

Research Article

Multiscale Residual Network Based on Channel Spatial Attention Mechanism for Multilabel ECG Classification

Shuhong Wang,^{1,2} Runchuan Li ,^{1,2} Xu Wang,^{1,2} Shengya Shen,³ Bing Zhou,^{1,2} and Zongmin Wang ^{1,2}

¹School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China

²Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou 450000, China

³Foreign Languages Department, Zhengzhou University of Economics and Business, Zhengzhou 450000, China

Correspondence should be addressed to Runchuan Li; runchuanli@foxmail.com and Zongmin Wang; zmwang@ha.edu.cn

Received 29 December 2020; Revised 1 March 2021; Accepted 1 April 2021; Published 4 May 2021

Academic Editor: Chien Yu Chen

Copyright © 2021 Shuhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automatic classification of ECG is very important for early prevention and auxiliary diagnosis of cardiovascular disease patients. In recent years, many studies based on ECG have achieved good results, most of which are based on single-label problems; one record corresponds to one label. However, in actual clinical applications, an ECG record may contain multiple diseases at the same time. Therefore, it is very important to study the multilabel ECG classification. In this paper, a multiscale residual deep neural network CSA-MResNet model based on the channel spatial attention mechanism is proposed. Firstly, the residual network is integrated into a multiscale manner to obtain the characteristics of ECG data at different scales and then increase the channel spatial attention mechanism to better focus on more important channels and more important ECG data fragments. Finally, the model is used to classify multilabel in large databases. The experimental results on the multilabel CCDD show that the CSA-MResNet model has an average F1 score of 88.2% when the multilabel classification of 9 ECGs is performed. Compared with the benchmark model, the F1 score of CSA-MResNet in the multilabel ECG classification increased by up to 1.7%. And, in the model verification on another database HF-challenge, the final average F1 score is 85.8%. Compared with the state-of-the-art methods, CSA-MResNet can help cardiologists perform early-stage rapid screening of ECG and has a certain generalization performance, providing a feasible analysis method for multilabel ECG classification.

1. Introduction

The mortality rate caused by cardiovascular diseases is still increasing, which has caused widespread concern in the health sector [1]. The abnormal detection of cardiovascular disease is electrocardiography (ECG), which uses electrodes placed on the skin to record the electrical activity of the heart over a period of time [2]. Abnormal ECG waveforms can reflect certain pathologies [3]. The doctor gives a clinical diagnosis result through detailed evaluation and analysis of the patient's ECG. However, the large number of ECGs collected in homes and hospitals every day may hinder doctors from reviewing the data in detail [4]. With the advent of computers and artificial intelligence technologies, more and more scholars have applied computers in the field

of cardiovascular disease diagnosis, using artificial intelligence methods to predict certain diseases in order to reduce the workload of doctors.

At present, there are several categories of ECGs, such as static ECG, dynamic ECG, and ECG collected from wearable ECG equipment. Considering the monitoring time, ECG monitoring can be divided into long-term monitoring and short-term monitoring [5]. The dynamic ECG belongs to long-term monitoring, while the static ECG and the ECG obtained through the wearable ECG acquisition device are short-term monitoring. Taking into account the number of leads of the ECG, it is divided into single-lead and multilead. Most of the multileads used in hospitals are mainly 12-lead. The 12-lead ECG is more comprehensive than the 1-lead or 2-lead ECG and will cover more ECG waveform information

[6]. Therefore, for some complex and changeable ECG categories, many current artificial intelligence algorithm studies have fused the signals of all 12 leads to more comprehensive ECG information, which achieves accurate automatic ECG classification.

The clinical ECG reflects the patient's heart condition over a period of time. During this period of time, there may be many different categories of abnormalities in the ECG at the same time. For example, Figure 1 shows a patient suffering from atrial premature beats (APB), premature ventricular contraction (PVC), and complete right bundle branch block (CRBBB). In the clinical diagnosis report in real life, the doctor will conduct a detailed and comprehensive review of the patient's ECG, and any abnormal waveforms and rhythms that may occur will be marked in the patient's ECG diagnosis results. This illustrates that there may be more than one heartbeat category in the patient's diagnostic results; a piece of ECG data record may correspond to multiple labels at the same time.

However, the current artificial intelligence diagnosis ECG algorithms only judge that the segment of the ECG belongs to a certain category of significant abnormality and finally ignore other general abnormal information. But this information is also a response to the patient's heart condition, and doctors can judge whether there may be other hidden heart problems based on this information. It is therefore of interest to find an efficient artificial intelligence algorithm for multilabel ECG classification. The multilabel classification is widely used in the field of target detection [7, 8] and the multilabel ECG classification is still in the early exploration stage and has not been effectively applied.

To solve the problem of multilabel ECG classification, this work presents an end-to-end deep learning model CSA-MResNet, which combines a multiscale residual network with a convolutional attention mechanism to perform nine multilabel ECG classifications on the clinical acquired China Cardiovascular Disease Database (CCDD) [9]. The main contributions of this paper are as follows: (1) A multiscale residual network CSA-MResNet model based on channel spatial attention mechanism is proposed, which classifies ECG records by extracting ECG features of different scales of convolution kernels and using the channel spatial attention mechanism to pay more attention to abnormal ECG fragments; (2) the proposed method was verified in the CCDD, and the F1 score reached 88.2%, which achieved better performance compared with benchmark methods; (3) CSA-MResNet was also tested on the HF-challenge dataset with an average F1 score of 85.8%, which proves that the proposed method has some generalization.

The rest of the paper is organized as follows: Section 2 briefly introduces the related works, the method is described in Section 3, Section 4 gives the experiment and results, and finally, Section 5 summarizes the full text.

2. Related Work

A detailed review of the patient's ECG by the clinician, mainly looking at rhythm and waveform abnormalities, requires extensive experience and medical theory, but it is

time-consuming and laborious to produce a clinically experienced cardiologist. With the advent of artificial intelligence, many computer research scholars use artificial intelligence technology for abnormal judgment on ECG, which is able to greatly reduce the workload and intensity of physicians, but it is still always difficult to realize precise classification because of the diversity and individual variability of ECG data. In the past few years, research scholars have conducted many studies on ECG intelligent classification [10–12] and proposed various ECG intelligent classification methods.

In the early days when artificial intelligence was widely used in medical research, some researchers used traditional machine learning methods to classify ECG. Acharya et al. [13] first denoised the ECG signal, then extracted the nonlinear features, and finally used K-NearestNeighbor (KNN) classifier to classify normal and myocardial infarction, with an average accuracy of 98.80%. Alickovic et al. [14] used multiscale principal component analysis (MSPCA) to denoise the ECG signal and used an autoregressive (AR) model for feature extraction and examined different classifiers for comparison. The highest accuracy of the five arrhythmia categories in the MIT-BIH arrhythmia database reached 99.93%. Lin et al. [15] extracted the normalized RR interval and wavelet morphological features to classify the normal, atrial premature beats and premature ventricular contraction in the MIT-BIH arrhythmia database. Pandey et al. [16] used a set of support vector machines to classify heartbeat into four categories. Rajesh et al. [17] used intrinsic mode functions to get the final features and used the AdaBoost classifier to classify heartbeat. Although the above work achieves better results, it requires manual feature extraction, which wastes time and energy for a database with huge ECG data.

Deep learning has been widely applied to the study of ECG classification because of its characteristic ability to automatically extract data features that eliminates the steps of traditional machine learning manual extraction for automatic learning of large amounts of data. Awni et al. [4] proposed the use of an end-to-end deep neural network (DNN) to classify 12 ECG categories from single-lead electrocardiographic signals, with the mean F1 score for DNN of 0.837 exceeding the mean of 0.780 by general cardiologists. Acharya et al. [18] developed a 9-layer deep convolutional neural network that achieved F1 scores of 0.83 for the detection of five different categories of ECGs. Kamaleswaran et al. [19] identified a 13-layer convolutional neural network (CNN) model that can be used to detect four beat categories, normal sinus rhythm, AF, other abnormal rhythms, and noise on single-lead short ECG recordings, achieving an average F1 score of 83%. Ullah et al. [20] changed the one-dimensional time series into two-dimensional spectra by short-time Fourier transform, and the proposed deep neural network model was a two-dimensional CNN composed of four convolution layers and four pooling layers to classify eight categories of the ECG in MIT-BIH arrhythmia database with the accuracy of 99.11%. Li et al. [21] used the BiLSTM-attention model to perform five different categories of ECG classification on the MIT-BIH



FIGURE 1: Record CCDD/96833 (0–5000): Atrial premature beats, premature ventricular contraction, and complete right bundle branch block.

arrhythmia database, which effectively improved the accuracy of cardiac beat classification. Yao et al. [2] proposed attention-based time-incremental convolutional neural network (ATI-CNN) achieving 81.2% overall classification accuracy for nine ECG categories.

All of the above research has achieved high precision, but it is known from clinical experience that a single ECG recording may contain more than one disease label at the same time and that physicians give more than one diagnostic result when making an ECG diagnosis, so the above studies based on single-label ECG recording cannot meet the needs.

A subset of investigators is currently beginning to explore ECG multilabel classification. Wang et al. [22] proposed an arrhythmia detection method based on the multiresolution representation of ECG signals by taking four different deep neural networks as four channel models for ECG vector representations learning and finally performed 34 kinds of ECG classification on a multilabel HF-challenge dataset with the $F1$ score of 92.38%, all higher than the results of individual channels. Cai et al. [23] developed multi-ECGNet to identify patients with multiple cardiac diseases at the same time, with an $F1$ score of 86.3% in identifying 55 arrhythmia classifications. Sun et al. [24] proposed a novel ensemble multilabel classification model to perform 7 kinds of multilabel ECG classification on the CCDD, and the final $F1$ score was 75.2%. All of the above provided feasible analysis methods for ECG multilabel

classification, but all of the above validation methods were performed on a single database, which could not reflect the generalizability of the model. Li et al. [25] proposed a model consisting of inception, GRU, and attention mechanism on a private multilabel ECG dataset for nine multilabel ECG classifications with an average $F1$ score over 88.6% and a maximum of 91.9% and also extracted the relevant category on the publicly available database PhysioNet for validation with an accuracy of 92.8% but only extracted 500 pieces of data from public databases and the results were not representative. Although the above methods accomplish multilabel ECG classification, there are still shortcomings in terms of accuracy and generalizability. Therefore, applying deep learning for multilabel ECG classification remains a challenge.

3. Methods

The multilabel ECG classification problem uses a deep learning model to automatically extract effective information from ECG records and predicts multiple categories for each record at the same time. In this work, the key issue is to construct an effective algorithm to classify 9 kinds of multilabel ECG signals. In response to this problem, this work proposes the CSA-MResNet architecture, which integrates the multiscale residual network and the channel

spatial attention mechanism structure and can automatically extract useful information from the 12-lead ECG signal to perform multilabel ECG classification.

3.1. Problem Formulation. The multilabel classification task of ECG signals is to automatically classify 9 ECG categories using clinical 12-lead ECG. The proposed model requires the input of 12-lead ECG records and the output of the possible prediction labels. In the original ECG record x_i , together with the corresponding reference label space $Y_i = \{y_1, y_2, \dots, y_q\}$, $0 \leq q \leq 8$ represents a label space with q possible class labels, where $x_i \in \mathbb{R}^{5000 \times 12}$ is a signal with

length 5000 and 12 leads. The multilabel ECG classification task is from the multilabel training set $D = \{(X_i, Y_i) | 1 \leq i \leq N\}$ learning function $f: x_i \rightarrow (y_1, y_2, \dots, y_q)$, where N is the total number of samples, and $0 \leq y_j \leq 1$, $1 \leq j \leq q$. The output threshold is set to 0.5. If some ECG predicts that the j th category prediction probability is greater than 0.5, then these signals are considered to belong to this j th category, if not, otherwise. The work uses CSA-MResNet as an ECG multilabel classifier. This work uses binary cross-entropy to measure the loss between the model output and the actual sample label, as shown in

$$\text{loss}(f(x_i), Y_i) = \text{mean} \left(\sum_{i=1}^q -\omega_i (Y_i \times \log f(x_i) + (1 - Y_i) \times \log(1 - f(x_i))) \right), \quad (1)$$

where q is the total number of arrhythmia categories, and $1 \leq i \leq q$; weight $\omega_i = 1/\log_{n_i}$ is the weight assigned to each category of arrhythmia; n_i is the number of samples of this category of arrhythmia. This allows the model to pay more attention to the categories of arrhythmias with a small number of samples, thereby reducing the impact on data imbalance to a certain extent.

3.2. Model Architecture. Figure 2 illustrates that the overall structure of the CSA-MResNet model is composed of a multiscale residual network and a channel spatial attention mechanism module. The network is mainly designed and constructed based on the experience of GoogleNet [26] and ResNet [27]. Both GoogleNet and ResNet are relatively mature neural networks, which can solve various image classification and target detection problems in the field of computer vision. The CSA-MResNet model contains 2 streams after the convolutional layer and the maximum pooling layer. Each stream has 32 layers of convolution and an average pooling layer, and residual connections are added on the basis of convolution to prevent the accuracy rate caused by deepening from degrading rapidly after saturation, and the channel spatial attention mechanism [28] is added before each pooling layer to obtain more important weights. The model is divided into two parts: feature fusion and multiscale feature fusion. The feature fusion includes convolution, channel spatial attention mechanism, and pooling layer. In this part, convolution is set to 15, and only the feature set of data in this scale is extracted. Then important information is filtered through the channel spatial attention mechanism, and finally the information is fused through the pooling layer. In the multiscale feature fusion part, two streams are assigned convolution kernels of different scales, which can extract convolution kernels of different scales. In each stream, a 32-layer residual network and channel spatial attention mechanism are set. Then, the features of two streams with different scales are aggregated through the pooling layer. Finally, the information on the

two scales is merged and sent to the fully connected layer for 9 multilabel ECG classifications.

3.2.1. Multiscale Residual Network. The multiscale residual network layer configuration is illustrated in Figure 3, where /2 means that the number of neurons passing through this layer is halved. $\times 3, @64$ represents the number of this structure multiplied by 3, and the convolution kernel number of this part is 64. ($\times 3 @ 64; \times 4 @ 128; \times 6 @ 256; \times 3 @ 512$) is the same as the above. The main structure of this model is the residual network of two streams. The residual network is used to prevent the accuracy degradation caused by the model depth being too deep. Two streams with different scales can pay attention to the ECG features of different data segments and different leads. Finally, the features extracted from different scales of the two streams can be merged to achieve a better classification effect. One of the streams in this model is ResNet, but on the basis of its structure, the first layer of the convolutional layer and the last layer of the fully connected layer are removed, and there are 32 convolutional layers. Different convolution kernel sizes are used in the residual structure of the two streams, and the convolution kernel size of one stream is set to 3, and the convolution kernel size of the other stream is set to 7. After each convolution layer, there are a Batch-Norm layer and a ReLU layer. BatchNorm [29] can speed up the convergence speed of the model, and ReLU [30] activation function can avoid gradient saturation problem.

3.2.2. Channel Spatial Attention Mechanism. To better focus on abnormal ECG data, a channel spatial attention mechanism is added to the model, which can focus on channel information and spatial information at the same time, compared with the ‘‘Squeeze-and-Excitation’’ (SE) module that only focuses on channels [31] which has better performance. The structure is shown in Figure 4, which can combine channel information and spatial information at the same time.

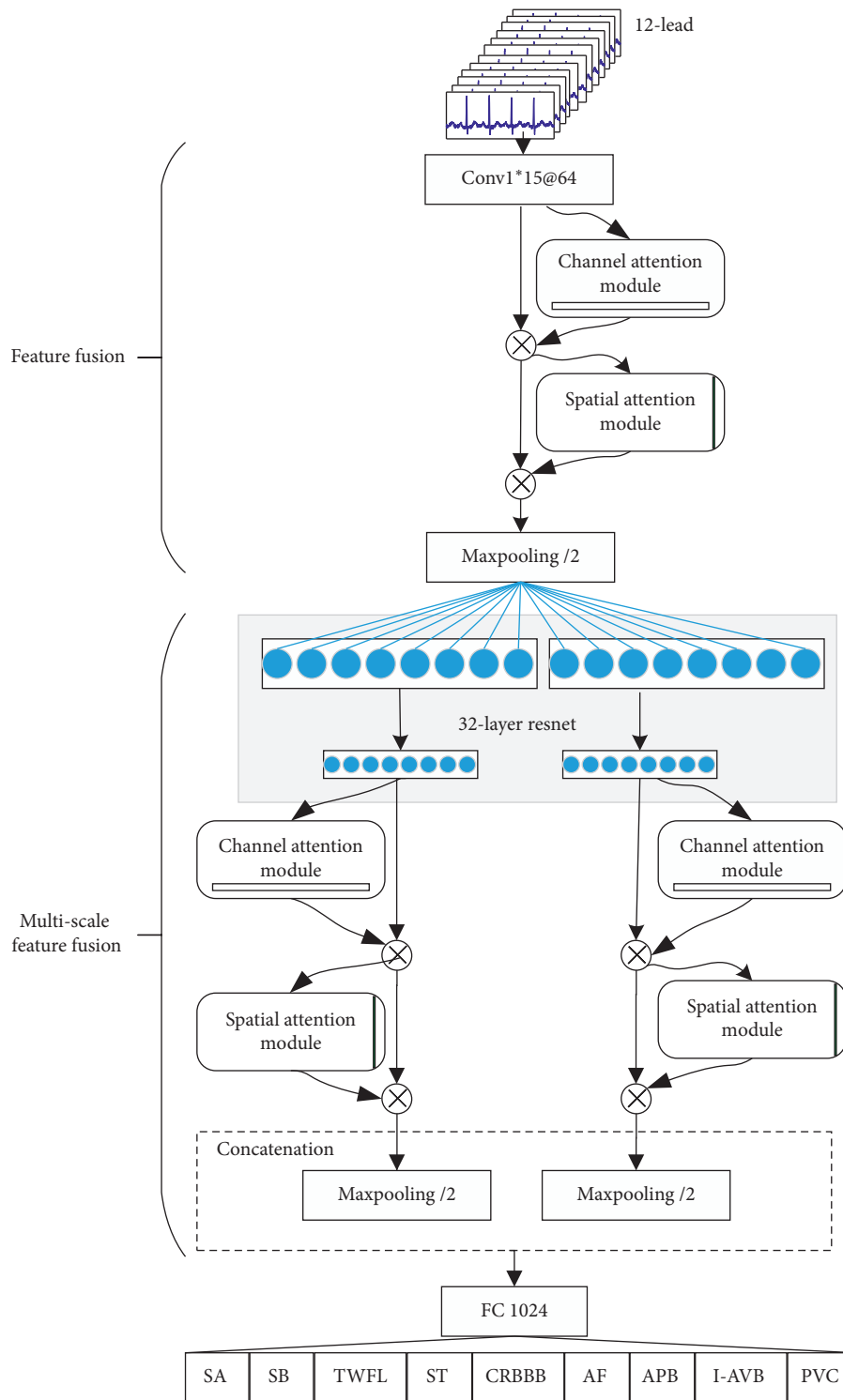


FIGURE 2: The architecture for CSA-MResNet.

The channel spatial attention mechanism includes two submodules, the channel attention mechanism (Figure 4(a)) and the spatial attention mechanism (Figure 4(b)). The channel attention mechanism obtains the channel attention map M_c through the selection of the channel, and in the

other spatial attention mechanism to the important part of the feature of the channel, the spatial attention map M_s is obtained. The input feature F passes through these two parts to obtain the detailed feature F'' . These two steps are represented by equations (2) and (3), respectively:

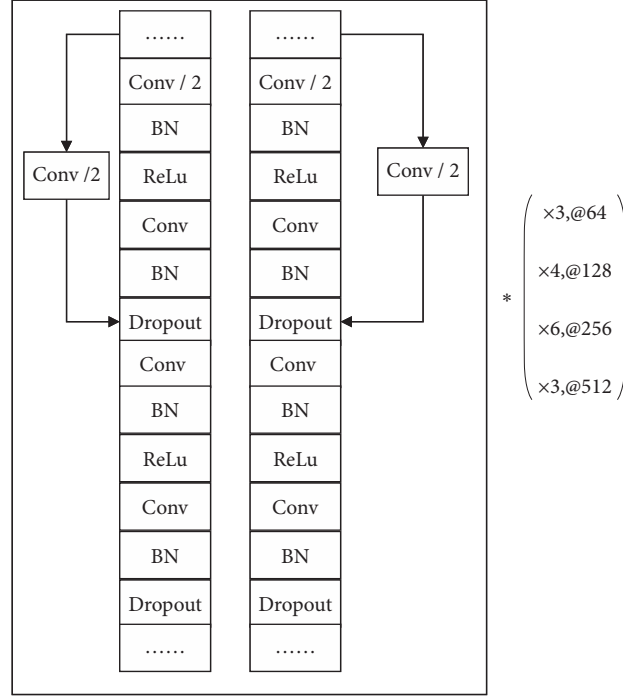


FIGURE 3: Layer configuration for multiscale residual network.

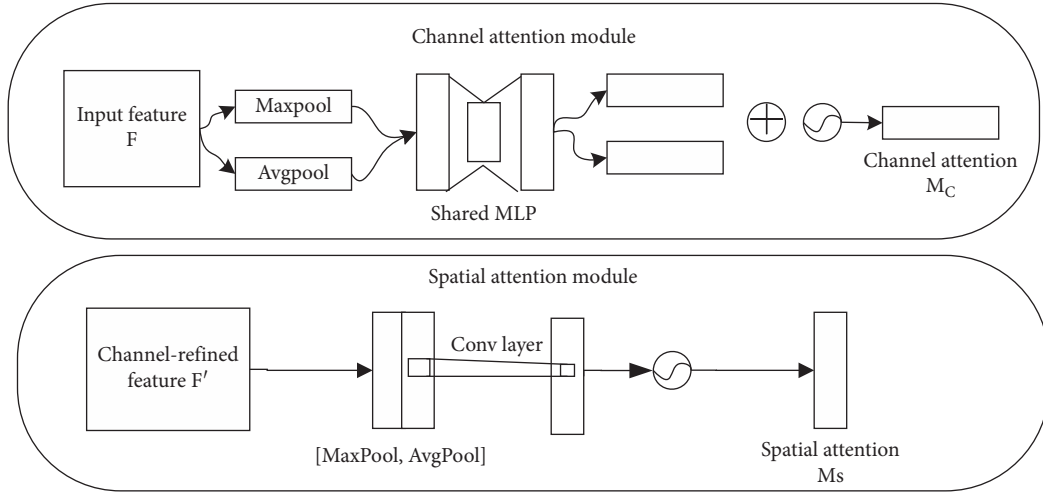


FIGURE 4: Channel convolution attention mechanism diagram (top (a): channel attention mechanism; bottom (b): spatial attention mechanism).

$$F' = M_c(F) \otimes F, \quad (2)$$

$$F'' = M_s(F') \otimes F'. \quad (3)$$

Figure 4(a) shows how the channel attention mechanism works, and its ability to channel the selection of input features allows the model to focus more on channels that are useful for the task. The parameters of this module were obtained by calculating the global average pooling and the global maximum pooling information about input features, followed by merging these two parts of information, in this

process both share the same fully connected network, and finally the spatial attention weights are compressed into 0-1 using the sigmoid activation function. This process can be shown as

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool)(F) + MLP(MaxPool)(F)) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))). \end{aligned} \quad (4)$$

Figure 4(b) shows how the spatial attention mechanism works, which can reduce the interference of other

nonimportant information on the same channel to the task and improve the accuracy of the model. Features that underwent a global maximum pooling and global average pooling of features output by the channel attention mechanism were convolved, using the sigmoid activation function to compress spatial attention weights to 0-1. This process can be shown as

$$\begin{aligned} M_s(F) &= \sigma\left(f^{1 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])\right) \\ &= \sigma\left(f^{1 \times 7}\left([F_{\text{avg}}^s; F_{\text{max}}^s]\right)\right). \end{aligned} \quad (5)$$

The channel attention mechanism focuses on the channels that contribute more to the ECG signal. The spatial attention mechanism assigns greater weight to more important information in different time periods of the ECG signal. The channel attention mechanism is a global application, and the spatial attention mechanism is local to the feature which plays an important role. Literature [28] shows that sequential placement has better performance than parallel placement, and the performance of channel priority is higher than spatial priority. Therefore, the attention mechanism is placed between the first convolution layer and the pooling layer and before the last pooling layer for the two streams, respectively. And the channel attention mechanism is prior to the spatial attention mechanism.

4. Experiments and Results

The data used in this work were obtained from a clinical 12-lead multilabel CCDD, including nine categories of ECG signals: (1) sinus arrhythmia (SA); (2) sinus bradycardia (SB); (3) T wave low and flat (TWFL); (4) sinus tachycardia (ST); (5) complete right bundle branch block (CRBBB); (6) atrial fibrillation (AF); (7) atrial premature beat (APB); (8) first-degree atrioventricular block (I-AVB); (9) premature ventricular contraction (PVC).

4.1. Environment. This work proposes model training and testing on XeonR Silver-4114CPU, 32 GB memory, and Geforce2080Ti graphics card. The server runs on the ubuntu 18.04 system, and the model is run on the PyTorch 1.2.0 framework.

4.2. Data. In this work, the CCDD multilabel dataset and HF-challenge multilabel dataset are used. In the experiment, contrasts to each indicator on different models on the CCDD are done to find the best model, followed by model generalization validation on the HF-challenge dataset. The number of ECGs used in detail is shown in Table 1.

4.2.1. CCDD. The multilabel ECG samples used in this work were collected from the China Cardiovascular Disease Database [9] (CCDD). The database contains 190,000 12-lead clinical multilabel ECGs' data, the sampling rate is 500 Hz, and the data duration is 10~20 seconds. In the CCDD, the ECG is divided into 12 primary disease types, and there are many secondary disease types. In order to

TABLE 1: The number of 9 disease types used in this work is in different databases.

	CCDD training	CCDD testing	HF-challenge testA
SA	2411	2507	54
SB	2903	2042	724
TWFL	4603	2015	262
ST	2747	1737	357
CRBBB	1892	1056	46
AF	1746	769	121
APB	1714	711	51
I-AVB	1534	634	9
PVC	1158	427	35
Total	17952	10635	1247

make better use of the advantages of deep learning, we choose to follow the two criteria of including as many types as possible and large amounts of data. And finally this work mainly selected 9 common disease categories in the CCDD for multilabel ECG classification (when selecting data, only ECG records containing these 9 categories of labels are retained; once a label other than the above 9 types of labels appears in an ECG record, these data will be deleted). The training set and test set were divided according to the literature [32], and the validation set was randomly selected from 10% of the training set. Data processing takes 5000 points of all the ECG signals, and to ensure the quality of the signal, the first 1.25 s is discarded, and the middle 10 s data are taken. If they are less than 10 seconds, delete the data directly. The detailed data information of the final experiment can be seen in Table 1.

4.2.2. HF-Challenge. The HF-challenge multilabel dataset is an ECG smart competition organized by the Tianchi platform [23]. The preliminary data contain 24106 records in the training set and 8036 in the test set A (testA), each record has 8 leads (mainly I, II and six limb lead V1~V6), the length is 10 seconds, a total of 55 categories, the sampling rate is 500 Hz, and the unit voltage is 4.88 microvolts. This work mainly used the 9 disease categories described in Section 4.2 in the testA and expanded each piece of data to a length of 5000 * 12 according to the official equation (6), which is Einthoven's law [33], and then multiplied by the unit voltage and converted it into real data for experiments. The ECG of cardiovascular disease may manifest in many ways. After the ECGs of different databases are marked by different doctors, the final marking description is not completely consistent due to the different degree of disease segmentation and expert marking standards, but the clinical manifestations of the disease are the same. For example, the use of flat T waves and low frequencies in the CCDD and the use of T wave changes in the HF-challenge testA are all abnormal T waveforms [34]. And the abnormal T wave is a typical manifestation of myocardial ischemia on the ECG [35]. Therefore, no distinction is made between T wave anomalies and T wave changes in this work. The distribution of the number of different categories of ECG is shown in Table 1.

$$\begin{aligned}
\text{III} &= \text{II} - \text{I}, \\
\text{aVR} &= -\frac{(\text{I} + \text{II})}{2}, \\
\text{aVL} &= \text{I} - \left(\frac{\text{II}}{2}\right), \\
\text{aVF} &= \text{II} - \left(\frac{\text{I}}{2}\right).
\end{aligned} \tag{6}$$

4.3. Training Setting. This work training uses clinical multi-label CCDD, grouping them into batches of 64 records and sending them into the model. The larger the batch, the faster the network model training, and at the same time, more memory is required. Through fine-tuning of hyperparameters, 64 were selected as the batch size. The weight initialization of the convolutional layer and the fully connected layer in CSA-MResNet uses the kaiming initializer [36] and the Adam optimizer [37] to accelerate the convergence of the network model. The default learning rate is from 0.001. In a total of 256 epochs, the learning rate is multiplied by 0.1 whenever 32, 64, and 128 epochs are encountered. The first convolution of the model is set to 15. In the two streams of the model, one of the convolution kernels is fixed at 7, and the convolution kernel size of the other stream is set differently, and the results are compared on the test set. The number of filters is performed in the order of 64, 128, 256, 512 according to different block combinations. The convolution step size of the residual connection is set to 2 to make the output size equal, and the step size of the remaining convolution and pooling is set to 1. Dropout [38] uses a ratio of 0.2 to prevent neurons from adapting to the training data. After the CSA-MResNet model is trained, the binary cross-entropy loss between the output of each batch and the actual label is calculated, and then backpropagation is performed. During the entire training period, save the model weight when the F_1 score of the validation set is the highest.

4.4. Result

4.4.1. Performance Metrics. This work evaluates the performance of 9 multilabel ECG classifications using the CSA-MResNet model on a large-scale test set. The work uses a multilabel classification method [39]; for the j^{th} class label y_j , the four basic quantities characterizing the binary classification performance can be defined as equations (7)~(10).

$$\text{TP}_j = |\{x_i | y_i \in Y_j, y_i \in f(x_i), 1 \leq i \leq N\}|, \tag{7}$$

$$\text{FP}_j = |\{x_i | y_i \notin Y_j, y_i \in f(x_i), 1 \leq i \leq N\}|, \tag{8}$$

$$\text{TN}_j = |\{x_i | y_i \notin Y_j, y_i \notin f(x_i), 1 \leq i \leq N\}|, \tag{9}$$

$$\text{FN}_j = |\{x_i | y_i \in Y_j, y_i \notin f(x_i), 1 \leq i \leq N\}|, \tag{10}$$

where $0 \leq j \leq 8$, and N is the total number of samples. TP represents the number of correctly classified records of a category. FP represents the number of records that belong to other categories and are incorrectly classified into a certain category. TN represents the number of records that actually belong to other categories and are finally classified into other categories. FN refers to the number of samples belonging to a certain class that were misclassified as in other classes.

Based on the above four quantities, typical classification indicators [40–42] on each class including specificity (Spe), precision (Pre), recall (Rec), accuracy (Acc), and F_1 score (F_1) are derived accordingly and defined as equations (11)~(15).

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{11}$$

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{12}$$

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

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{14}$$

$$F_1 = \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}}. \tag{15}$$

The F_1 score is a comprehensive index with a certain degree of stability, so it is used as the final evaluation index.

4.4.2. Optimal Size of Convolution Kernel. The size of the convolution kernel in deep neural networks is a key parameter that determines the performance of the model. Therefore, in this experiment, a comparison of different convolution kernel sizes is set. After the model structure is fixed, ensure that the convolution kernel size of one stream is 7, and the convolution kernel of the other stream is selected in (3, 5, 7, 9, 11), which are MResDNN-37, MResDNN-57, MResDNN-77, MResDNN-79, and MResDNN-711. For comparison to derive the most appropriate number of streams, three different convolution kernel scale sizes of MResDNN-357 were set. The above models all set the same hyperparameters such as optimizer, batch size, and learning rate. The classification results of different convolution kernel size models in the multilabel CCDD test set are shown in Table 2.

Table 2 compares the Acc, Sen, Pre, Rec, and F_1 of different convolution kernel size models in the classification of nine multilabel ECGs. The results show that when performing multilabel abnormal ECG classification, comparing the results of MResDNN-357 with the other three models, the average F_1 score is only 86.2%, which indicates that the classification performance of the three streams is the lowest. In the MResDNN-37, CRBBB increased by up to 3.7%, APB increased by up to 3.4%, and the final F_1 score was also the highest among the above models, reaching 86.8%, which may be because the greater

TABLE 2: Results of different convolution kernel size models on the CCDD testing (%).

	F1					
	MResDNN-37	MResDNN-57	MResDNN-77	MResDNN-79	MResDNN-711	MResDNN-357
SA	86.1	84.9	85.7	85.0	85.6	85.9
SB	93.0	93.2	92.9	93.0	93.2	93.0
TWFL	79.6	79.6	78.8	79.5	78.5	78.2
ST	95.3	95.6	95.6	95.2	95.8	95.5
CRBBB	97.5	95.1	96.6	96.2	95.6	93.8
AF	95.7	96.0	96.0	95.8	94.6	95.9
APB	61.7	60.7	58.3	58.9	59.0	58.4
I-AVB	78.6	78.4	79.1	79.2	76.6	77.6
PVC	60.4	66.6	72.8	75.3	59.0	67.3
Average						
Acc	96.8	96.7	96.8	96.7	96.7	96.7
Spe	98.6	98.5	98.6	98.5	97.6	98.5
Pre	89.7	88.8	90.3	88.6	83.1	88.9
Rec	84.0	84.2	83.5	84.6	89.4	83.6
F1	86.8	86.5	86.7	86.5	86.1	86.2

the difference in scale, the greater the degree of difference in obtaining ECG features. It can be seen from the results in the table that the final score of $F1$ of PVC and APB is lower, because the deep learning performance is greatly affected by the amount of data. PVC has fewer data pieces in the training set, and the model does not fully learn all the characteristics, while APB accounts for the largest proportion of all the data in the training set and the test set, and the data distribution affects the model performance to a certain extent. And TWFL has a low $F1$ score because of the largest proportion difference between the training set and the test set. It can better represent the more comprehensive characteristics of the ECG signal. According to the experimental results, the two convolution kernel sizes of 3 and 7 are selected as the benchmark in the two streams.

4.4.3. Optimal Position of Channel Spatial Attention Mechanism. The results of Section 4.4.2 show that the best result is when the convolution kernel size of one stream is fixed to 3 and the convolution kernel size of the other stream is 7, so the channel spatial attention mechanism is embedded on this benchmark. In order to select the embedding position of the channel spatial attention mechanism, the module is embedded in the residual block structure and named MCSA-ResDNN, and in the model CSA-MResNet proposed in this paper, the attention mechanism is placed between the first convolution layer and the pooling layer and before the last pooling layer for the two streams, respectively. Compare the above two models with the channel spatial attention mechanism and the model without this mechanism. The convolution size in the channel spatial attention mechanism is set to 7. The position comparison results of the channel spatial attention mechanism are shown in Table 3.

From the results in Table 3, it can be seen that the overall $F1$ score of MCSA-ResDNN is increased by 0.7% compared to MResDNN-37 without the channel spatial attention mechanism. The $F1$ score on the PVC category is increased by 10.8%; almost all the average indicators are higher than

TABLE 3: Results of different positions of channel spatial attention mechanism on the CCDD testing (%).

	F1		
	MResDNN-37	MCSA-ResDNN	CSA-MResNet
SA	86.1	86.9	87.3
SB	93.0	93.3	93.0
TWFL	79.6	79.7	80.5
ST	95.3	95.7	95.8
CRBBB	97.5	95.9	96.4
AF	95.7	95.6	96.9
APB	61.7	69.9	69.9
I-AVB	78.6	78.6	80.2
PVC	60.4	71.2	76.6
Average			
Acc	96.8	97.0	97.1
Spe	98.6	98.6	98.7
Pre	89.7	89.8	90.6
Rec	84.0	85.4	85.9
F1	86.8	87.5	88.2

MResDNN-37. It proves that using the channel spatial attention mechanism can improve performance to a certain extent. The comparison of the results from the different embedding positions of the channel spatial attention mechanism shows that the $F1$ scores of almost all categories of CSA-MResNet are better than MCSA-ResDNN, and the $F1$ scores on the SA category are increased by 1.2%. The $F1$ score of the PVC category has increased by 16.2%. From the above results, it can be seen that CSA-MResNet is the best model in the comprehensive index $F1$ score among the above models. From the model point of view, CSA-MResNet which adds channel spatial attention mechanism to the first two ends of multiscale residual network can pay more attention to more important channels and more important data fragments, while the MCSA-ResDNN which adds channel spatial attention mechanism to a multiscale residual network has better performance than MResDNN-37 which does not add attention mechanism. However, the channel spatial attention mechanism added to the residual structure

TABLE 4: The classification performance of the proposed model is verified in HF-challenge testA (%).

	Acc	Spe	Pre	Rec	F1
SA	93.7	94.6	37.9	72.2	49.7
SB	93.4	100	100	88.7	94.0
TWFL	84	99.7	95.7	25.2	39.9
ST	99.4	100	100	97.8	98.9
CRBBB	99.8	99.9	97.8	97.8	97.8
AF	99.8	99.9	99.2	99.2	99.2
APB	97.0	99.6	79.2	37.3	50.7
I-AVB	99.6	99.9	83.3	55.6	66.7
PVC	98.6	99.8	87.0	57.1	69.0
Average	96.1	99.1	94.0	78.9	85.8

TABLE 5: Comparison results of different research work on the CCDD.

Literature	ECG categories	Classifier	Performance
[32]	2	CBRNN	Spe = 76.32% Se = 75.52% Acc = 87.69%
[44]	2	Ensemble deep learning	Spe = 86.86 ± 3.51% Se = 80.23 ± 4.49% Acc = 84.84 ± 1.82%
[45]	2	LCNN	Spe = 83.84% Se = 83.43% Acc = 83.66%
[46]	2	Heart rate and LCNN fuse	Spe = 84.45% Se = 85.19% Acc = 84.77%
[47]	2	ResNet50	Spe = 91.63% Se = 87.73% Acc = 89.43%
[24]	7 multilabel	Ensemble multilabel classification model	Se (Rec) = 71.6% Acc = 75.2% Pre = 80.8% F1 = 75.2%
This work	9 multilabel	CSA-MResNet	Spe = 98.7% Se (Rec) = 85.9% Acc = 97.1% Pre = 90.6% F1 = 88.2%

destroys the performance of the original residual connection to a certain extent, resulting in the model ignoring the information of identity connection in learning and training, thus reducing the learning ability of the model. Therefore, the CSA-MResNet model with channel spatial attention mechanism added to both ends of multiscale residual network performs best.

4.4.4. Other Database Generalization Verification. There are many deep learning methods for intelligent ECG classification, but most previous studies used only one database to verify the results, and there is no independent cross-validation method. Generalization is the difficulty of traditional deep learning [43]. In the actual clinical environment, the ECG data obtained by different collection devices are

different, and the ECG data contain a variety of abnormal information about the human body. Whether a single database ECG can cover all ECG features is questionable, so generalization is worth analyzing.

In this work, we verify the CSA-MResNet deep learning model on the HF-challenge testA, use the same network structure and parameter test as Section 3.2, select the corresponding 9 diseases from the database, expand it to 12-lead ECG data according to equation (6), and test it on the dataset with the trained model. The results are shown in Table 4.

The results in Table 4 show that the CSA-MResNet model has lower average F1 scores, less than 60% in all, and lower generalization performance on the SA, TWFL, and APB categories. This may be because the ECG data for these categories are more complex or because of the small amount of experimental data, illustrating that these categories in the

CCDD cannot contain all the abnormal information. We can think that if samples of the abovementioned poor performance diseases are collected from other databases, and the HF-challenge dataset is generalized and verified after multidatabase training, the performance of the model on these types can be improved. But the mean $F1$ scores on the categories SB, ST, CRBBB, and AF were all greater than 90% and comparable to those tested on the CCDD. Finally, on average metrics, both Acc and Rec were lower than results on the CCDD, and Spe and Pre were higher than the CCDD average results. The average $F1$ score of 85.8% on this dataset and the average $F1$ score obtained on the CCDD do not differ much, demonstrating a certain generalization of the model proposed in this work.

4.4.5. Comparison with Previous Studies. Table 5 shows the contrasting results of different classification methods on the clinical CCDD between our method and others. The results show that CSA-MResNet end-to-end deep learning model is able to provide better classification performance and compare with other works discussed in this paper. The $F1$ score for multilabel classification of the final model was 88.2%, which provided a better method for multilabel classification.

Previous studies in the CCDD [32, 45], respectively, used CBRNN and LCNN models to conduct two normal and abnormal classifications on the database. Literature [44] used ensemble learning methods to identify both normal and abnormal categories, and [46] used heart rate to screen the ECGs and then use LCNN to conduct two normal and abnormal classifications. Literature [47] used ResNet50 for two classifications with a classification accuracy of 89.43%, demonstrating that the ResNet had a good performance on the CCDD. The above research works performed normal and abnormal classifications on the CCDD, and the classification performance indexes were expressed by the accuracy, and it can be seen that the best result was 89.43%, while literature [24] is the only experiment in a comparative research work using the CCDD for multilabel ECG classification. For multilabel classification experiments, it is more reasonable to use the average $F1$ score as an evaluation index due to the problem of sample imbalance. Literature [24] used the integrated multilabel classification model to achieve a final $F1$ score of 75.2% in 7 ECG categories. This work uses the CSA-MResNet model to classify 9 ECG multilabels, and the $F1$ score of this work is 88.2%, indicating that this model is more competitive than the latest methods.

5. Conclusion

In order to solve the problem of multilabel classification of clinical ECG, a multiscale residual deep neural network CSA-MResNet model based on the channel spatial attention mechanism for automatic multilabel classification is proposed in this work. The final $F1$ score is 88.2%. The model has a good recognition rate for multilabel ECG classification and provides a feasible analysis method for multilabel ECG classification. This work also verified the model on other HF-

challenge testA, and the final $F1$ score was 85.8%, proving that the model is generalizable. However, the experimental data do not include all ECG categories and more experimental data are needed. In the future work, we will pay more attention to this aspect of research.

Data Availability

The data used to support the findings of this study are included within the article. Further data can be requested from the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (grant number 2017YFB1401200); Key Science and Technology Project of Xinjiang Production and Construction Corps (grant number 2018AB017); Key Research, Development, and Dissemination Program of Henan Province (Science and Technology for the People) (grant number 182207310002); National Natural Science Foundation of China (grant number 61872324); Program for Science & Technology Innovation Talents in Universities of Henan Province (grant number 21HASTIT031); Training Plan for Young Backbone Teachers of Colleges and Universities in Henan (grant number 2019GGJS018); Key Scientific Research Project of Colleges and Universities in Henan Province (grant number 20A520035); Collaborative Innovation Major Project of Zhengzhou (grant number 20XTZX06013).

References

- [1] R. Li, S. Shen, G. Chen et al., "Multilevel risk prediction of cardiovascular disease based on adaboost plus RF ensemble learning," in *Proceedings of the 2019 5th International Conference on Electrical Engineering, Control and Robotics*, vol. 533, Guangzhou, China, January 2019.
- [2] Q. Yao, R. Wang, X. Fan et al., "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network," *Information Fusion*, vol. 53, 2019.
- [3] X. Zhang, R. Li, H. Dai, Y. Liu, B. Zhou, and Z. Wang, "Localization of myocardial infarction with multi-lead bidirectional gated recurrent unit neural network," *IEEE Access*, vol. 7, pp. 161152–161166, 2019.
- [4] A. Hannun, P. Rajpurkar, G. H. Tison et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, 2019.
- [5] P. Saman and J. Rubin, "Electrocardiogram monitoring and interpretation: from traditional machine learning to deep learning, and their combination," in *Proceedings of the Computing in Cardiology*, vol. 45, Maastricht, Netherlands, September 2018.
- [6] Y.-J. Chen, C.-L. Liu, V. S. Tseng, Y.-F. Hu, and S.-A. Chen, "Large-scale classification of 12-lead ECG with deep learning," in *Proceedings of the 2019 IEEE EMBS International*

- Conference on Biomedical & Health Informatics (BHI)*, May 2019.
- [7] T. Gong, B. Liu, Q. Chu, and N. Yu, "Using multi-label classification to improve object detection," *Neurocomputing*, vol. 370, pp. 174–185, 2019.
 - [8] S.-J. Shin, S. Kim, Y. Kim, and S. Kim, "Hierarchical multi-label object detection framework for remote sensing images," *Remote Sensing*, vol. 12, no. 17, p. 2734, 2020.
 - [9] J. Zhang, X. Liu, J. Dong et al., "CCDD: an enhanced standard ECG database with its management and annotation tools," *International Journal on Artificial Intelligence Tools*, vol. 21, no. 5, pp. 6721–7676, 2012.
 - [10] J. Wang, "A deep learning approach for atrial fibrillation signals classification based on convolutional and modified Elman neural network," *Future Generation Computer Systems*, vol. 102, 2019.
 - [11] S. W. E. Baalman, F. E. Schroevers, A. J. Oakley et al., "A morphology based deep learning model for atrial fibrillation detection using single cycle electrocardiographic samples," *International Journal of Cardiology*, vol. 316, pp. 130–136, 2020.
 - [12] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.
 - [13] U. R. Acharya, F. Hamido, V. K. Sudarshan et al., "Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads," *Knowledge-Based Systems*, vol. 99, pp. 146–156, 2016.
 - [14] A. Emina and S. Abdulhamit, "Effect of multiscale PCA de-noising in ECG beat classification for diagnosis of cardiovascular diseases," *Circuits Systems and Signal Processing*, vol. 34, no. 2, pp. 513–533, 2015.
 - [15] C.-C. Lin and C.-M. Yang, "Heartbeat classification using normalized RR intervals and morphological features," *Mathematical Problems in Engineering*, vol. 2014, 2014.
 - [16] S. K. Pandey, R. R. Janghel, and V. Vani, "Patient specific machine learning models for ECG signal classification," *Procedia Computer Science*, vol. 167, pp. 2181–2190, 2020.
 - [17] K. N. V. P. S. Rajesh and R. Dhuli, "Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier," *Biomedical Signal Processing and Control*, vol. 41, pp. 242–254, 2018.
 - [18] U. R. Acharya, S. L. Oh, Y. Hagiwara et al., "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, p. 389, 2017.
 - [19] R. Kamaleswaran, R. Mahajan, and O. Akbilgic, "A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using varying length single lead electrocardiogram," *Physiological Measurement*, vol. 39, no. 3, 2018.
 - [20] U. Amin, S. M. Anwar, M. Bilal et al., "Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation," *Remote Sensing*, vol. 12, no. 10, p. 1685, 2020.
 - [21] R. Li, X. Zhang, H. Dai, B. Zhou, and Z. Wang, "Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and BiLSTM-attention neural network," *IEEE Access*, vol. 7, pp. 109870–109883, 2019.
 - [22] D. Wang, M. Hua, D. Chen, H. Zhang, and L. Xu, "Automatic detection of arrhythmia based on multi-resolution representation of ECG signal," *Sensors*, vol. 20, no. 6, 2020.
 - [23] J. Cai, W. Sun, J. Guan, and I. You, "Multi-ECGNet for ECG arrhythmia multi-label classification," *IEEE Access*, vol. 8, pp. 110848–110858, 2020.
 - [24] Z. Sun, C. Wang, Y. Zhao, and C. Yan, "Multi-label ECG signal classification based on ensemble classifier," *IEEE Access*, vol. 8, pp. 117986–117996, 2020.
 - [25] D. Li, H. Wu, J. Zhao, T. Ye, and J. Fu, "Classification system of arrhythmias using 12-lead ECGs with a deep neural network based on an attention mechanism dagger," *Symmetry-Basel*, vol. 12, no. 11, 2020.
 - [26] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
 - [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
 - [28] S. Woo, J. Park, J.-Y. Lee et al., "CBAM: convolutional block attention module," in *Proceedings of the Computer Vision-ECCV 2018 15th European Conference*, pp. 3–19, Munich, Germany, September 2018.
 - [29] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning, PMLR 37*, Lille, France, July 2015.
 - [30] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, <https://arxiv.org/abs/1505.00853>.
 - [31] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
 - [32] E. K. Wang, L. Xi, R. P. Sun et al., "A new deep learning model for assisted diagnosis on electrocardiogram," *Mathematical Biosciences and Engineering: MBE*, vol. 16, no. 4, pp. 2481–2491, 2019.
 - [33] J. E. Madias, "A comparison of 2-lead, 6-lead, and 12-lead ECGs in patients with changing edematous states," *Chest*, vol. 124, no. 6, pp. 2057–2063, 2003.
 - [34] L. Chhabra and D. H. Spodick, "Hypothermia masquerading as pericarditis: an unusual electrocardiographic analogy," *Journal of Electrocardiology*, vol. 45, no. 4, pp. 350–352, 2012.
 - [35] M. D. Jacobsen, G. S. Wagner, L. Holmvang et al., "Clinical significance of abnormal T waves in patients with non-ST-segment elevation acute coronary syndromes," *The American Journal of Cardiology*, vol. 88, no. 11, 2001.
 - [36] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," 2015, <https://arxiv.org/abs/1502.01852>.
 - [37] D. Kingma and B. Jimmy, "Adam: a method for stochastic optimization," 2014, <https://arxiv.org/abs/1412.6980>.
 - [38] N. Srivastava, G. Hinton, A. Krizhevsky et al., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
 - [39] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.
 - [40] Y. Wang, B. Xiao, B. Li, W. Li, and M. Xu, "Pay attention and watch temporal correlation: a novel 1-D convolutional neural network for ECG record classification," in *Proceedings of the 2019 Computing in Cardiology Conference*, Biopolis, Singapore, September 2019.

- [41] R. Li, S. Ji, S. Shen et al., "Arrhythmia multiple categories recognition based on PCA-KNN clustering model," in *Proceedings of the 2019 8th International Symposium on Next Generation Electronics (ISNE)*, Zhengzhou, China, October 2019.
- [42] T. Xie, R. Li, S. Shen et al., "Intelligent analysis of premature ventricular contraction based on features and random forest," *Journal of Healthcare Engineering*, vol. 2019, 2019.
- [43] 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, no. 10, 2020.
- [44] L. Jin and J. Dong, "Ensemble deep learning for biomedical time series classification," *Computational Intelligence and Neuroscience*, vol. 2016, 2016.
- [45] L. Jin and J. Dong, "Deep learning research on clinical electrocardiogram analysis," *Scientia Sinica Informationis*, vol. 45, no. 3, pp. 398–416, 2015.
- [46] H. Li and L. Jin, "An ECG classification algorithm based on heart rate and deep learning," *Space Medicine & Medical Engineering*, vol. 29, no. 3, pp. 189–194, 2016.
- [47] S. Guo, Q. Lv, B. Liu, Y. Lin, and R. Li, "Deep convolutional neural networks for electrocardiogram classification," *Proceedings of the 2nd International Conference on Healthcare Science and Engineering*, vol. 536, pp. 57–69, 2019.