Received April 30, 2020, accepted May 20, 2020, date of publication June 10, 2020, date of current version June 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001284

Multi-ECGNet for ECG Arrythmia Multi-Label Classification

JUNXIAN CAI[1], WEIWEI SUN[2], JIANFENG GUAN[1], (Member, IEEE), AND ILSUN YOU[3], (Senior Member, IEEE)

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Nephrology, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China
[3]Department of Information Security Engineering, Soonchunhyang University, Asan 31538, South Korea

Corresponding author: Ilsun You (ilsunu@gmail.com)

This work was supported in part by the National Key Research and Development Program under Grant 2018YFE0206800, and in part by the Soonchunhyang University Research Fund.

ABSTRACT With the development of various deep learning algorithms, the importance and potential of AI + medical treatment are increasingly prominent. Electrocardiogram (ECG) as a common auxiliary diagnostic index of heart diseases, has been widely applied in the pre-screening and physical examination of heart diseases due to its low price and non-invasive characteristics. Currently, the multi-lead ECG equipments have been used in the clinic, and some of them have the automatic analysis and diagnosis functions. However, the automatic analysis is not accurate enough for the discrimination of abnormal events of ECG, which needs to be further checked by doctors. We therefore develop a deep-learning-based approach for multi-label classification of ECG named Multi-ECGNet, which can effectively identify patients with multiple heart diseases at the same time. The experimental results show that the performance of our methods can get a high score of 0.863 (micro-F1-score) in classifying 55 kinds of arrythmias, which is beyond the level of ordinary human experts.

INDEX TERMS ECG, arrythmia, multi-label classification, depthwise separable convolution, SE module.

I. INTRODUCTION

Electrocardiography is the process of producing an Electrocardiogram (ECG) which is a graph of voltage versus time of the electrical activity of the heart [1] using electrodes placed on the skin. ECG is the simplest and most efficient way to diagnose heart-related diseases and has been used in medical practice since 1903 [2]. After more than 100 years of development, ECG has played an important role in medicine, and many related advanced and easy-to-use clinical devices have been developed. However, medical imaging is still a field that requires a lot of manpower (expert knowledge and experience) to identify. How to use AI especial deep learning methods to carry out efficient and accurate medical diagnosis and free the manpower input of doctors is still a topic worthy of in-depth study.

Generally speaking, the method of measuring the ECG is to place the electrodes in different parts of the human body and connect them to the positive and negative electrodes of the electrocardiograph by wires. This method of electrical connection for recording ECG is called lead ECG. Different electrode positions and connection methods can form different leads. In the long-term clinical practice of electrocardiography, an international general Lead System established by Einthoven and now widely adopted has been formed, which is called the standard 12-Lead System [3]. The 10 electrodes in a 12-lead ECG are listed in Table1 [4].

For 12 leads, we have the following formulas below:

$$
\begin{aligned}
I &= LA - RA \\
II &= LL - RA \\
III &= II - I \\
aVR &= -(I + II)/2 \\
aVL &= I - II/2 \\
aVF &= II - I/2
\end{aligned} \tag{1}
$$

As a specialized medical field, the study of ECG requires a great deal of medical knowledge. If we use traditional machine learning methods for research, we have to use this professional medical knowledge to do a lot of feature

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

TABLE 1. Electrode placements.

Electrode name	Electrode placement
RA	On the right arm, avoiding thick muscle.
LA	In the same location where RA was placed, but on the left arm.
RL	On the right leg, lower end of inner aspect of calf muscle. (Avoid bony prominences)
LL	In the same location where RL was placed, but on the left leg.
V1	In the fourth intercostal space (between ribs 4 and 5) just to the right of the sternum (breastbone)
V2	In the fourth intercostal space (between ribs 4 and 5) just to the left of the sternum.
V3	Between leads V2 and V4.
V4	In the fifth intercostal space (between ribs 5 and 6) in the mid-clavicular line.
V5	Horizontally even with V4, in the left anterior axillary line.
V6	Horizontally even with V4 and V5 in the mid-axillary line.

extraction and feature engineering, which will greatly reduce the efficiency. As a most cutting-edge technology of machine learning, deep learning, especially CNN, can automatically extract important features through the superposition of multiple convolution layers, nonlinear variation layers, pooling layers, fully connected layers and normalized layers. Deep learning has made a great progress in image recognition, speech recognition, natural language processing and other fields, and we have reason to believe that a well-designed deep learning model can also achieve satisfactory results in ECG recognition.

Many current classification methods of ECG signals rely on the extraction of handmade features from ECG. This is done by using traditional feature extraction algorithms or drawing on human expertise. Then the extracted features are input into the generation or discrimination model to predict or classify the ECG. The extracted features are then classified by Support Vector Machine (SVM), decision tree models or other algorithms. Generally speaking, the machine learning methods are relatively fast in training and prediction, but the quality of feature extraction and feature engineering are difficult to quantify and prone to error, which will cause unstable factors to the final detection results.

In order to realize efficient, high-precision and highly automated end-to-end ECG detection, this paper proposes a Multi-ECGNet model based on one-dimensional convolution Resnet, depthwise separable convolution, and Squeeze-and-Excitation (SE) Module, which can be used for multi-label classification detection of ECG data. The datasets used in this paper contains 55 arrythmias which are symptoms related to heart diseases. For multi-label classification problems, we usually use micro-F1-score rather than accuracy as the evaluation criterion of the algorithm.

In general, the main contributions of this paper are as follows:

1) First, we put forward a whole set of analysis, modeling methods and research ideas of ECG detection with end-to-end deep learning model. It's proved that this method is superior to the common cardiologists in indicators through experiments.
2) Second, we propose an algorithm model named Multi-ECGNet integrating the advantages of ResNet, Squeeze-and-Excitation Module and Depthwise Separable Convolution, which is applicable to ECG detection.
3) Third, our method is a multi-label classification model, which can simultaneously detect 55 symptoms of heart disease, rather than most other relevant studies can only identify a single symptom.
4) Fourth, there are some wrong data and noise in the dataset, and the sample quantity distribution of 55 categories is seriously unbalanced. Therefore, based on focal loss, this paper also proposes corresponding solutions and improvement strategies of loss function.

The rest of this paper is organized as follows. Section 2 introduces the progress and achievements of ECG detection in recent years. Section 3 presents the proposed method and algorithm of the classification model. Section 4 introduces the specific implementation process. Section 5 presents our experimental process and results, and finally Section 6 concludes this paper and discuss the future work.

II. RELATED WORK

In recent years, researches on the combination of medical treatment and AI have gradually come to people's vision. Many efforts from large companies, universities and research institutes are focusing on this topic due to its potential scientific, social and economic values.

Before deep learning was fully developed, some scholars and researchers used traditional machine learning methods to classify ECG tests. In 2009, Kim *et al.* proposed a novel arrhythmia classification algorithm [5] which has a fast learning speed and high accuracy, and uses Morphology Filtering, Principal Component Analysis (PCA) and Extreme Learning Machine (ELM). The proposed algorithm can classify six beat types: normal beat, left bundle branch block, right bundle branch block, premature ventricular contraction, atrial premature beat, and paced beat. At that time, the algorithm was greatly faster than the algorithm using (Back Propagation Neural Network) BPNN, (Radial Basis Function Network) RBFN and SVM and got better experimental results than others.

Due to the limitations of machine learning methods, more researches in recent years have focused on the methods of deep learning, such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). Yildirim [6] proposed a new model called DBLSTM-WS based on wavelet sequence of deep bidirectional LSTM network for classification of ECG signals. The experimental results of five different types of heartbeat obtained from the (Massachusetts Institute of Technology, Boston's Beth Israel Hospital) MIT-BIH Arrhythmia Database showed that the recognition performance of DBLSTM-WS model is as high as 99.39

Baloglu *et al.* [7] propose a deep learning model based on CNN with end-to-end structure on standard 12-lead ECG signals for the diagnosis of myocardial infarction, with accuracy and sensitivity of all leading ECG signals exceeding 99.00%. Based on 1000 ECG signal fragments from MIT-BIH Arrhythmia Database, Yildirim *et al.* [8] designed a new one-dimensional CNN model involving 17 ECG categories (normal sinus rhythm, 15 types of arrhythmia, and pacemaker rhythm) with an overall accuracy of 91.33%.

Mathews *et al.* were with the other train of thought, and they used a single lead ECG signal data, with the Restricted Boltzmann Machine (RBM) and Deep Belief Network (DBN) [9], to detected ventricular and room on the rhythm of the heart, and DBN can obtain higher average ventricular ectopic beats on recognition rate (93.63%) and supraventricular ectopic beats recognition rate (95.57%) with the low sampling rate of 114 Hz.

These methods, whether based on machine learning or LSTM, CNN, RBM, DBM, have achieved good results on the MIT-BIH dataset, but all of them have their own limitations and tend to focus on one or a few diseases. And they didn't do any in-depth research on multi-label classification, and we know that patients with heart disease tend to have multiple heart symptoms at the same time, and these manifestations or symptoms are not independent.

In fact, there are many kinds of arrythmia symptoms. The research of this paper is based on 55 kinds of arrythmias. After proper training and adjustment, general deep learning methods can solve multi-classification problems with fewer categories. The solution space of a multi-classification problem such as 10 classifications is 10, but in the case of 55 multi-label classifications, the solution space of the problem will reach 255, which is undoubtedly a difficult challenge. At the same time, the correlation between the 55 labels, the data distribution, and the accuracy of the data labels will bring greater problems, which are the targets of this study. We propose a Multi-ECGNet network architecture by incorporating ResNet, depthwise separable convolution, SE Module design and weighted binary crossentropy improved based on focal loss as a multi-label classification loss. At the same time, Multi-ECGNet network model draws on the idea of label smoothing to deal with possible labeling errors. The following sections of this paper will detail the specific content and implementation process of these methods.

III. METHOD TO ELECTROCARDIOGRAMS MULTI-LABEL CLASSIFICATION

A. DATASETS AND DATA ANALYSIS

The dataset used in this paper is from the 2019 Tianchi Hefei High-Tech Cup ECG Human-Machine Intelligence Competition. The dataset contains 32,142 cases, of which 24,106 are used as training sets and 8,036 as test sets. Each sample has 8 leads, namely I, II, V1, V2, V3, V4, V5 and V6. The data of the remaining 4 leads can be derived according to the formulas, so we can get a total of 12 leads. Each sample has a sampling frequency of 500 Hz, a length of 10 seconds and a unit voltage of 4.88 microvolts. So each of our sample cases is a 5,000-row, 12-column matrix.

In this task, we need to perform multi-label classification on 55 arrythmias, and its solution is a one-hot matrix of length 55. In the evaluation criterion of multi-label classification, we use micro-F1 instead of precision or recall, and its calculation method refers to (2).

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{2}$$

In order to better solve and analyze the problem, we first need to analyze and visualize the data to help understand.

The age distribution of the data set samples is specific from children to the elderly, as shown in Fig. 1. The ages of some samples are missing, which does not affect the age distribution diversity of the dataset.

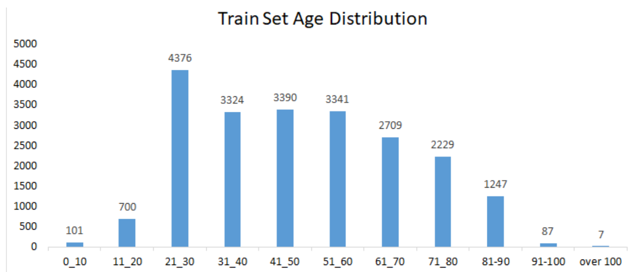

FIGURE 1. Train set age distribution.

The dataset contains 55 different categories of arrhythmias in all, each of which corresponds to one or more arrhythmias and different ECG waveform. The distribution of these 55 categories of arrhythmias was seriously unbalanced, in which the largest category of arrhythmias has 16,918 samples and the smallest one has 10 samples. The distribution of these arrhythmias is shown in Fig. 2. We can clearly get that the number of ECG of the first 10 categories of arrhythmias accounts for nearly 80% of the total dataset, and uneven data distribution will pose a higher challenge to the ability of algorithm and model, which is also one of the problems to be solved in this paper.

Each sample data contains 12 leads with 500 Hz sampling frequency during 10 seconds. Through visualization as shown in the Fig. 3, we can get that the waveforms of 12 leads present the characteristics respectively, which contain sinus rhythm, sinus arrhythmia, QRS low voltage, long QT, abnormal ECG example. For non-medical experts, it difficult to figure out the meanings of waveforms and their differences among different waveforms. Even for medical experts, according to the study published in Nature Medicine by Hannun *et al.* [10],

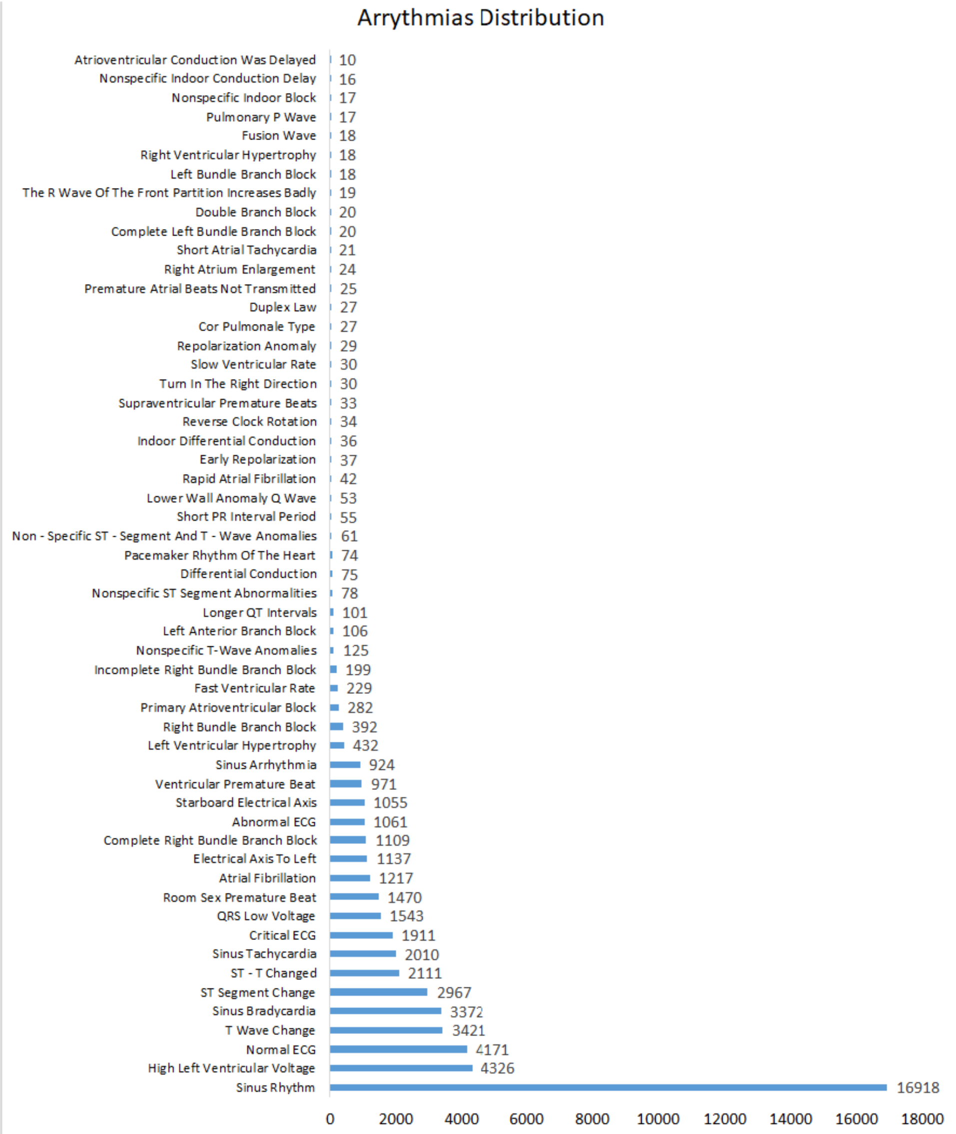

FIGURE 2. **Arrythmia distribution.**

the micro-F1-score of common cardiologists in judging cardiac arrhythmia is only 0.780 currently, which also indicates that even with the experience of experts, misdiagnosis is inevitable and even frequent. While using deep learning end-to-end methods to detect patients with arrhythmias and help doctors make clinical diagnoses would be a technological leap for humanity.

When we further observe these 55 categories arrhythmias, we will find that there are often correlations (mutual exclusion, inclusion, etc.) between certain categories. For example, sinus rhythm and QRS low voltage often appear at the same time, and the right axis deviation and the left axis deviation will not coexist. In view of the characteristics of the multi-label classification task, it is very necessary for us to dig deep into the potential correlation among each category. In order to quantify this correlation, we define the correlation coefficient $corr(x, y)$ as shown in (3), which represents the degree of association between the x label and the y label. After calculating the correlation coefficient among 55 labels, we can get a 55×55 relationship matrix. The thermal diagram of the relationship matrix is shown in Fig. 4, from which we can roughly see the relationship among these labels.

$$Corr(x, y) = \frac{x \cap y}{x \cup y} \tag{3}$$

B. 1-D CONV FOR ECG

CNNs have been proved to be a very effective neural network model in computer vision tasks such as image recognition and target detection. CNNs obtain the spatial characteristics of feature information through the sliding window mechanism of convolution kernel. Typically, in a computer vision task, the input image is a three-dimensional piece of information,

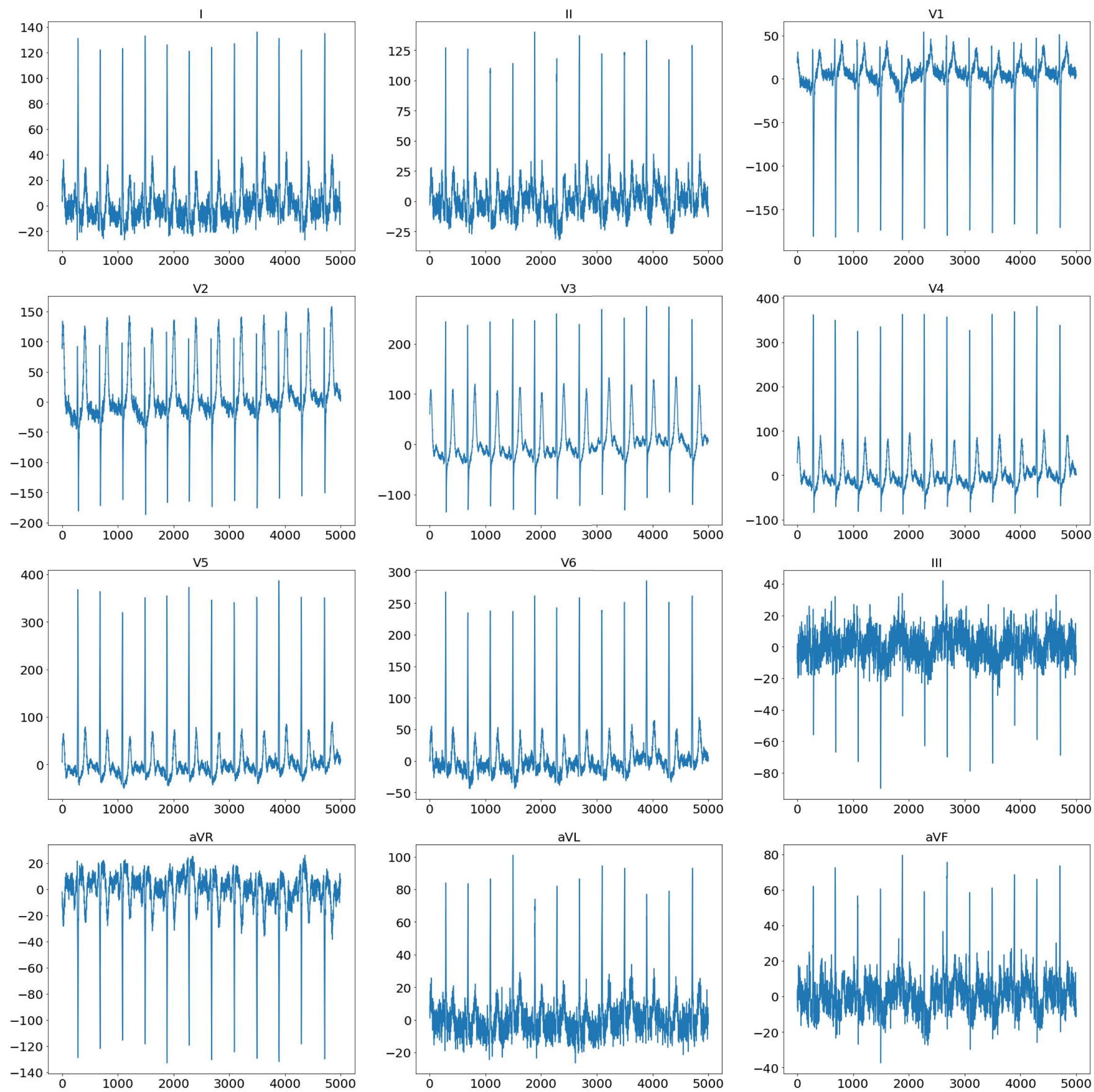

FIGURE 3. 12 leads of ECG.

including the length and width which represent the content of the image, and RGB channel which represents the color information of the image. So, when we convolve the image, what we need is a 2-D convolution kernel.

As shown in Fig. 5, we can find that the abnormal ECG signal often appears in a local area of the signal and reflects the characteristics of a space, which is similar to the principle of image recognition. Kiranyaz *et al.* [11] also mentioned the effectiveness of 1-D convolution in the ECG detection applications, rather than 2-D convolution which is commonly used in computer vision tasks. At the same time, 1-D convolution can be very efficient to extract characteristic information of ECG lead finally to classify. However, this research simply stacked the convolution layers for experiment, and hadn't come up with a very typical model or algorithm.

The principle and calculation process of 1-D convolution is similar to the ordinary operation of convolution. The convolution kernel slides on the feature map, and then the convolution kernel multiplies the value of the corresponding position in the feature map and adds it up. Finally, we get the new feature map. For example, as shown in Fig. 6, the original feature map is [3,1,1,5,6] and the convolution kernel is [-2,2,1]. We use the convolution operation with paddings to calculate the values of 3,10,3,-1,14, 2,-12 for each position. Finally, we can get the feature map as shown in Fig. 7.

Multi-channel 1-D convolution is very suitable for multi-lead ECG detection tasks. Different channels correspond to different leads, and each lead itself is a 1-D information so that using 1-D convolution is just right. For these reasons, we can transfer the models that are

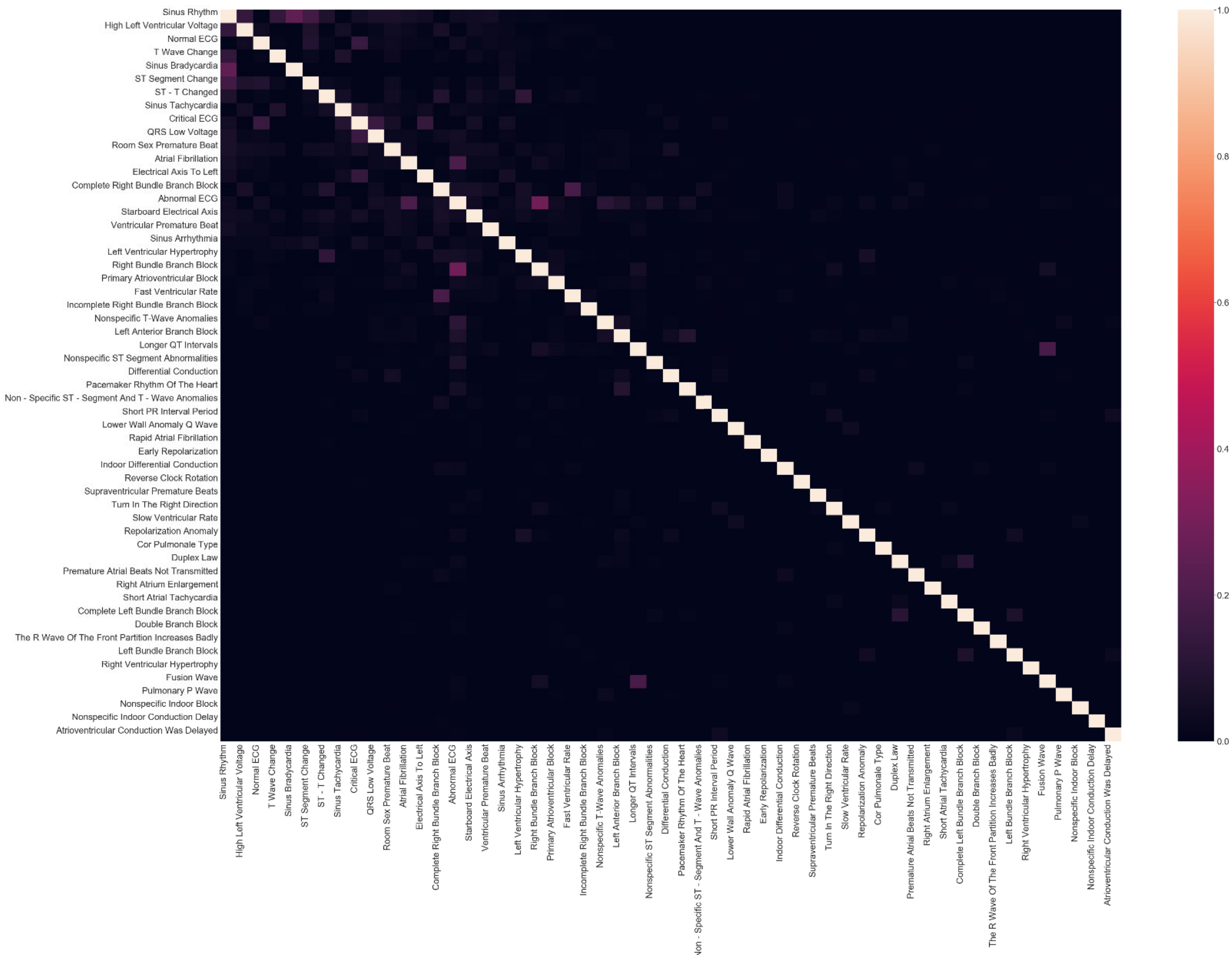

FIGURE 4. Heatmap of arrhythmia correlation.

commonly used in computer vision classification to classify ECGs with a few modifications.

C. MULTI-ECGNet

Above, we have shown that convolutional neural networks not only perform well in computer vision tasks, but also play a key role in ECG detection. After years of development and efforts of numerous researchers, the convolutional neural network has emerged AlexNet, VGG, Inceptions, ResNet, DenseNet, SENet and other models. However, which of these models is better suited to the task area of ECG testing?

Byeon *et al.* have done experiments by using Physikalisch Technische Bundesanstalt (PTB) ECG database to compare the performance of ResNet, DenseNet and Xception in ECG [12]. The conclusion is that Xception and ResNet perform better than DenseNet in most experiments and ensemble CNNs have shown higher than single CNN.

So how about we combine the neural architecture characteristics of ResNet and Xception? The ResNet proposed by He *et al.* solved the problem that the gradient disappearing in the DNN model during the back propagation through Shortcut Connections structure and Residual Learning [13]. The greatest contribution of the Xception model proposed by Chollet *et al.* is the depth separable convolution structure, which uses the point wise convolution to enhance the information exchange between channels, and greatly reduces the number of parameters and computation even speeds up the convergence of the model [14].

The 12 leads of ECG correspond exactly to the 12 channels of convolution, so the lead can be regarded as the channel of inputs in the architecture design of neural network. While all of these leads are crucial in an ECG detection, in fact, only one or parts of them are the decisive factors. Most things that happen in the world follow the 2-8 rule, and it is the small pieces of information that are most important to the whole. So which lead is the dark horse?

Maybe the Attention mechanism can help. In fact, relevant research has been published. The work of Hu *et al.* proposed the SENet network architecture, in which the most important contribution is the SE Block [15]. Firstly, The SE module performs *Squeeze* operation on the feature map obtained by convolution to obtain channel-level global features. And then, SE module performs *Excitation* operation on the global features to learn the relationship among each channel and the weights of different channels. Finally, the weights multiply the original feature map to get the final features. Essentially,

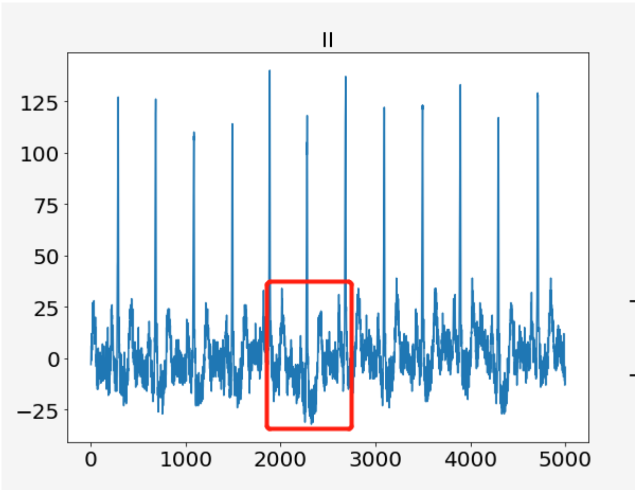

FIGURE 5. Abnormal signal of ECG.

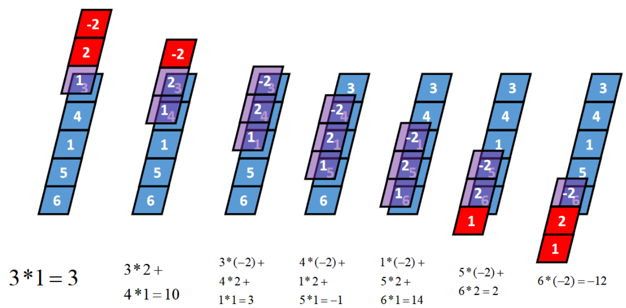

FIGURE 6. Computational process of 1-D convolution.

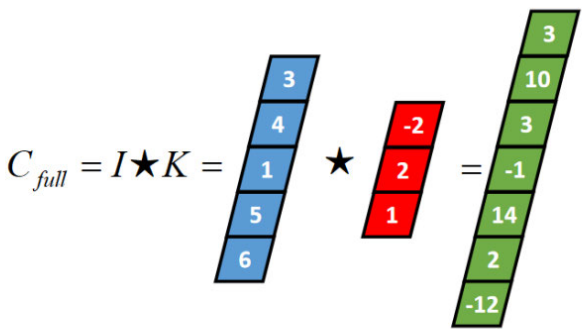

FIGURE 7. Feature map after convolution.

the SE block performs *attention* operations on the channel dimension. This attention mechanism allows the model to pay more attention to the channel features with the largest amount of information, while suppressing those unimportant channel features. Another point is that the SE block is universal, which means that we can also use it in the ECG detection network.

Therefore, as shown in Fig. 8, our model is based on the integration of ResNet, Xception and SE module, fully absorbing the advantages of various architectures, and can be used in ECG detection.

This proposed architecture has the following key features:

1) Inverted pyramid convolution kernel arrangement: When processing ECG signals in the early stage, a longer convolution kernel is used, and the size of the convolution kernel is continuously reduced in the later stage, which can play the role of multi-scale feature extraction and also reduce Part of the amount of calculation.
2) The combination of deep separable convolution and SE module: Both of them can strengthen and extract the correlation of different channel information, and the combination of them can achieve better results.
3) Feature fusion: Fusion of age and gender information into the final feature map has a certain effect on classification.

IV. IMPLEMENTATION

A. DATA PREPROCESSING

Considering that the original ECG dataset may have some errors such as noise during measurement, incorrect labeling, uneven data distribution, missing data and others, which will lead to unexpectedly bad results in subsequent model training if the data are not cleaned and preprocessed properly.

In data preprocessing, this paper uses the following methods to solve the problems faced.

1) DATA AUGMENTATION

When ECG raw data are collected, noise may be caused by abnormal fluctuations in the data due to equipment accuracy or the physical condition of the subject. In order to mitigate the impact of these noises, the usual method is data augmentation. The specific measure of data augmentation often needs to be combined with the actual situation. For the noise of ECG, a feasible method is to randomly add Gaussian additive noise and Gaussian multiplicative noise on the basis of the original signal.

2) LABEL SMOOTHING

ECG label data is produced by professional personnel, but it is inevitable that there will still be a small number of erroneously marked data in large data sets. During training, if there are wrong labels, the training results will be biased. In order to alleviate this situation, Christian Szegedy *et al.* first proposed the label smoothing method in 2016 [16]. This method uses the weighted average of the hard target and the uniform distribution on the label as the soft target, which can significantly improve the generalization ability and learning speed of various neural networks. Later, Hinton *et al.* explained the role and application scenarios of label smoothing in more depth [17].

In a classification problem, if a hard target is used as the optimization objective, the cross entropy function can be written as:

$$H(y, p) = \sum_{k=1}^{K} -y_k log(p_k) \tag{4}$$

If the correct class y_k is 1, otherwise it is 0, p_k is the possibility that the neural network predicts that it is the kth

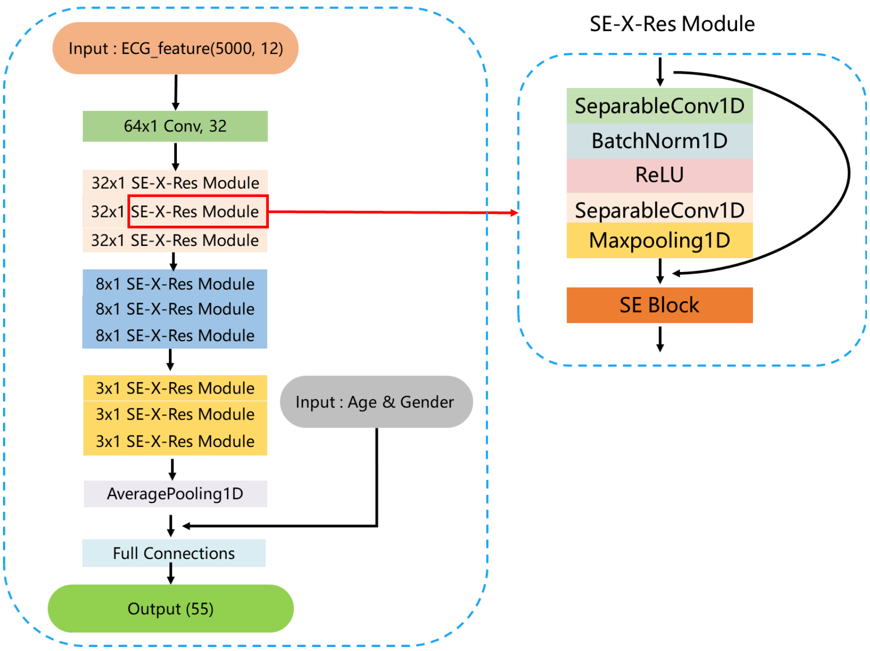

FIGURE 8. Architecture of Multi-ECGNet.

class. If we use a soft target, we can replace y_k with the following equation (5):

$$H_k^{LS} = y_k(1-\alpha) + \alpha/K \tag{5}$$

Label smoothing encourages the result after the penultimate layer activation function to be close to the template of correct class, and also away from the template of wrong class. It is precisely because of this mechanism that it is very appropriate to use label smoothing to handle incorrectly labeled ECG labels here.

3) COMPLETION OF MISSING DATA

In the original data set, part of the sample was missing in age or gender. In this paper, linear regression is used to fill the missing values of age. While for gender, male, female and missing are marked as [0, 1, 2].

B. LOSS FUNCTIONS

According to the specific scenarios and problems that we are facing, classification tasks usually include: binary-classification tasks, multi-classification tasks, multi-label-classification tasks.

1) BINARY-CLASSIFICATION TASKS

There are two categories in this classification task. The input is represented by the feature vector x, and the output is yes and no with $y = 0\ or\ 1$. The Binary-Classification task assumes that each sample is set with only one label 0 or 1.

2) MULTI-CLASSIFICATION TASKS

Multi-classification task means that there are multiple categories in the classification task, and each sample is set with one and only one label among $[0, 1, 2, \ldots, n]$.

TABLE 2. Functions of classification tasks.

Classification	Activation function	Loss function
Binary-Classification	Sigmoid	binary_crossentropy
Multi-Classification	Softmax	categorical_crossentropy
Multi-label-classification	Sigmoid	binary_crossentropy

3) MULTI-LABEL-CLASSIFICATION TASKS

Each sample of the multi-label classification can have multiple labels, and each label does not exist separately. For example, an apple can be a plant and also a fruit. The label for each sample can be expressed as $[0, 1 \ldots 0, 1]$, which means whether the sample belongs to a certain category. For different tasks, in actual model training, we need to adopt different activation functions and loss functions, as shown in Table 2.

Usually the loss function used in the multi-label classification tasks is *binary_crossentropy*, which will output the probability corresponding to each label in the output layer. In the ECG detection tasks, we are faced with the problem that how to determine the probability distribution of each sample in 55 categories arrhythmias. However, this problem is not so simple. As mentioned in the Section III Datasets and Data Analysis, these 55 categories arrhythmia samples are extremely unevenly distributed. The model will tend to judge the result as a larger category if we don't deal with unbalance dataset, which is obviously not a reasonable situation.

In order to solve the sample imbalance problem, the most common method is to increase the weight of the corresponding label when calculating the loss function according to the number of samples of each label. Specifically, the larger the number of samples, the lower the label weight, and the smaller the number of samples, the higher the weight.

The binary cross-entropy loss, weight, and weighted cross-entropy loss formula are as follows:

$$BCE = -\sum_{i}^{n} [y_i * logx_i + (1 - y_i) * (1 - logx_i)] \tag{6}$$

$$W_i = \frac{1}{log(n_i + 1)} \tag{7}$$

$$BCE_w = -\sum_{i}^{n} W_i * [y_i * logx_i + (1 - y_i) * (1 - logx_i)] \tag{8}$$

The weighted binary cross-entropy loss can alleviate the problem of category imbalance to a certain extent, but it only focuses on the number of samples, but fails to distinguish some easily distinguishable samples from ambiguous samples. In other words, among multiple samples of the same label, some samples can be classified almost obviously, but others often make mistakes. In order to solve this problem, Lin *et al.* [18] proposed Focal Loss based on the weighted loss function. This function can reduce the weight of easy-to-classify samples, so that the model is more focused on the hard-to-classify samples during training. The formula of Focal loss can be expressed as follows:

$$FL = -\sum_{i}^{n} W_i * (1 - x_i)^{\gamma} [y_i * logx_i + (1 - y_i) * (1 - logx_i)] \tag{9}$$

γ represents an adjustment factor. When $\gamma = 0$, the function is the ordinary cross-entropy loss function. While γ increases, the degree of adjustment will also increase.

Due to the effectiveness of Focal loss, applying it to the task of ECG detection can also play a very good role, and the relevant results can be seen in subsequent experiments.

V. EXPERIMENTS AND RESULTS

A. DIVIDE THE VALIDATION SET

In the dataset, we have 24106 samples as the training set and 8036 samples as the test set. In order to make full use of the samples in the training set, firstly, we divide the verification set with the training set and the verification set according to the 9: 1 ratio. Especially, we ensure that each part contains at least 1 sample of each category.

B. DATA PREPROCESSING

The specific method of how to preprocess the original data has been mentioned in the fourth part. Including data enhancement, label smoothing, missing value processing, and signal value standardization processing.

C. TRAINING PARAMETER CONFIGURATION

When training, we configured the following hyperparameters for experimentation:

- Optimizer:Adam
- Learning Rate:origin lr 1e-3 with decay
- Batchsize:32,64,128
- Epoch:100
- egularization strategy: Earlystopping,L2

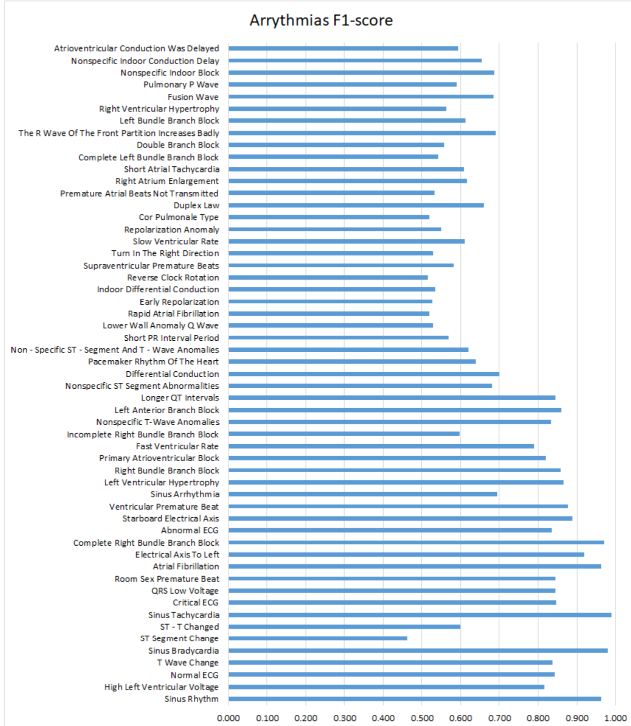

FIGURE 9. Arrythmias micro-F1-score.

In the experiments, we got the highest micro-F1-score of 0.863. At the same time, the micro-F1-score of each ECG arrythmia was calculated, as shown in Fig. 9. We can see that the scores of Sinus Rhythm, Sinus Bradycardia, Sinus Tachycardia, Complete Right Bundle Branch Block, and Atrial Fibrillation are reaching 0.962, 0.980, 0.991, 0.971, 0.963 and these labels take large proportion of dataset. Arrhythmias such as ST Segment Change, Atrioventricular Conduction Was Delayed, and Nonspecific Indoor Conduction Delay perform poorly. But these proportions are relatively small. In general, our model can reach a high-level performance in ECG detection, which has exceeded the performance of ordinary cardiologists in judging arrythmia in 0.780 [10], but we still have a gap in the recognition of some diseases, which needs further study.

At the same time, we also compare the differences between the original ResNet, ResNet + Xception, ResNet + SE and the Multi-ECGNet(SE+Xception+ResNet) of this paper. As shown in Fig. 10, when using the original ResNet34 for training, the training curve begins to converge gradually around 50 epochs. Although it performs best at the beginning, it is later surpassed and eventually reaches 0.834 at 100 epoch, which is used as a benchmark. In order to prove the effectiveness of Xception and SE module, and their effect on cross-channel information fusion and extraction, we use depthwise separable convolution in Xception and SE module in two sets of experiments to compare. We find the performance improvement is about 0.019 and 0.025, respectively. Furthermore, we add the depthwise separable convolution and the SE module together, which is the Multi-ECGNet

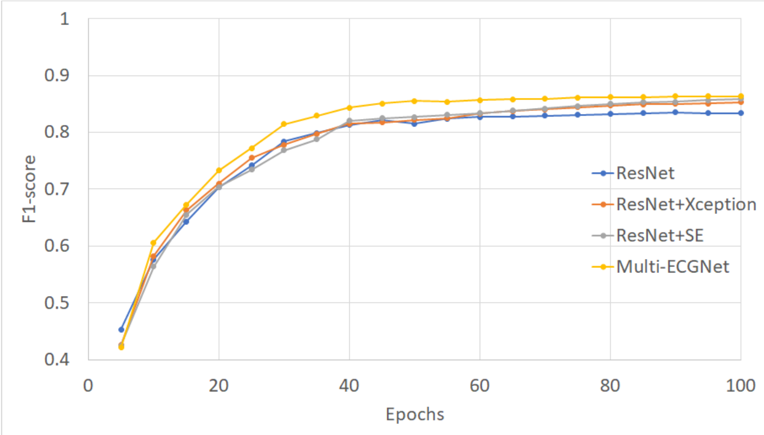

FIGURE 10. **Traning curves of ResNet, ResNet+Xception, ResNet+SE, Multi-ECGNet.**

model proposed in this paper. Combining the advantages of these two, the best performance improvement finally reaches 0.029 which means we get 0.863.

The experiment resuts prove that the Multi-ECGNet proposed in this paper can achieve the current high level of performance in the ECG detection, which has exceeded the recognition standard of ordinary cardiologists, and can assist doctors to detect patients in clinical diagnosis and treatment.

VI. CONCLUSION

In this paper, we study the method of arrythmia multi-label classification based on the ECG 12-lead data set. Through the preprocessing of data, the improvement of the loss function and Multi-ECGNet model proposed in this paper by integrating depthwise separable convolution and SE block, we got the best performance of 0.863 by using micro-F1 as the evaluation criterion, which shows that the proposed method can reach a high-level performance in the arrythmia detection field. Of course, we think that replacing the backbone network with the latest network model such as EfficientNet is likely to achieve better results, which is also what we will do in the future.

The methods and experimental results of this paper also have some limitations. Our data set is still not enough to cover all arrythmias and more diverse populations. And our experimental results have not yet been clinically verified, which requires further research with medical experts.

ACKNOWLEDGMENT

(Junxian Cai and Weiwei Sun contributed equally to this work.) The authors would like to thank the anonymous reviewers for their valuable comments which helped them to improve the content, organization, and presentation of this paper.

REFERENCES

[1] L. S. Lilly, *Pathophysiology of heart disease: A collaborative project of medical students and Faculty*, 5th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2013.

[2] M. Rivera-Ruiz, C. Cajavilca, and J. Varon, "Einthoven's string galvanometer: The first electrocardiograph," *Texas Heart Inst. J.*, vol. 35, no. 2, p. 174, 2008.

[3] F. G. Yanowitz. *Ecg Learning Center Created by Eccles Health Sciences Library at University of Utah*. Accessed: Apr. 1, 2020. [Online]. Available: https://ecg.utah.edu/lesson/1

[4] EMTRESOURCE.COM. *12-Lead Ecg Placement Guide With Illustrations*. Accessed: Apr. 4, 2020. [Online]. Available: https://www.cablesandsensors.com/pages/12-lead-ecg-placement-guide-with-illustrations

[5] J. Kim, H. Shin, K. Shin, and M. Lee, "Robust algorithm for arrhythmia classification in ECG using extreme learning machine," *Biomed. Eng. OnLine*, vol. 8, no. 1, p. 31, 2009.

[6] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018.

[7] U. B. Baloglu, M. Talo, O. Yildirim, R. S. Tan, and U. R. Acharya, "Classification of myocardial infarction with multi-lead ECG signals and deep CNN," *Pattern Recognit. Lett.*, vol. 122, pp. 23–30, May 2019.

[8] Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, pp. 411–420, Nov. 2018.

[9] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Comput. Biol. Med.*, vol. 99, pp. 53–62, Aug. 2018.

[10] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, p. 65, 2019.

[11] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[12] Y. Byeon, S. Pan, and K. Kwak, "Ensemble deep learning models for ecg-based biometrics," in *Proc. Cybern. Informat. (K&I)*, Feb. 2020, pp. 1–5.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[17] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 4696–4705.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

JUNXIAN CAI was born in Fujian, China, in 1995. He received the B.S. degree in the Internet of Things engineering from the University of Science and Technology Beijing, China, in 2017. He is currently pursuing the master's degree in computer science with the Beijing University of Posts and Telecommunications (BUPT). His research interests include machine learning, deep learning, and the combination of interdisciplinary subjects such as medical science for about two years.

WEIWEI SUN received the B.S. degree in traditional Chinese medicine from the Henan University of Traditional Chinese Medicine, Zhengzhou, China, in 2005, and the Ph.D. degree in the nephropathy of Chinese medicine from the Beijing University of Chinese Medicine, Beijing, China, in 2010. From 2017 to 2018, she was a Research Scholar in dynamical biomarkers (DBiom) at the Beth Israel Deaconess Medical Center (BIDMC)/Harvard Medical School. From 2010 to 2015, she was an Attending Physician with the Dongzhimen Hospital, Beijing University of Chinese Medicine, where she has been an Assistant Professor since 2016. She has authored more than 20 articles. Her research interests include nephropathy and the analysis of complex physiological signals.

JIANFENG GUAN (Member, IEEE) received the B.S. degree in telecommunication engineering from Northeastern University, Shenyang, China, in 2004, and the Ph.D. degree in communication and information system from Beijing Jiaotong University, Beijing, China, in 2010.

From 2010 to 2015, he was a Lecturer with the Institute of Network Technology, Beijing University of Posts and Telecommunications, where he has been an Assistant Professor since 2016. He has authored more than 100 articles and holds more than 70 inventions. His research interests include future network architecture, network security, and mobile Internet. He serves as a TPC Member of INFOCOM MobilWorld, in 2011 and from 2015 to 2017, MobiSec, from 2016 to 2019, ICCE 2017, ICC 2018 CCNCP, GLOBECOM 2018 CCNCPS, and WCNC 2019. He was a recipient of several Best Paper Awards from the ACM Mobility Conference 2008, IC-BNMT 2009, and Mobisec 2018. He is also a Reviewer of journals such as the IEEE Transactions on Vehicular Technology, TB, CC, CN, JNCA, FGCS, IEEE Access, INS, SCN, IJSSC, IJAHUC, and JoWUA.

ILSUN YOU (Senior Member, IEEE) received the M.S. and Ph.D. degrees in computer science from Dankook University, Seoul, South Korea, in 1997 and 2002, respectively, and the second Ph.D. degree from Kyushu University, Japan, in 2012. From 1997 to 2004, he was at Thin Multimedia Inc., Internet Security Company Ltd., and Hanjo Engineering Company Ltd., as a Research Engineer. He is currently an Associate Professor with the Department of Information Security Engineering, Soonchunhyang University. Especially, he has focused on 4/5G security, security for wireless networks, mobile Internet, the IoT security, and so forth while publishing more than 180 articles in his research areas. He is a Fellow of the IET. He is currently serving or has served as the General Chair or a Program Chair of international conferences and workshops such as WISA, MobiSec, from 2016 to 2019, AsiaARES, from 2013 to 2015, MIST, from 2009 to 2017, MobiWorld, from 2008 to 2017, and so forth. He is the Editor-in-Chief of the *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications* (JoWUA). He serves on the Editorial Board of *Information Sciences* (INS), the *Journal of Network and Computer Applications* (JNCA), IEEE Access, *Intelligent Automation and Soft Computing* (AutoSoft), the *International Journal of Ad Hoc and Ubiquitous Computing* (IJAHUC), *Computing and Informatics* (CAI), and the *Journal of High Speed Networks* (JHSN).