Research Article

ECG Signal Detection and Classification of Heart Rhythm Diseases Based on ResNet and LSTM

Qiyang Xie,[1,2] **Xingrui Wang,**[1] **Hongyu Sun,**[1] **Yongtao Zhang,**[3] **and Xiang Lu**[1]

[1]*College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao 266590, China*
[2]*School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China*
[3]*Department of Cardiology, Affiliated Hospital of Qingdao University, Qingdao, 266000 Shandong, China*

Correspondence should be addressed to Xiang Lu; sdustlx@sdust.edu.cn

Arrhythmia is one of the most threatening diseases in all kinds of cardiovascular diseases. It is important to achieve efficient and accurate automatic detection of arrhythmias for clinical diagnosis and treatment of cardiovascular diseases. Based on previous research on electrocardiogram (ECG) automatic detection and classification algorithm, this paper uses the ResNet34 network to learn the morphological characteristics of ECG signals and get the significant information of signals, then passes into a three-layer stacked long-term and short-term memory network to get the context dependency of the features. Finally, four classification tasks are implemented on the PhysioNet Challenge 2017 test dataset by using the softmax function. The activation function is changed from the ReLu function to the mish function in this model. Negative information of ECG signals is considered in the training process, which makes the model have more stable and accurate classification ability. In addition, this paper calculates and compares the average information entropy of correctly classified samples and incorrectly classified samples in the test set. Moreover, it eliminates the impact of obvious signal abnormalities (redundancy or loss) on the model classification results, to more comprehensively and accurately explain the classification effect and performance of the model. After eliminating the possibility of abnormal signal, the ResNet34-LSTM3 model obtained an average F1 score of 0.861 and an average area under the receiver operating characteristic curve (ROC) of 0.972 on the test dataset, which indicates that the model can effectively extract the characteristics of ECG signals and diagnose arrhythmia diseases. Comparing the results of the ResNet34 model and ResNet-18 model on the same test dataset, we can see that the improved model in this paper has a better classification and recognition effect on ECG signals as a whole, which can identify atrial fibrillation diseases more effectively.

1. Introduction

With the increasing pressure on people's lives and work, cardiovascular disease has gradually become one of the important diseases threatening human life and health. According to the report of the World Health Organization, the mortality of cardiovascular disease ranks first among all kinds of diseases, accounting for 33.3% of other diseases. Arrhythmia is a kind of cardiovascular disease with a high incidence rate and high risk in all cardiovascular diseases. Atrial fibrillation (AF) is the most common arrhythmia disease. The clinical manifestations of patients are atrial arrhythmia or ineffective contractions. These diseases often occur in the elderly population and have a high incidence rate and long course. It is easy to cause heart failure, stroke, and other complications, which pose a serious threat to the safety of patients. Therefore, early and accurate detection of this kind of arrhythmia is an important challenge in clinical work. At present, the main tool for arrhythmia diagnosis is the electrocardiogram (ECG). By analyzing the ECG signal of patients, medical workers can make an accurate diagnosis of different types of arrhythmias. However, this kind of manual detection method relying on the clinical experience and a lot of professional knowledge of medical workers is often prone to make mistakes [1], and it also needs to invest a lot of manpower and energy. With the continuous development and maturity

of computer technology and electronic information technology, the task of using a computer to analyze ECG signals to realize automatic detection of arrhythmia has become a research hotspot at this stage, which can provide a more effective and reliable diagnosis basis for medical workers, thereby alleviating the investment in human resources [2].

The existing ECG classification algorithms usually include signal preprocessing, such as wavelet transform and manual feature extraction, but the amount of computation will increase the delay of the real-time classification system. In recent years, deep learning algorithm with their advantages of automatic learning features is increasingly used in the field of health care, such as medical image recognition and segmentation, time series data monitoring, and analysis. At present, the outstanding algorithm can establish an end-to-end DNN network to learn the characteristics of ECG records by using the extensive digital characteristics of ECG data, which saves a lot of signal preprocessing steps. Because the performance of DNN increases with the amount of training data, this method can make good use of the extensive digitization of ECG data.

The rest of this paper is organized as follows. The second section reviews the related research. Datasets and methods are described in the third section. The fourth section introduces and analyzes the experimental results. The fifth section summarizes the advantages and disadvantages of this method and puts forward the prospects.

2. Related Work

The common classification task of automatic detection of ECG signals usually has three steps, which are preprocessing, signal, feature extraction, and identification classification [3]. Since ECG signals are acquired using an ECG acquisition recorder, the original signal would be mixed with several noise and invalid signals. In general, low-pass filters, wavelet transform, and other relatively classical denoising methods are used in the preprocessing step. After signal preprocessing, feature extraction of the signal is performed. The traditional feature extraction methods use the discrete Fourier transform or wavelet transform to extract the morphological features of time series signals [4, 5], such as slope, amplitude, peaks, interval, and other characteristic information, and compose the feature vector addition to all types of traditional machine learning algorithms, such as principal component analysis and independent component analysis. More efficient, reliable, and compact eigenvectors can be obtained from ECG signals. These traditional feature extraction algorithms need to provide hand-crafted or feature-specific implications. However, the selection and combination of features often require expertise, and the selection process is time consuming [6]. With the development of deep learning theory, researchers worldwide began to use deep learning algorithms to automatically extract features of interest from data.

In a deep learning-based arrhythmia detection study, Kiranyaz et al. [7] developed a convolutional neural network (CNN) classification algorithm based on one-dimensional convolution for the corresponding disease class of ECG,

which can accomplish the basic classification tasks but has low sensitivity for arrhythmia classification of sveb type. Rajpurkar et al. [8] proposed a convolutional neural network algorithm with residual structure, which utilized the electrocardiographic signal collected from a single-lead wearable device for the detection of arrhythmia information [9] and used the AlexNet network as input bispectral spectrum of ECG signal, and the experiment finally got an average accuracy of 91.3%.Mostayed et al. [10] proposed a recurrent neural network algorithm; they trained the 12 lead ECG signal inputs into a model composed of two bidirectional long short-term memory (LSTM) networks to detect pathologies in the signal. Yildirim [11] used wavelet transform to decompose the ECG signal into a wavelet sequence, then entered into a two-way LSTM model for training and classification, and obtained a recognition accuracy of 99.39% under ideal conditions. Subsequently, Saadatnejad et al. [12] proposed a lightweight feature automatic extraction method combining wavelet transform with LSTM network, which could realize continuous real-time classification of electrocardiographic signals. Feng et al. [13] proposed a 16-layer convolutional neural network and combined it with a long-term memory network to realize multichannel classification, which achieved 95.4% accuracy in classifying myocardial infarction disease in the PTB database.

In addition to the above deep learning algorithms that directly utilize the one-dimensional ECG data for training, literature [14] transformed three adjacent beats in the ECG signal into a two-dimensional coupling matrix, and this matrix obtained the correlation between signal beat and morphological information [15, 16] Jun et al. [17] converted each beat in the signal into a two-dimensional gray-scale image, which was then taken as input to a 2D convolutional neural network. Then, such 2D methods need to convert 1D cardioelectrical signals into 2D information, which also occupies harder disk space while increasing the computational cost. In conclusion, many existing algorithms suffer from complicated preprocessing processes [18, 19] and high time costs [17].

3. ECG Dataset Introduction and Resnet34-LSTM3 Classification and Detection Method

Based on the end-to-end network characteristics, this study tries to combine a 34-layer ResNet network (ResNet34) with three stacked LSTM networks (LSTM-3) in combination with previous experience. Moreover, this model does not need too complex procedures such as signal preprocessing and manual feature extraction, and it uses the ResNet34 network to learn the morphological features of electrocardiographic signals and acquire significant information of the signal (the features extracted by the network are mainly the deep-level abnormal waveform feature information contained in the F wave, P wave, and QRS complex in the ECG signal). The context dependence of features was then acquired utilizing a three-layer stacked LSTM network. Finally, a multiclassification task on the PhysioNet challenge 2017 (https://physionet.org/challenge/2017/) test dataset was implemented through the softmax function. This model

utilizes the max pooling layer, dropout layer, and batch normalization layer several times to optimize the calculations and improve the classification accuracy. At the same time, it is intended to change the activation function from the ReLu function to the Mish function so that the model takes into account the negative value information of the ECG signal in the training process, while making the model more stable. In addition, this paper uses the model to classify ECG signals and calculates and compares the average information entropy of correctly classified samples and incorrectly classified samples. It eliminates the impact of obvious signal abnormalities (redundancy or loss) on the model classification results later, to more comprehensively and accurately explain the classification effect and performance of the model.

3.1. Introduction to ECG Datasets. The dataset used in the experiment is from the PhysioNet Challenge 2017 Short Single-Lead ECG AF Classification Competition website. The training set contains 8528 single-lead ECG signal sampling records, ranging from 9 seconds to slightly more than 60 seconds; each ECG sample has a sampling frequency of 300 Hz and has been band-pass filtered by the AliveCor device. Each sample contains a mat file of the corresponding ECG and a hea file containing waveform information, while all ECG samples are classified by human cardiologists into four categories: normal (N), atrial fibrillation (A), other rhythm (O), and noise (~).

The data division of training set and test set is shown in Figure 1, wherein there are 8528 data in the training set and 852 data in the test set. More details of the training set are shown in Table 1, where SD stands for standard deviation and med stands for median. Figure 2 shows examples of ECG waveforms in four categories (lasting 20 seconds) from top to bottom, with normal rhythm-like normal (N), atrial fibrillation (A), other rhythm (O), and noise (~) from left to right.

3.2. ECG Data Preprocessing. To train the built deep learning model more efficiently, the sequence length of each input network needs to be fixed. For this reason, this study first traversed all ECG signal samples in the dataset, finding the largest sequence length and defined as max_length. On the other hand, because the majority of ECG signal sample points in the dataset are around 9000 (the sampling time is about 30 seconds), and a considerable number of samples are about 18000, so for samples with sampling points close to max_length/2, if the number of sample points is larger than 9000, only the first 9000 sample points of this sample will be taken. If the number of sample points for this sample is less than 9000, then the sample is null-filled so that its sequence length reaches 9000. Similarly, for samples with sampling points close to max_length, if the number of sampling points of this sample is greater than 18000, only the first 18000 sampling points of this sample will be taken. If the number of sampling points of this sample is less than 18000, then the sample will be zero-filled so that its sequence length reaches 18000. The ECG signal samples processed above are later referred to as normalized samples, and the process is shown in Figure 3.
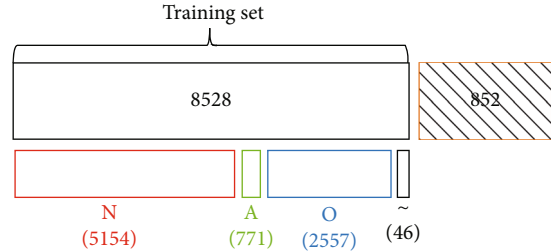

Figure 1: Data set composition diagram.

Table 1: Data table for the training set.

| Type | Number | Time length (s) | | | | |
		Mean	SD	Max	Med	Min
Normal (N)	5154	31.9	10.0	61.0	30	9.0
AF (A)	771	31.6	12.5	60	30	10.0
Other (O)	2557	34.1	11.8	60.9	30	9.1
Noisy (~)	46	27.1	9.0	60	30	10.2
Total	8528	32.5	10.9	61.0	30	9.0

Category vectors currently contain four different labels, namely N, A, O, and ~, and each ECG sample corresponds to a label identified by a human cardiologist. In this study, each normalized sample was divided into trunc_samp input sequences of the same length. The label specification of each input sequence is consistent with that of the original sample [20], where trunc_samp is defined as

$$trunc\_samp = \text{int} \left(max\_length / step \right). \tag{1}$$

In the experiment, the step is set to 256 and int is an integer operation.

The shape of the final input matrix is (normalized number of samples) × trunc_samp, step, 1), where 1 indicates that a single input sequence is one-dimensional and the final output matrix shape is (normalized number of samples) × trunc_samp, 4), of which 4 represents the four types of labels.

3.3. ResNet34-LSTM3 Classification and Detection Method

3.3.1. ResNet34-LSTM3 Model Structure. The ResNet34-LSTM3 network model consists of ResNet34 and LSTM-3. The ResNet34 network is used to extract the feature information of different levels of ECG signals, and the skip structure in the network is used to avoid network degradation such as gradient disappearance and training accuracy degradation due to too large network depth. LSTM-3 stacked network has the feature of capturing information related to the sequence in time. Therefore, the context dependencies of the features can be extracted by the input eigenvector of the ResNet34 network and output to the LSTM-3 network. Several maximum pooling layers, batch normalization layers, and dropout layer are arranged in the network to optimize the calculation and improve the classification accuracy. Considering the negative information of ECG signals, the Mish function is used as the activation function in the
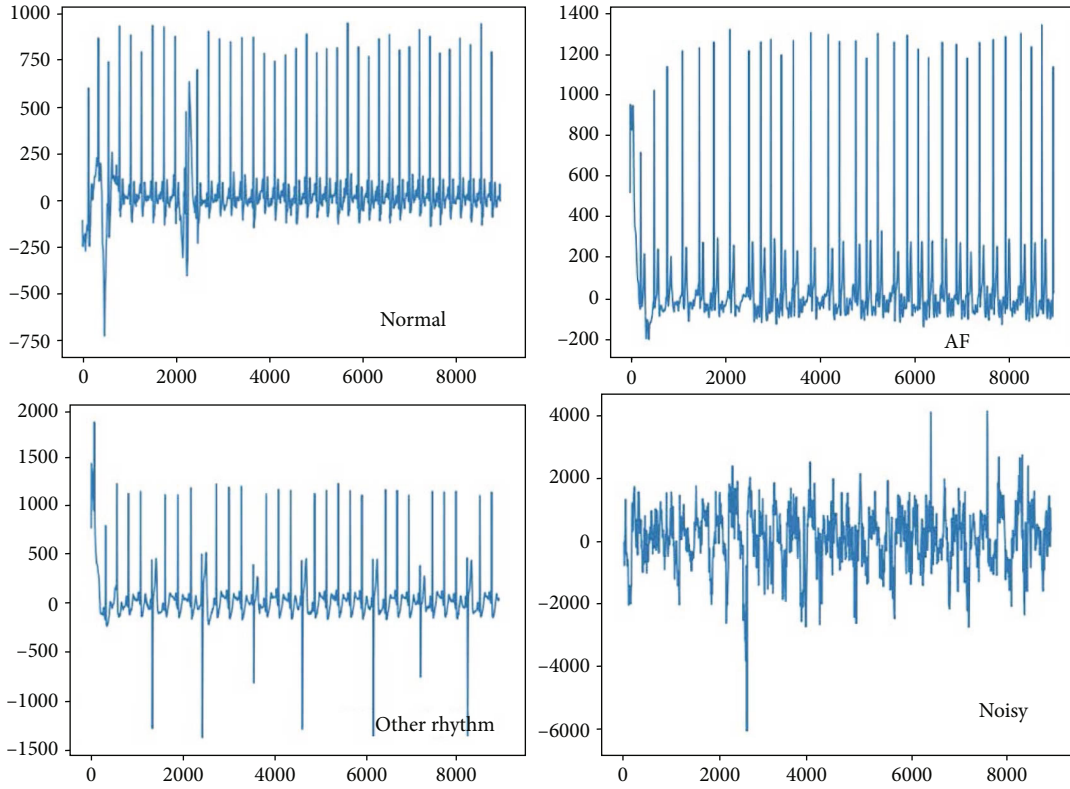
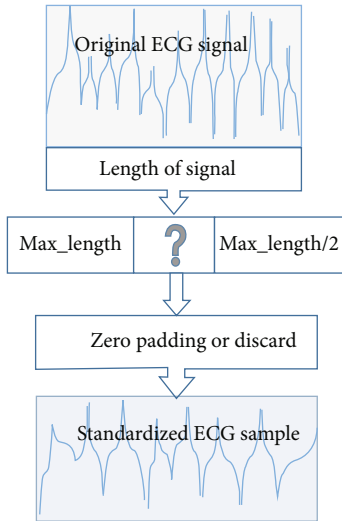Figure 2: Examples of ECG waveforms.

Figure 3: Normalized ECG sample generation diagram.

model. The network structure diagram of the ResNet34-LSTM3 model is shown in Figure 4.

3.3.2. ResNet34 Network Architecture. A general deep convolution network is one that stacks more network layers to better extract spatial features at different levels from the signal sequence or image provided. However, it has been found that deep CNN models are difficult to train. Because with the increase of network depth, the training accuracy will first rise and reach saturation and then continue to increase; the

network depth will lead to a decrease in accuracy, that is, the network begins to degenerate [21]. To overcome the degeneration problem, the deep residual network is used in this study to stabilize the training accuracy of the model while increasing the network depth. Compared with other types of deep CNN models such as VGGs and AlexNet, the deep residual network solves the network degradation problem by adding a skip structure, as shown in Figure 5.

The problem of deep network degradation is due to the existence of the nonlinear activation function ReLu, which causes a lot of important information loss for each activation layer from input to output, making this process almost irreversible [22]. The purpose of the residual structure is to enable the deep convolution network to have an equal mapping capability. In this way, when the network is deepened, at least the performance of the deep convolution network and the shallow network are balanced. It is difficult for existing neural networks to fit the potential identity mapping function $H(x) = x$, but if the network is designed as $H(x) = F(x) + x$ (as shown in Figure 5), that is, the identity mapping is directly part of the network in the residual structure, and the network is directly fitted to the residual function $F(x) = H(x) - x$, $F(x) = 0$, the identity mapping $H(x) = x$ can be obtained more quickly, thus solving the degeneration problem of deep convolution network [22].

At the same time, the output function of the residual structure is $H(x) = F(x) + x$, and the constant 1 in the derivative results of x from $(dH(x)/dx = dF(x)/dx) + 1$ and $H(x)$ can also alleviate the possible disappearance of gradients in the deep network when reverse propagation occurs.
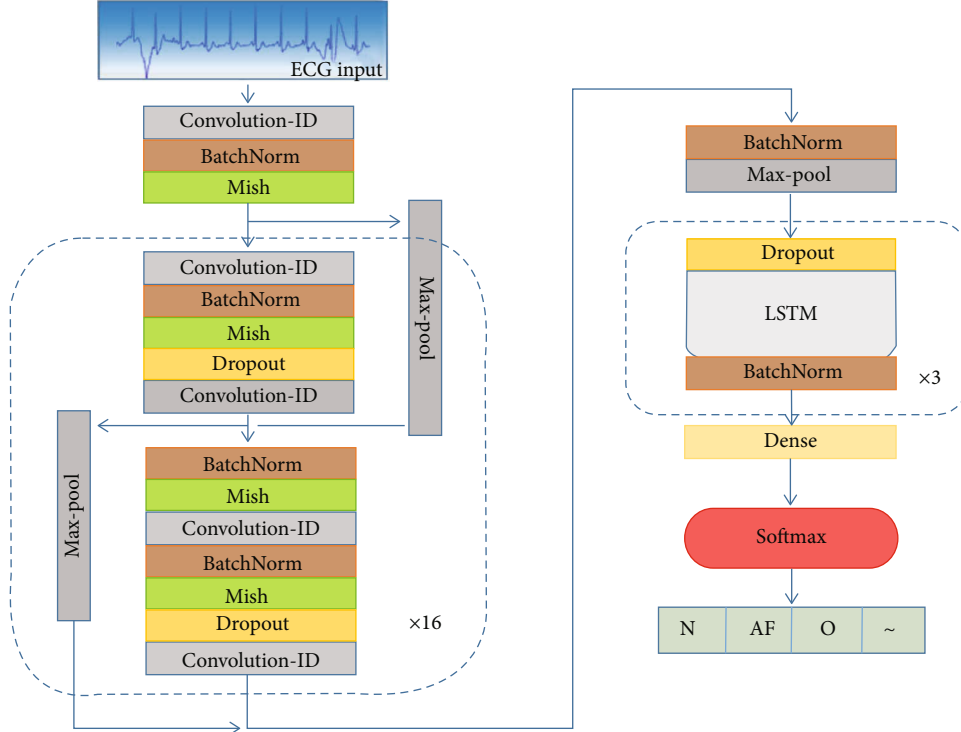
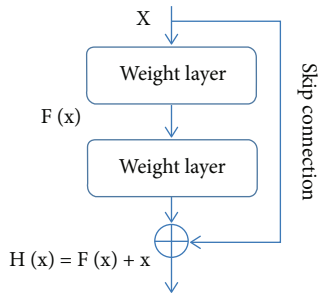FIGURE 4: ResNet34-LSTM3 network structure diagram.

FIGURE 5: Residual structure diagram.

This study uses ResNet34 to extract the characteristics of different levels of input ECG signals. As shown in Figure 4, the ResNet34 network is composed of the signal input layer, one-dimensional convolution layer, BN layer (batch normalization unit), activation layer, dropout layer, and maximum pooling layer as a whole. The convolution layer has the characteristics of weight sharing and local connectivity, which can be used to extract the local characteristics of ECG signals. The formula for calculating one-dimensional convolution is as follows:

$$c^l = \sum_{i=0}^{m} \left(w_i^l x_i^l \right) + b^l. \tag{2}$$

w^l and b^l are the weight and offset of l layer and m is the convolution kernel size.

The batch normalization layer normalizes the distribution of data features at each level, which guarantees that the input feature distribution has the same mean and variance and makes the change of model loss values and gradients more stable [23]. The BN calculation formulas are as follows:

$$\text{Input}: \beta = \{x_{1\cdots m}\},$$

$$\mu_\beta = \frac{1}{m} \sum_{i=1}^{m} x_i,$$

$$\sigma_\beta^2 = \frac{1}{m} \sum_{i=1}^{m} \left(x_i - \mu_\beta \right)^2,$$

$$\widehat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \varepsilon}},$$

$$y_i = \gamma \widehat{x}_i + \beta \overset{\text{def}}{=} BN_{\gamma,\beta}(x_i),$$

$$\text{Output}: \{y_i = BN_{\gamma,\beta}(x_i)\}.$$

$$\tag{3}$$

From the above formulas, the BN layer first calculates the mean μ_β and variance σ_β^2 of each minibatch data, then normalizes the data to mean 0 and variance 1 (where ε is to prevent variance from being zero). Finally, two parameters that can be learned (scaling parameter γ and offset parameter β) as output are used for linear change. According to that, some useful feature information is lost after the data is normalized. Therefore, the introduction of linear change will restore the model to a certain extent.

The activation layer can make the model fit nonlinearly and have the ability to classify. Many previous studies have used the Relu function (formula (4)) as the activation function. However, using the Relu function will lose negative information of ECG signals, resulting in a poor classification effect. Therefore, this paper flexibly uses functions as the Mish activation function (formula (5)). The two activation function curves are shown in Figure 6. From Figure 6, it is clear that the function has similar nonlinear ability as the Relu function, while retaining a small amount of negative information in the ECG signal, so that the classification performance of the network is better.

$$ReLu = \begin{cases} x, x > 0, \\ 0, x \le 0, \end{cases} \tag{4}$$

$$Mish = x \times \tan h(\ln (1 + e^x)). \tag{5}$$

To preserve the significant information of each layer of ECG signals and reduce the complexity of network calculation, a maximum pooled layer with a step of 1 and a core size of 2 is added to the network. In addition, the dropout layer is added to the network to randomly discard part of the information to prevent the model training from overfitting.

3.4. LSTM-3 Network Structure. A LSTM network is a time series model that can extract time domain characteristics from any sequence data [24]. Compared with recursive neural networks, LSTM can solve the problem of gradient disappearance in long-term sequence learning, thus improving the learning ability of models. The structure of the LSTM unit is shown in Figure 7.

The equations for calculating the internal parameters of LSTM cells are as follows:

$$f_t = \sigma\left(w_{fx}x_t + w_{fh}h_{t-1} + b_f\right), \tag{6}$$

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i), \tag{7}$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o), \tag{8}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tan h(w_{cx}x_t + w_{ch}h_{t-1} + b_c), \tag{9}$$

$$h_t = o_t \times \tan h(c_t). \tag{10}$$

In Equations (6)–(10), w is the weight parameter, b is the deviation, σ is the Sigmoid function, h_t is the hidden state of the current unit, and the subscripts of w and b represent the weights and deviations of three different gates, respectively. i_t, f_t, c_t, and o_t are input gates, forgetting gates, cell states, and output gates, respectively. The tan h is a hyperbolic tangent function.

As shown in Equation (6), the forgetting door controls the input of information from the previous unit. It determines how much information needs to be retained or transmitted to the next unit. The input door controls the input of new information from the outside. It determines how much new information should be used. The current unit state can be obtained by combining the output of the updated forgetting door with the input door as shown in Equation (9). The
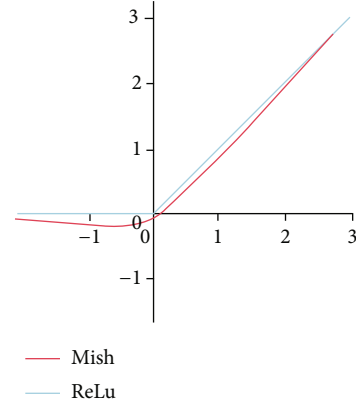

FIGURE 6: Mish function and ReLu function graph.

hidden state of the current cell is calculated from the cell output and the latest cell state.

Based on the time series advantages of LSTM networks, this study uses a three-layer stacked LSTM network after the ResNet34 network to extract context dependencies in ECG signal characteristics. Each LSTM network contains the same number of LSTM units, which is set to 256 in this paper. The schematic diagram of the single-layer LSTM network structure is shown in Figure 8.

In the LSTM-3 network, the output sequence of the previous LSTM network constitutes the input sequence of the next LSTM network, with one BN layer and dropout layer added between each two LSTM networks. Assuming the eigenvector of the output of the ResNet34 network is a, the learning process of the LSTM-3 network can be represented by the following:

$$\begin{aligned} h_t^1, c_t^1 &= LSTM^1\left(h_{t-1}^1, c_{t-1}^1, a_t\right), \\ h_t^2, c_t^2 &= LSTM^2\left(h_{t-1}^2, c_{t-1}^2, h_t^1\right), \\ h_t^3, c_t^3 &= LSTM^3\left(h_{t-1}^3, c_{t-1}^3, h_t^2\right). \end{aligned} \tag{11}$$

In the above formulas, LSTM represents an operation function of the LSTM layer, which is used to process the feature sequence, the sequence number $\{1, 2, 3\}$, representing the sequence number of three successively connected LSTM layers, and H and C are the hidden state and layer state components of the corresponding LSTM layer.

3.5. Network Output Layer Design. After the output of the LSTM-3 network, a fully connected layer with 1024 neurons is connected. Finally, the four classifications of the input ECG signal are implemented by the softmax function. The softmax formula is as follows:

$$P(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{4} e^{x_j}}. \tag{12}$$

$P(x_i)$ is the predicted probability distribution of x_i belonging to all possible classes. j is an accumulative variable, ranging from 1 to 4 (total number of categories).
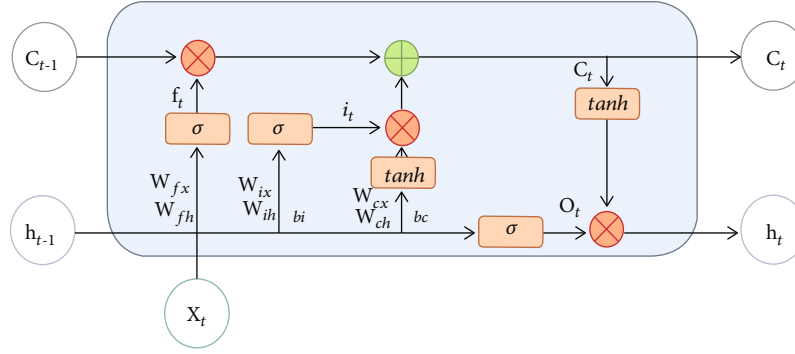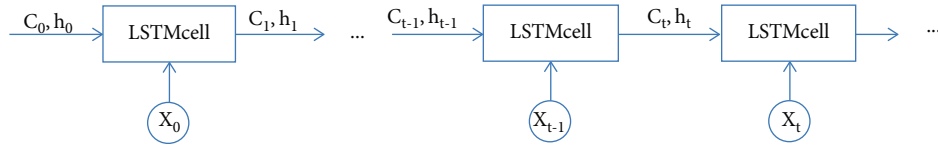
FIGURE 7: LSTM unit structure diagram.

FIGURE 8: Schematic diagram of single-layer LSTM network structure.

3.6. Information Entropy Verification. The concept of information entropy is used to describe the uncertainty of an information source. Shannon, the father of information theory, proposed in his paper that "any information has redundancy, and the size of redundancy is related to the occurrence probability or uncertainty of each symbol in the information." Shannon, with the help of the concept of thermodynamics, called the average amount of information after eliminating the redundancy in information as "information entropy." In the experiment, the sampling value of each ECG sample is uncertain, which can be measured according to its occurrence probability. If the probability of sampling value is large, the uncertainty is small and the amount of information provided is small; on the contrary, the uncertainty is large.

In the calculation of the average information entropy of ECG samples, it is assumed that n sampling values can appear in a certain ECG sample to transmit information: $U_1 \cdots U_i \cdots U_n$, The corresponding probability is $P_1 \cdots P_i \cdots P_n$. And generally, it can be considered that the occurrence of various sampling values is independent of each other. At this time, the uncertainty of the single sampling value of ECG signal sample is $-\log(P_i)$, and its average information entropy is E. The calculation formula is as follows:

$$H(U) = E[-\log(P_i)] = -\sum_{i=1}^{n} P_i \log(P_i). \tag{13}$$

After the trained Resnet34-LSTM3 model completes the classification task on the test set, this paper calculates the average information entropy of the correctly classified sample signal and the incorrectly classified sample signal, respectively. Then, the average information entropy of the two kinds of samples is compared. If the average information entropy of the correctly classified sample signals is

TABLE 2: Hardware equipment.

Hardware type	Hardware name
CPU	AMD Ryzen 5 5600 H with Radeon graphics
Memory	16.0 GB
GPU 0	NVIDIA GeForce RTX 3050Ti Laptop GPU
GPU 1	AMD Radeon (TM) graphics

significantly higher or lower than the latter, it shows that the misclassification of signal samples by the model may be caused by the anomaly of these sample signals themselves. If the average information entropy of the two types of sample signals is the same, it shows that the misclassification of signal samples by the model is caused by the factors of the model itself.

This paper calculates and compares the average information entropy of sample signals, which can eliminate the impact of obvious signal anomalies (redundancy or loss) on the classification results of the model, to more comprehensively and accurately explain the classification effect and performance of the model.

4. Training and Results

The model was trained and evaluated using the training and test datasets provided by the official website of the PhysioNet Challenge 2017 Short Single-Lead ECG AF Classification Competition.

The development IDE used in this study was the PyCharm Professional Edition, and the compilation environment was Python 3.6. The models were trained and tested using the Kerns 2.3.1 framework with TensorFlow 2.0.0 backend. The hardware equipment based on the whole experiment process is shown in Table 2.
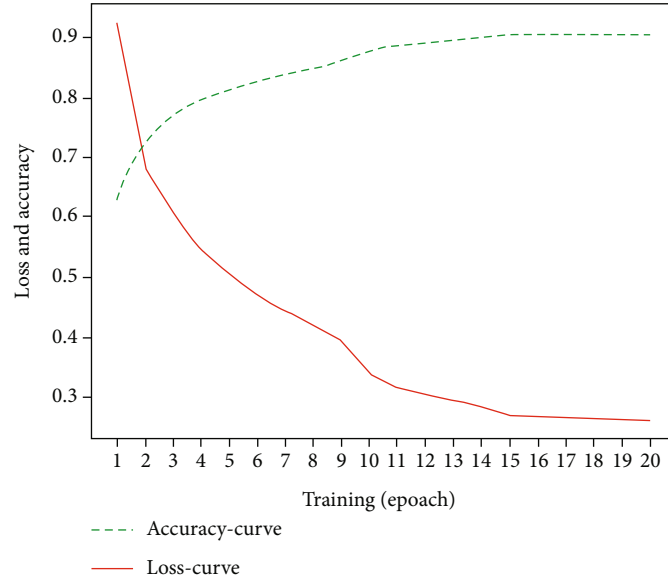
FIGURE 9: Loss curve and accuracy curve of model training.

4.1. Model Training. The maximum epoch of network training is set to 50, and the batch size is 32. Using the Adam optimizer to update the network weight, the initial learning rate is set to 0.001. In the training process, if the accuracy of the model on the verification set is not increased by two consecutive epochs, the learning rate is reduced to 10 times of the original, and the minimum learning rate is set to 10^{-6}. The initial length of the one-dimensional convolution kernel is set to 16, the initial number of convolution kernels in each convolution layer is set to 32, and the number of convolution kernels is doubled after every two convolution layers. The convolution kernel weight is initialized by the normal distribution.

To prevent the model from overfitting in the training process, if the various indicators of the model are not optimized after 8 epochs, the training of the model will be stopped in advance. The loss value curve and accuracy curve of the model in the training process are shown in Figure 9. It can be seen from Figure 9 that the loss value curve and accuracy curve of the model have converged before 20 epochs training.

4.2. Assessment Results. After the model training is completed, the average information entropy of the model was calculated after classification on the test set, as shown in Table 3. The average information entropy of correctly classified sample signal and incorrectly classified sample signal is 8.9088 and 8.9057, respectively. Therefore, the sample signals participating in the model classification test are not obviously abnormal.

After ensuring that there is no obvious abnormality in the sample signal, the overall average precision, recall, F1 score, specificity, and negative predictive value (NPV) of the model on the test set were calculated, as shown in Table 4.

TABLE 3: Average information entropy record table.

Average information entropy of correctly classified sample signals	Average information entropy of incorrectly classified sample signals
8.9088	8.9057

TABLE 4: Classification index score table.

	Precision	Recall	F1 score	Specificity	NPV	Support
A	0.839	0.739	0.786	0.994	0.988	72
N	0.909	0.873	0.890	0.962	0.945	486
O	0.776	0.805	0.790	0.952	0.960	259
~	0.967	0.989	0.978	0.968	0.990	35
Average	0.873	0.852	0.861	0.969	0.971	

It can be seen from the table that the overall average precision, recall, F1 score, specificity, and NPV of ResNet34-LSTM3 classification detection method in the test set are 87.3%, 85.2%, 86.1%, 96.9%, and 97.1%, respectively.

The F1 score values of the model for noise (~), normal rhythm (N), atrial fibrillation (A), and other rhythm (O) are 0.978, 0.890, 0.786, and 0.790, respectively. It shows that the model can recognize the noise signals in the test set well and eliminate the interference of noise in the sample under the training data set with very limited sample size. At the same time, it also has a good classification effect for normal rhythm (N), atrial fibrillation (A), and other rhythm (O). In addition, the specificity scores of the model for all kinds of samples are higher than 0.95, which indicates that the model has good recognition ability for negative cases.

To better evaluate the classification ability of the model for ECG signals, the ResNet34-LSTM3 model in this paper is compared with the ResNet34 model and ResNet18 model,
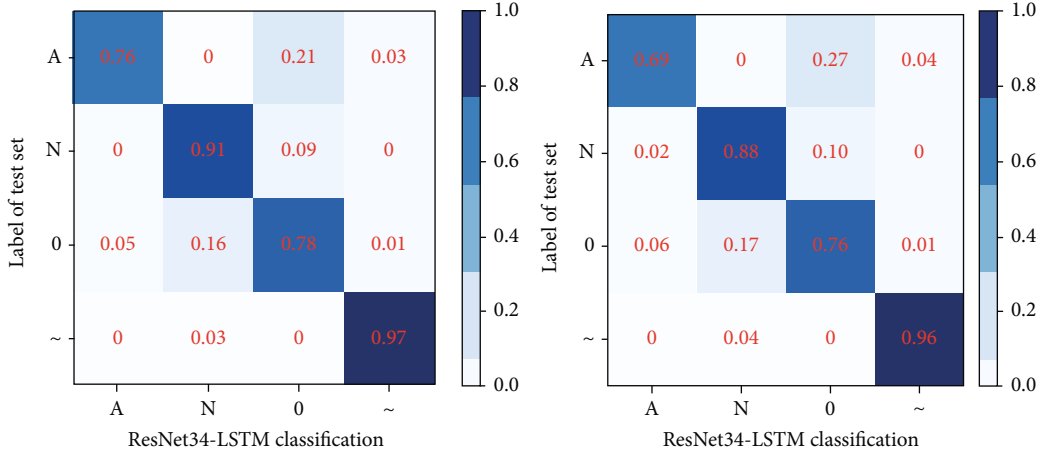
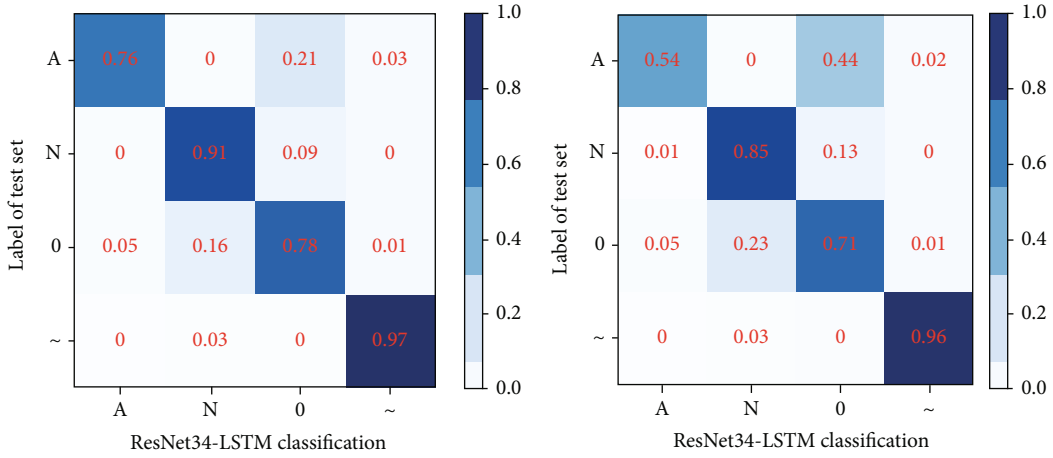FIGURE 10: Comparison of the ResNet34-LSTM3 and ResNet34 confusion matrix.

FIGURE 11: Comparison of the ResNet34-LSTM3 and ResNet18 confusion matrix.

which has better classification ability for ECG signals. The confusion matrix comparison of three different models is obtained as shown in Figures 10 and 11. At the same time, the F1 score and AUC value of the three models for four kinds of heart rate classification are obtained as shown in Tables 5 and 6. The experimental data obtained in the two tables are based on the test data set provided by the PhysioNet Challenge 2017 Short Single-Lead ECG AF classification competition official website.

According to the data in Tables 5 and 6, the overall F1 score average and AUC average of ResNet34-LSTM3 model in the test set are 0.861 and 0.972, respectively, both of which are higher than the other two classification models. Therefore, it can be shown that the ResNet34-LSTM3 model has better classification and recognition effect on ECG signals as a whole. The F1 score and AUC value of the ResNet34-LSTM3 model for atrial fibrillation (A) were 0.786 and 0.967, respectively, which are higher than those of ResNet34 model (0.777 and 0.959), indicating that the improved model can better identify atrial fibrillation (A) diseases.

TABLE 5: F1 scores of three models.

Type	F1 (ResNet34-LSTM3)	F1 (ResNet34)	F1 (ResNet18)
A	0.786	0.777	0.713
N	0.890	0.892	0.885
O	0.790	0.789	0.753
~	0.978	0.978	0.982
Average	0.861	0.859	0.833

The F1 scores and AUC values of the ResNet34-LSTM3 model and ResNet34 model for normal rhythm (N), other rhythm (O), and noise (~) are the same, which indicates that the improved model can still recognize other rhythm signal samples well, and there is no decline in classification ability. The overall F1 score and AUC of ResNet34-LSTM3 model in the test set are significantly higher than those of the ResNet18 model, which shows that the ResNet34-LSTM3 model in this paper is significantly better than the ResNet18 model in the classification ability of ECG signal.

Table 6: AUC scores of the three models.

Type	AUC (ResNet34-LSTM3)	AUC (ResNet34)	AUC (ResNet18)
A	0.967 (0.962-0.972)	0.959 (0.954-0.965)	0.954 (0.949-0.959)
N	0.979 (0.978-0.981)	0.979 (0.978-0.981)	0.965 (0.963-0.967)
O	0.945 (0.942-0.949)	0.946 (0.943-0.949)	0.887 (0.883-0.892)
~	0.998 (0.998-0.999)	0.999 (0.998-0.999)	0.998 (0.997-0.998)
Average	0.972	0.971	0.951

5. Conclusions

In this paper, based on the ResNet34 network, a three-layer stacked long-term and short-term memory networks are added, and the Mish function is used as the activation function. The final improved model can obtain the context dependence of the feature and retain the negative information in the ECG signal. The average F1 score of 0.861 and the average AUC value of 0.972 are obtained by the improved ResNet34-LSTM3 model on the PhysioNet challenge 2017 test dataset, which shows that the model can effectively extract the characteristics of ECG signals and diagnose arrhythmia diseases. Comparing the evaluation results of the previous ResNet34 model and ResNet18 model on the same test dataset, it can be seen that the improved model has a better classification and recognition effect on ECG signals as a whole, and it can more effectively identify arrhythmias such as atrial fibrillation, which will provide a more effective and reliable diagnostic basis for medical workers.

There are some important limitations in this study. The input dataset of the experiment is PhysioNet challenge 2017 Short Single-Lead ECG signal, which provides a limited signal compared with the standard 12 lead ECG signal. Therefore, whether the ResNet34-LSTM3 model classification performance is better in the 12 lead ECG signals remains to be determined. In addition, when the algorithm is used clinically, it may be limited by the duration of ECG signals, and the application of all kinds of algorithms, including the one presented algorithm, must eventually tailor specific ECG signal pretreatment methods for the target clinical application. Therefore, in the next stage of the study, we consider segmenting the signal to supplement the signal segment by copying other electrocardiogram signals in the same category in order to maximize the use of information. In the future, we will conduct experiments with more types of ECG data to prove the performance of our model.

In a word, the ResNet34-LSTM3 network model in this paper can distinguish the signals with different concentric laws in Short Single-Lead ECG signals, and its classification performance is also better than that of the predecessors in partial scores. If more tests are carried out in the clinical environment, this method may help medical workers improve the efficiency and accuracy of ECG clinical interpretation.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] Z. Golrizkhatami and A. Acan, "ECG classification using three-level fusion of different feature descriptors," *Expert Systems with Applications*, vol. 114, pp. 54–64, 2018.

[2] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," *Biomedical Signal Processing and Control*, vol. 57, article 101819, 2020.

[3] J. S. Huang, B. Q. Chen, N. Y. Zeng, X. C. Cao, and Y. Li, "Accurate classification of ECG arrhythmia using MOWPT enhanced fast compression deep learning networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, 2020.

[4] O. T. Inan, L. Giovangrandi, and G. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Transactions on Bio-Medical Engineering*, vol. 53, no. 12, pp. 2507–2515, 2006.

[5] Tran Hoai Linh, S. Osowski, and M. Stodolski, "On-line heart beat recognition using Hermite polynomials and neuro-fuzzy network," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 4, pp. 1224–1231, 2003.

[6] F. Murat, O. Yildirim, M. Talo, U. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review," *Computers in Biology and Medicine*, vol. 120, article 103726, 2020.

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

[8] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," 2017, https://arxiv.org/abs/1707.01836.

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

[10] A. Mostayed, J. Luo, X. Shu, and W. Wee, "Classification of 12-lead ECG signals with bi-directional LSTM network," 2018, https://arxiv.org/abs/1811.02090.

[11] Z. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Computers in Biology & Medicine*, vol. 96, pp. 189–202, 2018.

[12] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," *IEEE Journal of Biomedical & Health Informatics*, vol. 24, no. 2, pp. 515–523, 2020.

[13] K. Feng, X. Pi, H. Liu, and K. Sun, "Myocardial infarction classification based on convolutional neural network and recurrent neural network," *Applied Sciences*, vol. 9, no. 9, p. 1879, 2019.

[14] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018.

[15] H. Li, J. Liu, R. Liu, N. Xiong, K. Wu, and T. H. Kim, "A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis," *Sensors*, vol. 17, no. 8, article 1792, 2017.

[16] H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu, and N. Xiong, "Spatio-temporal vessel trajectory clustering based on data mapping and density," *IEEE Access*, vol. 6, pp. 58939–58954, 2018.

[17] T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y.-H. Kim, "ECG arrhythmia classification using a 2-D convolutional neural network," 2018, https://arxiv.org/abs/1804.06812.

[18] L. Dan, J. Zhang, Z. Qiang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Dalian, China, 2017.

[19] M. Jangra, S. K. Dhull, and K. K. Singh, "ECG arrhythmia classification using modified visual geometry group network (mVGGNet)," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 3, pp. 3151–3165, 2020.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems.*, vol. 60, pp. 84–90, 2017.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520, Salt Lake City, UT, USA, 2018.

[23] Y. Li, N. Wang, J. Shi, X. Hou, and J. Liu, "Adaptive batch normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109–117, 2018.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.