database | **DL: Plain-CNN**<br>**MSF-CNN A**<br>**MSF-CNN B** | **93.41(Plain-CNN)**<br>**96.38(MSF-CNN A)**<br>**98.00(MSF-CNN B)** | **87.61(Plain-CNN)**<br>**91.82(MSF-CNN A)**<br>**96.17(MSF-CNN B)** | **89.73(Plain-CNN)**<br>**92.58(MSF-CNN A)**<br>**96.38(MSF-CNN B)** |

**Abbreviations:**

**Heartbeat types:** S: Supraventricular ectopic beat; V: Ventricular ectopic beat; F: Fusion beat; Q: Unknown beat; N: any heartbeat not in the S, V, F, Q classes or normal beat; PVC: Premature ventricular contraction beat; PAC: Premature atrial contraction beat; L: Left bundled branch blocks; R: Right bundled branch blocks; V: Premature ventricular contractions; A: Atrial premature beats; P: Paced beats; VEB: Ventricular ectopic beats; SVEB: Supraventricular ectopic beats.

**Approaches:** ML: Machine learning, DL: Deep learning, SVM: Support vector machine; DNN: Deep neural network, CNN: Convolutional neural network; MFSWT: Slice wavelet transform; PSO: Particle swarm optimization; PPNN: Probabilistic process neural network; DRNNs: Deep recurrent neural networks; BGRU: Bidirectional gated recurrent unit; KNN: k-Nearest Neighbor; MTG: Multi-Task Group.

## REFERENCES

[1] National Heart Lung and Blood Institute, Types of Arrhythmias, 2011 [Online]. Available: https://www.nhlbi.nih.gov/health/health-topics/topics/arr/types. (Accessed 5 July 2017).

[2] U.R. Acharya, J.S. Suri, J.A.E. Spaan, S.M. Krishnan, Advances in Cardiac Signal Processing, 2007.

[3] Arrhythmia irregular heartbeat center, Heart Disease and Abnormal Heart Rhythm (Arrhythmia), 2017. [Online]. Available: https://www.medicinenet.com/ arrhythmia_irregular_heartbeat/article.htm.

[4] S.M. Mathews, K. Chandra, K.E. Barner, "A novel application of deep learning for single-lead ECG classification," Computers in Biology and Medicine, vol. 99, pp. 53–62, Jun. 2018.

[5] S. Preejith, R. Dhinesh, J. Joseph, and M. Sivaprakasam, "Wearable ECG platform for continuous cardiac monitoring," in Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE, pp. 623–626, Oct. 2016.

[6] B. Murugesan, V. Ravichandran, and K. Ram, "ECGNet: Deep Network for Arrhythmia Classification," IEEE Instrumentation and Measurement Society, pp. 623–626, Jun. 2018.

[7] American National Standards Institute, Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms, 2012.

[8] R.J. Martis, U.R. Acharya, H. Adeli, "Current methods in electrocardiogram characterization," Comput. Biol. Med, vol. 48, no.1, pp. 133-149, May. 2014.

[9] U.R. Acharya, S.L. Oh, Y. Hagiwara, "A Deep Convolutional Neural Network Model to Classify Heartbeats," Computers in Biology and Medicine, vol. 89, no.1, pp. 389-396, Oct. 2017.

[10] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, "Deep Residual Learning for Image Recognition," In CVPR, 2017.

[11] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. Presented at CVPR,2015. [Online]. Available: arxiv.org/abs/1409.1556.

[12] Gao Huang, Zhuang Liu, Kilian Q. Weinberger. Densely Connected Convolutional Networks. Presented at CVPR,2016. [Online]. Available: arxiv.org/abs/1608.06993.

[13] G. Cai, Y. Wang, L. He, and M. Zhou, "Unsupervised Domain Adaptation with Adversarial Residual Transform Networks," IEEE Transactions on Neural Networks and Learning Systems, DOI: TNNLS.2019.2935384, online 2019.

[14] T. D. Pham, K. Wardell, "A. Eklund and G. Salerud, "Classification of short time series in early Parkinson's disease with deep learning of fuzzy recurrence plots," IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 6, pp. 1306-1317, November 2019.

[15] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Presented at CVPR,2015. [Online]. Available: arXiv:1506.01497.

[16] J. Dai, Y. Li, K. He, and J. Sun. R-FCN: Object Detection via Region-based Fully Convolutional Networks. Presented at NIPS, 2016. [Online]. Available: arXiv:1605.06409.

[17] Y. Tian, X. Li, K. Wang and F. Wang, "Training and testing object detectors with virtual images," IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 2, pp. 539-546, Mar. 2018.

[18] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y. Deformable Convolutional Networks. Presented at CVPR, 2017. [Online]. Available: arXiv:1703.06211.

[19] Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. Presented at CVPR, 2017. [Online]. Available: arXiv:1706. 05587.

[20] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia. Pyramid Scene Parsing Network. Presented at CVPR, 2017. [Online]. Available: arXiv:1612.01105,2017.

[21] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask R-CNN. Presented at ICCV, 2017. [Online]. Available: arXiv:1703.06870.

[22] S.K. Berkaya, A. K. Uysal, E.S. Gunal, "A survey on ECG analysis," Biomedical Signal Processing and Control, vol. 43, pp. 216-235, May. 2018.

[23] M.A. Awal, S.S. Mostafa, M. Ahmad, M.A. Rashid, "An adaptive level dependent wavelet thresholding for ECG denoising," Biocybern Biomed. Eng, vol. 34, no.4, pp. 238-249, Mar. 2014.

[24] L.Q. Wang, "Research on ECG Waveform Detection and Arrhythmia Classification[D]," Hebei University of Technology, Tianjin, 2014.

[25] N. Sen, C. Chandrakar, "Development of a Novel ECG signal Denoising System Using Extended Kalman Filter," IJAREEIE. vol. 3, pp. 6896-6901, 2014.

[26] B.S. Gayal, F.I. Shaikh, "Denoising of ECG signal using undecimated wavelet transform," IJAREEIE. vol. 3, pp. 7200-7208, 2014.

[27] R. Rodrigues, P. Couto, "A Neural Network Approach to ECG Denoising," Available online: arXiv:1212.5217, 2012.

[28] Md. A. Kabir, C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," Biomedical Signal Processing and Control, vol. 7, pp. 481---489, 2012.

[29] Y.C. Yeh, W.J. Wang, C.W. Chiou, "Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals," Measurement, vol. 42, no.5, pp. 778-789, Jun. 2009.

[30] Y.C. Yeh, W.J. Wang, 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.

[31] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, S. Luo, "ECG beat detection using filter banks," IEEE Trans. Biomed. Eng., vol. 46, pp. 192–202, 1999.

[32] B. Abibullaev, H.D. Seo, "A new QRS detection method using wavelets and artificial neural networks," J. Med. Syst. vol. 35, pp. 683–691, 2011.

[33] M. Korurek, B. Dogan, "ECG beat classification using particle swarm optimization and radial basis function neural network," Expert Syst. Appl., vol. 37, pp. 7563–7569, 2010.

[34] A. Martínez, R. Alcaraz, J.J.Rieta, "Application of the phasor transform for automatic delineation of single-lead ECG fiducial points," Physiol. Meas., vol. 31, pp. 1467–1485, 2010.

[35] Y. Kutlu, D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," Comput. Method Program Biomed., vol.105, no. 3, pp. 257–267, 2012.

[36] E. J. da S. Luz, W. R. Schwartz, G. C. Chavez, et al., "ECG-based heartbeat classification for arrhythmia detection: A survey," Computer Methods and Programs in Biomedicine, vol. 127, pp: 144-164, 2015.

[37] A. M. Alqudah, "An enhanced method for real-time modelling of cardiac related biosignals using Gaussian mixtures," Journal of medical engineering & technology, vol. 41, no. 8, pp. 600-611, Oct. 2017.

[38] T.Y. Li, and M. Zhou, "ECG Classification Using Wavelet Packet Entropy and Random Forests," Entropy, vol. 18, no. 8, pp. 285-300, Aug. 2016.

[39] A. M. Alqudah, I. Abuqasmieh, A. Badarneh and H. Alquran, "Developing of robust and high accurate ECG beat classification by combining Gaussian mixtures and wavelets features," Australasian physical & engineering sciences in medicine, vol. 42, no. 1, pp. 149-157, Jan. 2019.

[40] M. Hammad, A. Maher, K. Q. Wang, F. Jiang, and M. Amrani, "Detection of Abnormal Heart Conditions Based on Characteristics of ECG Signals," Measurement, vol. 125, pp. 634-644, Sep.2018.

[41] S. Kiranyaz, T. Ince, M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[42] T. J. Jun, H. J. Park, Y. H. Kim, "Premature ventricular contraction beat detection with deep neural networks," in 15th IEEE International Conference on Machine Learning and Applications, pp. 859–864, 2016.

[43] U. R. Acharya, S. L. Oh, Y. Hagiwara, et al., "A Deep Convolutional Neural Network Model to Classify Heartbeats," Computers in Biology and Medicine, vol. 89, pp 389-396, 2017.

[44] H. Alquran, A. M. Alqudah, I. Abu-Qasmieh, Al-Badarneh, S. Almashaqbeh, "ECG classification using higher order spectral estimation and deep learning techniques," Neural Network World, vol. 29, no. 4, pp: 207-219, Aug. 2019.

[45] A.L. Goldberger, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," Circulation 101 (23) (2000) e215–e220.

[46] J. Schmidhuber, "Deep Learning in neural networks: an overview," Neural Netw, vol. 61, pp. 85-117, Jan.2015.

[47] A. Krizhevsky, I. Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network", NIPS Curran Associates Inc,2012.

[48] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzman machines," Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814, 2010.

[49] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," In CVPR, 2015.

[50] K. Simonyan, A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In ICLR, 2015.

[51] K. Luo, J. Li, Z. Wang, A. Cuschieri, "Patient-specific deep architectural model for ECG classification," J. Healthcare Eng., vol. 2017, May 2017.

[52] S. Raj, K. C. Ray, "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias," Expert Systems with Applications, vol. 105, pp. 49–64, Sep. 2018.

[53] N. D. Feng, S. H. Xu, Y. Q. Liang, K. Liu, "A Probabilistic Process Neural Network and Its Application in ECG Classification," IEEE Access, vol. 7, pp. 50431 – 50439, Apr. 2019.

[54] A. A. S. León, J. R. N. Alvarez, "1D Convolutional Neural Network for Detecting Ventricular Heartbeats," IEEE Latin America Transactions, vol. 17, no. 12, pp. 1970 – 1977, Dec. 2019.

[55] H. M. Lynn, S. B. Pan, P. Kim, "A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks," IEEE Access, vol. 7, pp. 145395-145405, Sep. 2019.

[56] S. H. Xu, J. J. Li, K. Liu, L. Wu, "A Parallel GRU Recurrent Network Model and Its Application to Multi-Channel Time-Varying Signal Classification," IEEE Access, vol. 7, pp. 118739 - 118748, Sep. 2019.

[57] H. Yang, Z. Q. Wei, "Arrhythmia Recognition and Classification Using Combined Parametric and Visual Pattern Features of ECG Morphology," IEEE Access, vol. 8, pp. 47103 - 47117, Mar. 2019.

[58] S. S. S. Xu, M. W. Mak, C. C. Cheung, "Towards End-to-End ECG Classification with Raw Signal Extraction and Deep Neural Networks," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1574 - 1584, Jul. 2019.

[59] J. Huang, B. Chen, B. Yao and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," IEEE Access, vol. 7, pp. 92871-92880, 2019.

[60] Q. J. Lv, H. Y. Chen, W. B. Zhong, Y. Y. Wang, J. Y. Song, S. D. Guo, L. X. Qi, C. Y.C. Chen, "A Multi-Task Group Bi-LSTM Networks

Application on Electrocardiogram Classification," IEEE Journal of Translational Engineering in Health and Medicine, vol. 8, pp. 1900111-1900121, Feb. 2020.

[61] S. Saadatnejad, M. Oveisi, M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 515 – 523, Feb. 2020.

[62] B. Hou, J. Yang, P. Wang and R. Yan, "LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1232-1240, April 2020.

[63] G. M. Wang, J. F. Qiao, J. Bi, W. J. Li, M. C. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," IEEE Transactions on Automation Science and Engineering, vol. 16, no.2, pp. 874-885, 2019.

[64] W. A. Khan, S. H. Chung, H. L. Ma, et al., "A novel self-organizing constructive neural network for estimating aircraft trip fuel consumption," Transportation Research Part E: Logistics and Transportation Review, vol.132, pp. 72-96, 2019.

[65] E. J. Palomo, E. López-Rubio, "The growing hierarchical neural gas self-organizing neural network," IEEE transactions on neural networks and learning systems, vol. 28, no. 9, pp. 2000-2009, 2016.

[66] Y. Yu, S. Gao, Y. Wang, and Y. Todo, "Global optimum-based search differential evolution," IEEE/CAA J. Autom. Sinica, vol. 6, no. 2, pp. 379-394, Mar. 2019.

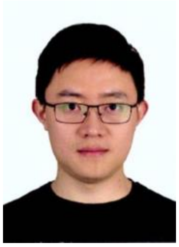
[67] Q. Kang, X. Song, M. Zhou, and L. Li, "A Collaborative Resource Allocation Strategy for Decomposition-Based Multiobjective Evolutionary Algorithms," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol.49, no. 12, pp. 2416-2423, Dec. 2019.

[68] K. Z. Gao, Z. G. Cao, L. Zhang, Z. H. Chen, Y. Y. Han, and Q. K. Pan, "A review on swarm intelligence and evolutionary algorithms for solving flexible job shop scheduling problems," IEEE/CAA J. Autom. Sinica, vol. 6, no. 4, pp. 875-887, July 2019.

[69] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation and prediction," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 2, pp. 601 - 614, Feb. 2019.

[70] G. S. Hong, "A Cost-Sensitive Deep Belief Network for Imbalanced Classification," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 1, pp. 109-122, Jan. 2019.

**HAO DANG** received the M.S. degree in pattern recognition and intelligent system from the Henan University of Technology, Zhengzhou, China, in 2016. He is currently pursuing the Ph.D. Degree in control science and engineering with the School of Automation, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include the pattern recognition, intelligent systems, machine learning, and so on.



**XINGXIANG TAO** received the M.S. degree in logistics engineering from the School of Information, Beijing Wuzi University, Beijing, China, in 2017. He is currently pursuing the Ph.D. Degree in control science and engineering with the School of Automation, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include the machine learning, the pattern recognition and intelligent systems.

**YARU YUE** received the M.S. degree in Detection Technology and Automatic Equipment from the Beijing Information Science and Technology University, Beijing, China, in 2018, where he is currently pursuing the Ph.D. degree in pattern recongnition and intelligent system from School of Autumation, Beijing University of Posts and Teleconmmunications, Beijing, China. His research interests include the machine learning, computer vision, neural networks, and so on.

**DANQUN XIONG** received the B.S. degree from the Department of Clinical Medicine, Nanchang University, Nanchang, China, in 2013, and the M.S degree in internal medicine from medicine school of Tongji University, Shanghai, China, in 2016. He is currently an attending doctor in Department of Cardiology of Jiading District Central Hospital Affiliated Shanghai University of Medical and Health Sciences. His research focus on the detection and diagnose of arrhythmia.

**XIANGDONG XU** received the B.S. degree from the Department of Clinical Medicine, Nanchang University,Nanchang, China, in 1997, He is currently an Professor, and a Master Supervisor and director of in Department of Cardiology of Jiading District Central Hospital Affiliated Shanghai University of Medical and Health Sciences. He has published more than 20 articles. His research focus on the usage and challenge of innovative technology in General Practice Medicine, such as Machining Learning, Sequencing, and Big-Data.

**XIAOGUANG ZHOU** received the M.S. degree from the Department of Precision Instrument, Tsinghua University, in 1984, and the Ph.D. degree in engineering from the Tokyo University of Agriculture and Technology, Japan. He was a Visitor Professor with the Tokyo University of Agriculture and Technology from 2001 to 2002, and a JSPS Researcher with Tokyo University from 2013 to 2014. He is currently a Professor, and a Doctoral Supervisor with the School of Automation, Beijing University of Posts and Telecommunications. He also serves as the Director of the Engineering Research Center of Information Networks, Ministry of Education. He is the author of over 10 books, over 100 articles, and over 16 inventions. His research interests include control theory and its application in engineering, deep learning, computer vision, Internet of Things and automated logistics system, and mechatronics technology. He is a permanent member of the Chinese Association of Automation/Manufacturing Technology Committee and the China Institute of Communications/Equipment manufacturing technical Committee.