

A ResNet based multiscale feature extraction for classifying multi-variate medical time series

Junke Zhu¹, Le Sun^{1*}, Yilin Wang¹, Sudha Subramani², Dandan Peng³,
and Shangwe Charmant Nicolas¹

¹ School of Computer Science, and Department of Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing 210044, China

[e-mail: 201983290338@nuist.edu.cn, LeSun1@nuist.edu.cn, 20191220029@nuist.edu.cn, charmantshangwe@yahoo.fr]

² Victoria University, Australia

[e-mail: Sudha.Subramani@vu.edu.cn]

³ School of Computer Science and Network Engineering, Guangzhou University, Guangdong, China

[E-mail: 1112106006@e.gzhu.edu.cn]

*Corresponding author: Le Sun

*Received February 9, 2022; revised March 30, 2022; accepted April 24, 2022;
published May 31, 2022*

Abstract

We construct a deep neural network model named ECGResNet. This model can diagnosis diseases based on 12-lead ECG data of eight common cardiovascular diseases with a high accuracy. We chose the 16 Blocks of ResNet50 as the main body of the model and added the Squeeze-and-Excitation module to learn the data information between channels adaptively. We modified the first convolutional layer of ResNet50 which has a convolutional kernel of 7 to a superposition of convolutional kernels of 8 and 16 as our feature extraction method. This way allows the model to focus on the overall trend of the ECG signal while also noticing subtle changes. The model further improves the accuracy of cardiovascular and cerebrovascular disease classification by using a fully connected layer that integrates factors such as gender and age. The ECGResNet model adds Dropout layers to both the residual block and SE module of ResNet50, further avoiding the phenomenon of model overfitting. The model was eventually trained using a five-fold cross-validation and Flooding training method, with an accuracy of 95% on the test set and an F1-score of 0.841. We design a new deep neural network, innovate a multi-scale feature extraction method, and apply the SE module to extract features of ECG data.

Keywords: Multi-scale convolutional feature extraction methods, ResNet50 structure, Squeeze-and-Excitation Modules

1. Introduction

Cardiovascular disease is the number one cause of death worldwide and is a common disease among middle-aged and elderly people over the age of 50, posing a serious threat to human health[1]. The number of people with cardiovascular disease in China has exceeded 330 million. A large number of scholars have invested in the optimization of ECG systems and sensors to closely integrate healthcare with the Internet of Things[2] (e.g., algorithm optimization[3], protocol improvement[4] and error correction[5]). However, the average time from onset to death of malignant diseases is 12 minutes, with high prevalence, high disability, and high mortality. Even with the application of today's most advanced treatment methods, more than 50% of people with the disease are still unable to take care of themselves completely[6]. The number of people who die from cardiovascular disease is as high as 15 million each year worldwide[7]. Therefore, the importance of accurate clinical identification and diagnosis of cardiovascular disease cannot be overstated, and the ECG signal test is the most commonly used test to assess cardiac function.

The ECG signal reflects the continuous changes in the heart over a period of time. The 12 leads of the ECG include six limb leads (I, II, III, aVR, aVL, aVF) and six chest leads (V1~V6). A complete cycle of ECG signal consists of QRS, P, and T waves together. The start, end, peak, trough, and interval of the waveform respectively record detailed information about the state of heart activity, which provide an important basis for analysis in the diagnosis of heart disease. While ECG signals are susceptible to interference, a person's waveform is different at different times[8]. In traditional medicine, doctors have to visually capture the changes in the waveform during each cycle and correlate the waveform between cycles, remove the effects of interference in the ECG signal and extract the main features to determine the patient's heart disease. The variety of heart diseases and the potential for multiple diseases to be reflected in a single ECG signal, as well as the high level of expertise required to read an ECG, make it extremely difficult for doctors to accurately diagnose a patient's disease by reading an ECG[9]. The current accuracy rate for human cardiologists is roughly 78%, with plenty of room for improvement[10].

In recent years there has been a significant amount of research work using deep learning techniques to intelligently classify ECG signals to assist doctors in making diagnostic decisions. Rasmus et al.[11] used an end-to-end model combining CNN and a recurrent neural network RNN to classify ECG data into atrial fibrillation and normal sinus heart rate in 2019. They achieved 98.98% sensitivity and 96.95% specificity for the two conditions by 5-fold cross-validation, respectively. In 2020 Li et al.[12] proposed the use of deep residual networks for ECG diagnosis, developing a 31-layer 1D residual convolutional neural network. The net includes four residual blocks, each consisting of three 1D convolutional layers, three batch normalization layers, three rectified linear unit layers, and a fully connected layer. They used the global public dataset MIT-BIH as the training dataset. The classification accuracy for the 2-lead signal is up to 99.38%. The four main methods of deep learning for ECG classification today are CNN, RNN, LSTM, and ResNet. CNN reduces the complexity of the network through three major strategies: local perceptual field, weight sharing, and downsampling. CNN is now widely used in the field of medical imaging diagnosis. RNNs, which work extremely well for speech recognition and text inference, are more concerned with the temporality of ECG data and are more sensitive to short sequence data. LSTM is an improvement on RNNs, which allows the network to handle long-range dependencies. LSTM used in ECG classification generally achieves better results than the other. ResNet is our most promising neural network model. The introduction of the residual block can better solve the problem of

gradient disappearance and gradient explosion. As a result, ResNet allows us to build a deeper network to explore the subtle changes and correlations of ECG data between different leads, further improving the accuracy of the model.

An improved multi-scale decomposition enhanced residual convolutional neural network proposed by Cao et al., [13] requires the data to be decomposed by a wavelet framework and then reconstructed into confident good samples. The three FDResNet were then coupled into a multi-scale decomposition-enhanced residual convolutional network by migration learning. The final accuracy of 92.1% was obtained. The data pre-processing of this method is very complex. A series of operations such as noise reduction and decomposition reconstruction will take up a lot of computational resources. However, the accuracy rate needs to be improved. The end-to-end ResNet framework developed by Deepankar et al., [14] can only classify into four categories based on ECG fragments, the combined CNN and RNN approach used by Rasmus et al., [11] can only classify into two categories based on ECG data. The number of classifications in some models is too small to be adequate. Sherin et al., [15] used ResNet to classify ECG detection, but only single-lead data were processed. There is a discrepancy between this model and the 12-lead channels of the ECG. In addition, LSTM and RNN are extremely difficult to train and require extremely high hardware requirements. Google, Meta, and other companies [16] have adopted Attention-based models to replace RNN and its variants to achieve approximate accuracy while greatly reducing training costs.

To solve the problem of gradient disappearance and gradient explosion and to build deep neural networks, we successfully built and trained a 50-layer and deeper network model using the ResNet structure proposed by Kaiming He [17] in 2015. We aimed to process and classify 12-lead ECG data and adopted the approach of the SE module in the paper Squeeze-and-Excitation Networks published by Jie et al., [18] in 2017, which effectively processed the relationship between channels by performing Squeeze and Excitation operations on each channel, making the processing of 12-lead ECG data to be processed well. In order to save computing resources as much as possible, we propose a multi-scale convolutional feature extraction method based on the idea of a feature pyramid network (FPN) [19] to replace the tedious data pre-processing work. By combining this module with the main body of the ResNet network structure, the computing resources and training hardware requirements are reduced to a very low level.

In this paper, we propose the ECGResNet model. The model uses the 16 Blocks of ResNet50 as the main body of the model. We add an SE module in each residual block to handle the channel-to-channel relationship. And we combine convolution with convolutional kernels of 8 and 16 respectively as a multi-scale convolutional feature extraction method to replace the original convolutional layer of ResNet with a convolutional kernel of 7. Finally we use a fully-connected layer to integrate age, gender, and other factors. The main innovations of this paper are as follows:

(1) A multi-scale convolutional feature extraction method is proposed, where the input data does not require tedious data pre-processing work, such as noise reduction.

(2) After extensive experiments, ResNet50 was used as the main body of the model, which is moderately deep and effectively avoids the phenomenon of gradient disappearance and gradient explosion. Combining the modified SE module with ResNet50 enables the model to handle 12-lead input data. Through extensive experiments, the effectiveness of our proposed classification algorithm is demonstrated.

(3) A fully connected layer is used, taking into account patient gender and age data.

Section 2 describes the related work. Section 3 introduces our proposed ECGResNet model. Section 4 shows our experimental setup and structure. Section 5 concludes this question finally.

2. Related Work

2.1 Literature Review

In 2019 Annisa et al., [20] proposed an RNN-based deep learning sequence modeling method for the classification of ECG rhythm signals. The performance of this model was compared with an improved RNN followed by an LSTM model, and the results showed that the improved LSTM outperformed the standard RNN in all cases with the same network model hyperparameters. In the same 2019, Andrew Ng's team [21] proposed a solution to apply deep learning techniques to the field of ECG diagnosis by developing a deep neural network. The network used is a 1-dimensional 34-layer convolutional neural network. The network can diagnose heart rate irregularities based on single-lead ECG signals. The ECG data can obtain powerful classification performance without extensive pre-processing such as Fourier transform or wavelet transform, achieving the expected results. The dataset used by Ng's team contains a sample of 91,232 single-lead data from 53,877 adult patients. A 30-second recording from each sample was taken for labeling as the dataset. The deep neural network was trained using this dataset. The final model means F1-score (the summed mean of positive predictive values and sensitivities) reached 0.837, surpassing the mean cardiologist (0.780). Deepankar et al., [14] used deep residual networks (DRN) to classify different ECG segments into four categories, namely AF, normal rhythm, other non-noisy symptomatic ECG rhythms, and noisy rhythms. The method achieved an F1-score of 0.88 on the test set, outperforming CNNs and RNNs. Cao et al., [13] proposed an improved multi-scale decomposition enhanced residual convolutional neural network. The network re-segmented the original ECG recordings with large differences in length into short samples of 9s and decomposed and reconstructed the segment samples into sub-signal samples of different scales using a derived wavelet framework decomposition. Three FDResNet with good performance were coupled into a multi-scale decomposition-enhanced residual convolutional network using migration learning techniques. An average accuracy of 92.1% was achieved for single-lead signals based on a large ECG dataset provided by the 2017 PhysioNet/CinC competition.

The model built by our team finally chose ResNet50 as the main body, which contains more than 50 convolutional layers. This design significantly reduces the occurrence of gradient disappearance and gradient explosion compared with existing RNN and LSTM models due to the effect of residual blocks. Our team focuses on the intelligent diagnosis of 12-lead ECG data, which is more complex than the data handled by the existing methods described above. We have added the SE module as an attention mechanism [22] between channels, which complements ResNet to build deep neural networks more effectively. Inspired by Feature Pyramid Networks (FPN), a multi-scale feature extraction method consisting of two convolutional layers of different sizes is constructed. This method is more efficient than the improved multiscale decomposition enhanced residual convolutional neural network proposed by Cao et al., by reducing the cumbersome data pre-processing operations while achieving the same accuracy. Finally, taking additional factors such as gender and age into account, this detail was found to improve accuracy by 1.5% and F1-score by 1.2%.

2.2 Background Knowledge

The depth of the neural network is closely related to the performance of the model, with more layers enabling more complex feature extraction operations and, in theory, better performance. In actual experiments, when the depth of the network reaches a certain level, various indicators such as accuracy rate tend to saturate or decline, i.e., problems such as gradient disappearance and gradient explosion occur, which will make it extremely difficult

to train deep neural network models. ResNet is a milestone in the history of CNN imagery, with 19 layers in the famous VGG network in 2014 and 152 layers in ResNet in 2015. ResNet has effectively solved the degradation problem through its residual learning method, allowing the network to be as deep as it should be. In the end, it achieved five firsts, breaking the history of CNN models in ImageNet.

The ResNet network is a modification of the VGG19 network with the addition of a residual block implementing a short-circuiting mechanism. The residual structure formula is expressed as $x_{l+1} = x_l + F(x_l, W_l)$. The expression of any deep cell L feature can be obtained by recursion as $x_L = x_l + \sum_{i=0}^{L-1} F(x_i, W_i)$. x_L is the feature of an arbitrarily deep cell L. x_l is the feature of a shallow cell l. F is the residual function. $\sum_{i=0}^{L-1} F$ is the sum of the outputs of all previous residual functions. For backpropagation, assuming a loss function of E, the backpropagation chain rule yields: $\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} (1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, w_i))$. The equation consists of two parts: passing without a weighting layer $\frac{\partial \varepsilon}{\partial x_L}$, passing through the weighting layer $\frac{\partial \varepsilon}{\partial x_l} (1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, w_i))$. The former ensures that the signal can be transmitted directly back to any shallow layer x_i . The formula $\frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} F(x_i, w_i)$ is unlikely to take a value of -1 effectively avoiding phenomena such as gradient disappearance. Under the terms of this residual calculation method, ResNet does downsampling directly using a convolution with a step size of 2 and replaces the fully connected layer with a global average pool layer. Doubling the number of feature map layers whenever the size of the feature map is reduced by half so that the complexity of the network layers remains constant. The short-circuiting mechanism between every two layers forms residual learning. When the network is deeper, residual learning between three layers will be performed, with three convolutional kernels of 1x1, 3x3, and 1x1 respectively. After several experiments with different depths of ResNet in ECG signal classification and diagnosis, the shallow F1-score of ResNet18 and ResNet34 models floated around 70% and it was difficult to improve. While ResNet152 was too deep and the training time increased significantly but the result was not obvious. In summary, ResNet50 and ResNet101 are clearly preferable compared to the other depths.

The network structure of ResNet at each depth is shown in Fig. 1. ResNet50 has the first layer of 7x7x64 convolution, followed by 3, 4, 6, and 3 for a total of 16 blocks. Each block has 3 layers of 1x1, 3x3, and 1x1 convolution, for a total of 48 layers. ResNet101 has 17 more blocks than ResNet50, i.e., 51 more convolutional layers, for a total of 101 layers.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 1. ResNet architecture diagram[17]

3. Methods

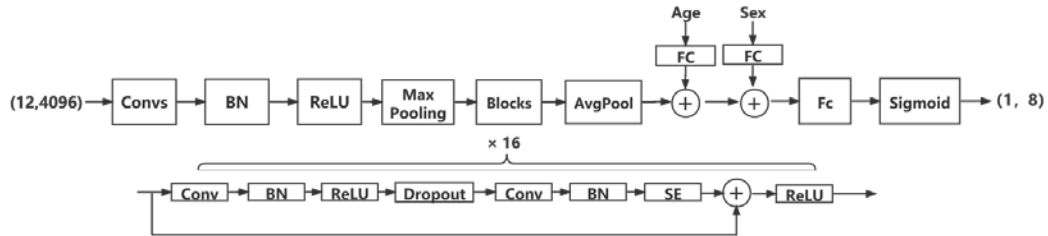


Fig. 2. ECGResNet Model structure diagram

We propose a ResNet-based multiscale CNNs feature extraction for classifying multi-variate medical time series, called ECGResNet, which enables the simultaneous diagnosis of multiple heart diseases. Fig. 2 illustrates the complete structure of ECGResNet. The model consists of a multi-scale feature extraction layer Convs, 16 Blocks, and 3 fully connected layers (Fc). Each Block contains 2 convolutional layers, a Dropout layer, and a Squeeze-And-Excitation module (SE module), which are explained in detail below. Firstly, the 12-lead ECG recordings with different sampling rates were resampled to the same frequency and re-intercepted to the same length, so that all recordings had the same frequency and length. The resampled ECG data were fed into ECGResNet, where the multi-scale convolutional feature extraction layer (Convs) was used for feature extraction. The extracted features are then combined and analyzed by 16 Blocks and fed into the final fully-connected layer in an overlay with the features extracted by the fully-connected layer (FC) for age and gender. Finally, the Sigmoid layer outputs the classification results in the form of probabilities. The details of the above process are explained in detail in the following section.

This neural network model ECGResNet was modified based on ResNet50, using the 16 Blocks of ResNet50 as the main body of the model data processing, with the following improvement points: (1) The convolutional kernels of 8 and 16 were added together as a multi-scale convolutional method for initial feature extraction; (2) an attention mechanism between channels was added to the residual blocks SE module and the addition of a Dropout layer to avoid overfitting; (3) the output section uses the FC multiple times to integrate patients considering age and gender.

The 12-lead ECG data for the input model ECGResNet is simply resampled to the same frequency, intercepted to the same length, and not subjected to any noise reduction. We consider the classification of cardiovascular diseases as a traditional multiclassification problem, on the basis of which we assume that the likelihood of having each disease is independent of each other. Therefore, the Softmax used for single classification problems such as cat and dog classification is not applicable here and we have modified it to a Sigmoid function, as shown in Fig. 3. The output of the ECGResNet model is processed by the Sigmoid function to obtain a 1×8 matrix with each number in the matrix between 0 and 1, indicating the likelihood of having each of the 8 diseases. If the probability of the corresponding disease is greater than or equal to 0.5, then the test subject is judged to have the disease and is represented by 1. If the probability of the corresponding disease is less than 0.5, the detector is judged not to have the disease, denoted by 0. The final ECGResNet model output is a 1×8 Matrix.

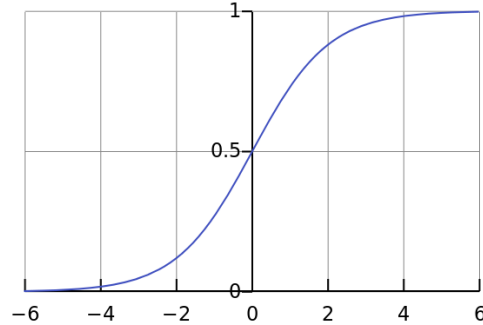


Fig. 3. Function images

Inspired by the idea of Feature Pyramid Networks (FPN) in Computer Vision (CV), the multi-scale convolutional layer *Conv8* uses convolutional layers with 8 and 16 convolutional kernels respectively for feature extraction of the input 12×4096 data, as shown in **Fig. 4**. This method replaces ResNet's original convolutional layer with 7 kernels and aims to focus on the overall trend of the ECG signal while also noticing subtle changes. The convolutional layer with a convolutional kernel of 8 has a step size of 2 and a padding of 3. The convolutional layer with a convolutional kernel of 16 has a step size of 4 and a padding of 7. The feature map from the convolutional layer with a convolutional kernel of 16 is up-sampled and superimposed on the feature map from the convolutional layer with a convolutional kernel of 8 in an additive manner. The result of the convolutional layer is then transformed into a 64×1024 feature map through the BN layer, RELU layer, and maximum pooling layer, which is used as input to the residual block. *Conv8* is used to represent a convolutional layer with a convolutional kernel of 8, *Conv16* represents a convolutional layer with a convolutional kernel of 16, and *upsample* represents an upsampling operation with a parameter of 2. The resulting output Y is shown in (1).

$$Y = \text{Conv8}(y) + \text{upsample}(\text{Conv16}(y)) \quad (1)$$

The input data to the ECGResNet model is a 12-lead ECG signal, which can be considered as 12 channels of 1-dimensional data. The input data is processed in the feature extraction layer of the model and transformed into 64 channels of 1D data, with the number of channels increasing up to 2048. The diagnosis of cardiovascular disease requires a comprehensive consideration of the correlation of information between leads, i.e., the relationship of data between channels, rather than focusing on data within a channel. Following the traditional attentional mechanism commonly used in 3-channel image recognition and performing Seq2seq models on the data within each channel would not be satisfactory. We have focused our attention on considering the channel-to-channel relationship by adding an SE module with a fully connected layer parameter r of 16 applicable to ECG data processing to each residual block of ResNet, as shown in **Fig. 5**. The SE module consists of two parts, Squeeze and Excitation. The global compressed feature volume of the data features is obtained through the Squeeze part. The weights of each channel in the data features are obtained through the Excitation part. And the weighted data features are used as input of the next part of the network. The weights are assigned to each channel individually by this method, thus determining the spatial relationships between the channels. To avoid frequent overfitting in the experiments, a Dropout layer is added between the two fully connected layers, with an optimal parameter of 0.2. The final structure of the SE module is shown in **Fig. 6**. Where C is the number of channels of input data, W is the length of each channel of data, and the size of

input data is $W \times C$. The data input to the SE module is compressed to $1 \times C$ size by global pooling, a global average pooling layer, and then connected to two full connection layers for scaling operation, with the parameters of the full connection layer set to 16. The first full connection layer $fc1$ scales the data to $1 \times 16C$. The Dropout layer will set $0.2 \times 16C$ value to zero. The second fully connected layer $fc2$ scales the data back to $1 \times C$. After the Sigmoid function, the output is C numbers between 0 and 1, which is used as the weight i for each channel, so the final output is shown in (2) and (3).

$$Input = fc1(pooling(input)) \tag{2}$$

$$Output = input \cdot sigmoid(fc2(Dropout(Input))) \tag{3}$$

To further avoid overfitting, a Dropout layer was added between each of the two residual blocks with a parameter setting of 0.5. Finally, gender and age were taken into account. The gender and age parameters were combined with the output of the residual blocks through a preliminary fully-connected layer, and the probability of occurrence of each disease was found jointly through the fully-connected layer and the Sigmoid function.

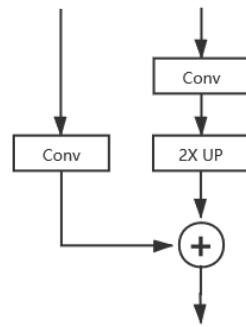


Fig. 4. Multi-scale convolutional methods

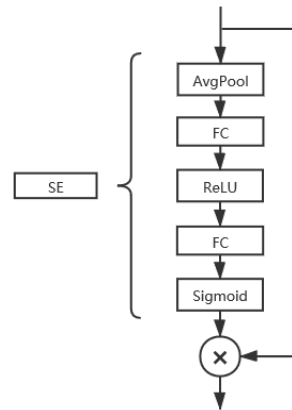


Fig. 5. Improved SE module

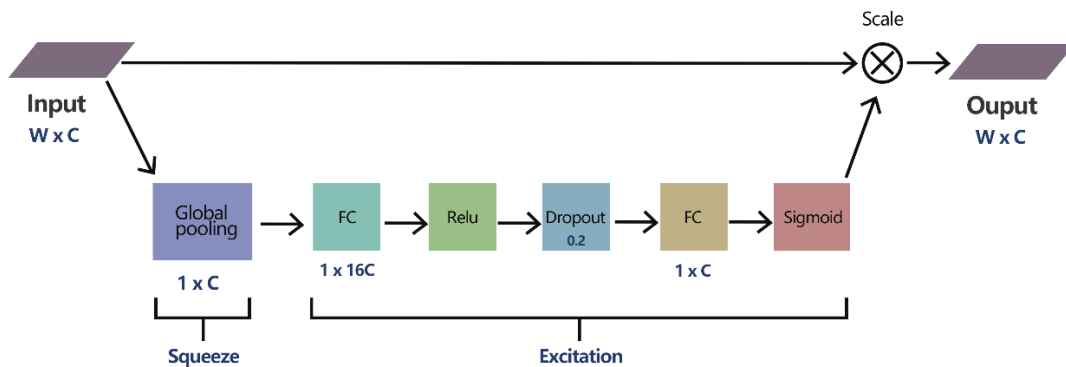


Fig. 6. Complete improved SE schematic

The loss function used for this multi-classification task was the binary cross-entropy loss function BCELoss [23]. The optimizer used in training was Adam, with an initial learning rate of 0.05. The learning rate was adjusted to 1/10 of the previous state when the F1-score [24] of the validation set was reduced during training.

4. Experiment

Our experimental data was prepared and trained on the same computer with an Intel Core i7-8700 CPU, a Geforce RTX 3080 GPU with 10GB of video memory, and 32GB of RAM.

The datasets used for training and validation of the model are saved in the Python WaveForm-DataBase (WFDB) package format. The datasets are 12-lead ECG recordings with a total of 45152 samples from Shaoxing People's Hospital and Ningbo First Hospital. Each recording has a frequency from 500Hz to 1000Hz and durations from 6s to 60s. Each data contains relevant information such as age, gender, and type of disease. Age is an integer from 1 to 100 and some data are missing age information and are labelled as NaN. Each dataset contains gender data female or male. Diseases are represented using a 9-digit code. Due to the limited dataset used, the number of samples for each disease ranged from 1043 to 16559. So as to achieve a more accurate disease determination, we only trained and validated the model for the top 8 common cardiovascular diseases in descending order of number (Including T-wave abnormalities, sinus bradycardia, left ventricular hypertrophy, premature atrial beats, prolonged QT interval, etc.). A single ECG signal data may contain one or more diseases.

4.1 Data pre-processing.

First, we resample all ECG signals to a minimum frequency of 500Hz and each ECG signal is intercepted to 4096 samples. Data with less than 4096 samples were padded with zeros, and data longer than the required number of samples were added to the dataset by randomly cropping in a window of 4096. The ages are normalized to the range [0,1] using a linear function. If age is unknown then it is uniformly represented by 0. Sex is set to 1 for males and 0 for females, and the type of disease in the dataset header file is represented as a matrix with 1 row and 8 columns. If the disease is present, it is indicated by a 1 in the corresponding subscript position, otherwise it is 0. Other diseases are considered not to be present.

ECG signals are characterized by weakness, low amplitude and low frequency, and are highly susceptible to interference and noise. Commonly used noise reduction methods are mainly wavelet transform and Fourier transform. Our team has conducted several experiments on the wavelet transform, using the continuous wavelet transform (CWT) to transform the time domain signal to the frequency domain on a scale of 9. db5 was chosen as the wavelet base, and the decomposed high frequency noise in layers 1 and 2 was set to 0. The wavelet coefficients in layers 3-9 were thresholded by a soft threshold formula. Using the wavelet transformed data to train the neural network, the optimal parametric model was eventually found to be almost unchanged in accuracy and F1-score from before after threshold tuning. Our team analyzed that the core features of the ECG data could be extracted effectively by avoiding noise through the multi-scale convolutional feature extraction method. Therefore, wavelet transform and other pre-processing methods can be omitted in this model, which will save the computational costs associated with data pre-processing.

4.2 Model development

We divided all datasets into a training set, test set, and validation set by 7:2:1 after randomly breaking them up, and used the k-fold cross-validation method to evaluate the model performance. Since the k-fold cross-validation method reallocates the training and validation sets, we set to guide the parameter (learning rate) adjustment by the accuracy of the validation set after each training cycle, and evaluate the model performance by the test set after every two training cycles. The learning rate is reduced when the F1-score of the validation set decreases during the training process. Training is stopped when the model's ACC score, F1-

score for the test set, does not improve in two consecutive cycles.

ECGResNet uses the Flooding training method proposed by Takashi et al., [25] in ICML 2020 in the training process, setting a threshold value. When the training loss value is greater than this threshold a normal gradient descent is performed, and when the training loss value is less than this threshold a gradient up operation is performed in reverse. This operation keeps the training loss value around a threshold and allows the model to be trained continuously. The core formula of the method is $\tilde{J}(\theta) = |J(\theta) - b| + b$, where b is the set threshold and J is the training loss. The flooding method folds the part below the threshold above the threshold, thus making the whole loss curve flatter and the generalization ability is somewhat improved. The method obtained good results in the training of our ECG classification model and complemented with the k-fold cross-validation method, effectively solving the overfitting problem caused by the k-fold cross-validation method due to the training set being partially repeated as the training set and the validation set being trained several times. The final validation set accuracy was improved by 1.32% and the F1-score was improved by 0.68%.

4.3 Threshold optimization

In the model development process, if we were to perform the threshold optimization operation as a single classification problem, we would need to perform threshold optimization for each of the 8 disease classifications. The search for the optimal threshold in an 8-dimensional space would be computationally time-consuming. We retreat to the second-best and optimize each threshold in the range, in steps of 0.01 with other thresholds fixed, based on the assumption that each disease is independent of the other.

After determining the threshold value by the above method, the threshold value in the Flooding method was optimized in steps of 0.001 in the range. The final experiment yielded optimal training results when the threshold value was set to 0.01634.

4.4 Scoring metrics

In the construction and development of the model, based on the assumption above that the diseases are independent of each other, we discuss each disease separately. The prediction of having a disease with the same outcome as the actual outcome is noted as TP. The accuracy of the diagnosis of having a disease, and not having a disease, is represented by the accuracy ACC value shown in (4). The summed average of the accuracy and recall is represented by the F1-score shown in (5). The model performance was tested by accuracy and F1-score.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

4.5 Classification performance

During the research and development phase of the models, the following four models were deployed and the accuracy ACC and F1-score were calculated for the models, as shown in [Table 1](#).

Table 1. Model experimental results

Methods	ACC	F1
A) Improved ResNet101+gender,age	0.930	0.755
B) Improved ResNet50 _{single conv} +gender,age	0.945	0.819
C) Improved ResNet50_{Multi conv}+gender,age	0.950	0.841
D) Improved ResNet50 _{Multi conv}	0.935	0.829

The experimental results are in line with the expected vision. The ResNet network structure is feasible in the detection of cardiovascular diseases. Multiscale convolutional extraction of features + ResNet50 (SE module) + gender and age are the optimal solutions for the classification of 8 major cardiovascular diseases with an accuracy of 0.950 on the test set and an F1-score of 0.841.

The cardiovascular disease classification task is usually labeled with a 1 for having the disease and a 0 for not having the disease. The sample of each disease category represents only a small proportion of the total data set. The higher the number of samples classified, the more unbalanced the proportion of positive and negative cases, and the lower the accuracy ACC reference value. Accordingly, the accuracy ACC and F1-score are of very different importance for models with different numbers of classifications. A comparison with the model proposed by our team, based on the number of models classified, is shown in **Tables 2** and **Table 3**.

Table 2. Accuracy (ACC) comparison results

Work	No of classes	ACC(%)
A) Acharya et al.[26]	5	94.03
B) Plawiak[27]	15	93.04
C) Melgani[28]	6	91.67
D) Kumar et al.[29]	3	92.16
E) ECGResNet	8	95.00

Table 3. F1-score comparison results

Work	No of classes	F1
A) Yingjing et al.[30]	27	0.538
B) Singstad et al.[31]	27	0.356
C) Zhaohan et al.[32]	4	0.864
D) Zhaohan et al.[33]	4	0.820
E) ECGResNet	8	0.841

Observing the results obtained, the model we have designed has better results in terms of both accuracy and F1-score. We believe that the combination of convolutional feature recognition and gender-age consideration through multiple scales is beneficial to the model. The input data used is a 12-lead ECG signal without noise reduction, which requires more

performance than a single-lead signal, but the model designed by our team still achieves satisfactory results.

5. Conclusion

We used bidirectional LSTM and Transformer architectures in our model and found that such models are not suitable for classification in ECG signals and are very prone to gradient disappearance. The result is not even as good as a 30-layer CNN stack. We eventually abandoned our focus on time series in favor of ResNet to effectively address the gradient disappearance, gradient explosion, etc. associated with deeper networks. ResNet obtained better results in the experiment, validating our idea. The traditional attention mechanism commonly used in the CV field does not work well in ECG signal processing. 12-lead ECG data can be thought of as 12 channels of 1-dimensional data. What our team believes should be looked at is the relationship between the data channels, i.e., the relationship between the data in each lead. We used the SE module to deal with the relationships between channels, which is essential for our model. Regarding the noise reduction processing of ECG signals, noise reduction methods such as wavelet transform and Fourier transform, which are effective in CNN models, have a subtle effect on the model we have designed. It is speculated that this may be due to the use of ResNet50 as the base model, which has a greater depth and is able to avoid regular noise interference in discriminating core features. Our team still believes that noise still has a certain impact on the accuracy of the model. Through several experiments, we have determined to use a multi-scale convolution method with 8 and 16 convolution kernels respectively to replace the first layer of convolution with 7 convolution kernels in ResNet50 for feature extraction and obtained the desired accuracy improvement. ResNet101 and ResNet152 theoretically enable the model to achieve higher accuracy, but the training time is longer and the improvement in accuracy is limited. Our team has therefore continued to use the more cost-effective ResNet50 as the basic structure of the model. In the experiments, we did not experiment with more disease classifications as the dataset sample was mainly focused on 8 diseases.

There are still many problems that need to be solved. The dataset we used is relatively homogeneous. The number of diseases that can be diagnosed by the model is small. No detailed generalizability researches of the model are carried out. In the future, we will add different datasets to test the model for validation. The parameters will be optimized using a more optimal collective tuning approach[34]. We will also try to apply the model to the diagnosis of other diseases, such as lung cancers[35]. We will also build confusion matrices and optimize the parameters separately for each disease to get better result.

Acknowledgement

This work is partially supported by the National Natural Science Foundation of China (Grants No 61702274), PAPD, NUIST Students' Platform for Innovation and Entrepreneurship Training Program (No 202110300085Y), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_1005).

References

- [1] L. Sun, Q. Yu, D. Peng, S. Subramani and X. Wang, "Fogmed: a fog-based framework for disease prognosis based medical sensor data streams," *Computers, Materials & Continua*, vol. 66, no.1, pp. 603–619, 2021. [Article \(CrossRef Link\)](#)
- [2] L. Sun, Y.L. Wang, Z.G. Qu, N.N. Xiong, "BeatClass: A Sustainable ECG Classification System in IoT-based eHealth," *IEEE Internet of Things Journal*, Aug. 2021. [Article \(CrossRef Link\)](#)
- [3] Z. G. Qu, Z. W. Cheng, W. J. Liu, X. J. Wang, "A novel quantum image steganography algorithm based on exploiting modification direction," *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 7981-8001, 2019. [Article \(CrossRef Link\)](#)
- [4] Z. G. Qu, Y. M. Huang and M. Zheng, "A novel coherence-based quantum steganalysis protocol," *Quantum Information Processing*, vol.19, no. 362, pp. 1-19, 2020. [Article \(CrossRef Link\)](#)
- [5] P. Lu, "A Position Self-Adaptive Method to Detect Fake Access Points," *Journal of Quantum Computing*, vol. 2, no. 2, pp. 119–127, 2020. [Article \(CrossRef Link\)](#)
- [6] World Health Organization. Cardiovascular diseases (CVDs)[EB/OL] (2017-05-17)[2019-10-13]. [Online]. Available: [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- [7] Yusuf, S., et al., "Global Burden of Cardiovascular Diseases," *Circulation*, 104(22), 2746-2753, 2001. [Article \(CrossRef Link\)](#)
- [8] Okada, M., "A digital filter for the QRS complex detection," *IEEE transactions on bio-medical engineering*, BME-26(12), 700-703, 1979. [Article \(CrossRef Link\)](#)
- [9] Ahlstrom, M. L., and W. J. Tompkins, "Digital filters for real-time ECG signal processing using microprocessors," *IEEE transactions on bio-medical engineering*, BME-32(9), 708-713, 1985. [Article \(CrossRef Link\)](#)
- [10] Rajpurkar P, Hannun A Y, Haghpanahi M, et al., "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," 2017.
- [11] Andersen RS, Peimankar A, Puthusserypady S., "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, 115, 465-473, 2019. [Article \(CrossRef Link\)](#)
- [12] Li Z, Zhou DS, Wan L, et al., "Heartbeat classification using deep residual convolutional neural network from 2 – lead electrocardiogram," *Journal of Electrocardiology*, 58, 105-112, 2020. [Article \(CrossRef Link\)](#)
- [13] Cao X C, Yao B, Chen B Q., "Atrial fibrillation detection using an improved multi – scale decomposition enhanced residual convolutional neural network," *IEEE Access*, 7, 89152-89161, 2019. [Article \(CrossRef Link\)](#)
- [14] Clifford, Gari D., et al., "AF classification from a short single lead ECG recording: The PhysioNet/Computing in cardiology challenge 2017," in *Proc. of 2017 Computing in Cardiology (CinC)*, IEEE, 2017. [Article \(CrossRef Link\)](#)
- [15] Mathews S M, Kambhamettu C, Barner K E., "A novel application of deep learning for single-lead ECG classification," *Computers in Biology and Medicine*, 99, 53-62, 2018. [Article \(CrossRef Link\)](#)
- [16] Vaswani, Ashish, et al., "Attention is all you need," *Advances in neural information processing systems*, 2017.
- [17] He, Kaiming, et al., "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2016.
- [18] Jie H, Li S, Gang S, et al., "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011-2023, 2020. [Article \(CrossRef Link\)](#)
- [19] Lin TY, Dollar P, Girshick R, et al., "Feature Pyramid Networks for Object Detection," in *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, 2017. [Article \(CrossRef Link\)](#)
- [20] Darmawahyuni A, Nurmaini S, Sukemi, et al., "Deep learning with a recurrent network structure in the sequence modeling of imbalanced data for ECG-rhythm classifier," *Algorithms*, 12(6),1-12, 2019. [Article \(CrossRef Link\)](#)

- [21] Rajpurkar, Pranav, et al., "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017. [Article \(CrossRef Link\)](#)
- [22] Sun L., Zhong Z., Qu Z. and Xiong N., "PerAE: An Effective Personalized AutoEncoder for ECG-based Biometric in Augmented Reality System," *IEEE Journal of Biomedical and Health Informatics*, Jan. 2022. [Article \(CrossRef Link\)](#)
- [23] Zhao, Zhibin, et al., "Adaptive lead weighted ResNet trained with different duration signals for classifying 12-lead ECGs," in *Proc. of 2020 Computing in Cardiology*, IEEE, 2020. [Article \(CrossRef Link\)](#)
- [24] Fujino, Akinori, Hideki Isozaki, and Jun Suzuki, "Multi-label text categorization with model combination based on f1-score maximization," in *Proc. of the Third International Joint Conference on Natural Language Processing: Volume-II*, 2008.
- [25] Ishida, Takashi, et al., "Do we need zero training loss after achieving zero training error?," *arXiv preprint arXiv:2002.08709*, 2020. [Article \(CrossRef Link\)](#)
- [26] Acharya UR, Fujita H, Lih O S, et al., "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *INFORMATION SCIENCES*, 405, 81-90, 2017. [Article \(CrossRef Link\)](#)
- [27] P Pławiak, "Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals," *Swarm and Evolutionary Computation*, 39, 192-208, 2018. [Article \(CrossRef Link\)](#)
- [28] Melgani F, Bazi Y, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *IEEE transactions on information technology in biomedicine*, 12(5), 667-677, 2008. [Article \(CrossRef Link\)](#)
- [29] Kumar RG, Kumaraswamy YS, "Investigating cardiac arrhythmia in ECG using random forest classification," *International Journal of Computer Applications*, 37(4), 31-34, 2012. [Article \(CrossRef Link\)](#)
- [30] Feng, Yingjing, and Edward Vigmond, "Deep Multi-Label Multi-Instance Classification on 12-Lead ECG," in *Proc. of 2020 Computing in Cardiology*, IEEE, 2020. [Article \(CrossRef Link\)](#)
- [31] Singstad, Bjørn-Jostein, and Christian Tronstad, "Convolutional Neural Network and Rule-Based Algorithms for Classifying 12-lead ECGs," in *Proc. of 2020 Computing in Cardiology*, IEEE, 2020. [Article \(CrossRef Link\)](#)
- [32] Xiong, Zhaohan, et al., "ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network," *Physiological measurement*, 39(9), 094006, 2018. [Article \(CrossRef Link\)](#)
- [33] Xiong, Zhaohan, Martin K. Stiles, and Jichao Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *Proc. of 2017 Computing in Cardiology (CinC)*, IEEE, 2017. [Article \(CrossRef Link\)](#)
- [34] Wenjun Tan, Peifang Huang, Xiaoshuo Li, Genqiang Ren, Yufei Chen, Jinzhu Yang, "Analysis of Segmentation of Lung Parenchyma Based on Deep Learning Methods," *Journal of X-Ray Science and Technology*, 2021(1), 945-959, 2021.
- [35] Wenjun Tan, Luyu Zhou, Xiaoshuo Li, Xiaoyu Yang, Yufei Chen, Jinzhu Yang, "Automated vessel segmentation in lung CT and CTA images via deep neural networks," *Journal of X-Ray Science and Technology*, 29(6), 1123-1137, 2021. [Article \(CrossRef Link\)](#)



Junke Zhu Mr Junke Zhu is a student in the School of Computer and Software, Nanjing University of Information Science and Technology, China. He is interested in computer vision and intelligent healthcare. He has some basic knowledge of machine learning and deep learning, and is able to modify and write artificial neural networks independently.



Le Sun Dr. Le Sun is a Professor in the School of Computer and Software, Nanjing University of Information Science and Technology, China. She received her PhD degree in 2016, from Victoria University, Australia, and received the Chancellor Citation Best PhD Award. In 2018, she was a visiting scholar of Michigan State University. She has published over 30 papers on high quality conferences and journals, e.g., IEEE Transactions on Services Computing, ACM Transactions on Internet Technology, IEEE Access, and Future Generation Computer Systems. Her research interests include medical data stream mining, services computing, and fog computing.



Yilin Wang Miss Yilin Wang is a master student in the School of Computer and Software, Nanjing University of Information Science and Technology, China. She received her bachelor degree from Nanjing University of Information Science and Technology, China, in 2019. Her research interests include smart medical and data mining.



Sudha Subramani Dr. Sudha Subramani is a Lecturer in Information Technology Discipline, Victoria University. She received B.E and M.E degree in computer science from Anna University, India in 2010 and 2012, respectively. She completed Ph.D. degree in College of Engineering and Science, Victoria University (VU), Australia in 2019 and she was a recipient of VU International Postgraduate Research Scholarship during her PhD. She was worked as Research Ambassador @VU. She has done summer internship in Robert Bosch, Japan, from 2016 to 2017, and worked on a project to develop disease assistance tool for disease management in tomato green houses. Her research interests include Artificial Intelligence, Social Media Data Analytics, Machine Learning, Deep Learning, and Natural Language Processing. She has published her research works on high quality conferences and journals like IEEE Access, BMC, HIS and WISE. She received the Gold Medal for her academic excellence in Masters program. She was awarded Oracle cloud credits of US 100,000\$ in 2020 for recognition of her research in “AIforGood” and in further collaboration with Oracle Research.



Dandan Peng Miss Dandan Peng is a doctor in the School of Computer Science and Network Engineering, Guangzhou University, China. She received her master degree in Nanjing University of Information Science and Technology, China, in 2021. Her research interests include Wise Medical, such as human physiological signal processing.



Shangwe Charmant Nicolas Mr. SHANGWE Charmant Nicolas is a master's student in the school of computer and software, Nanjing University of Information Science and Technology, China. He received his Bachelor's degree from Nanjing University of Information Science and Technology, China, in 2021. His research interest is Medical Data Mining.