IJLAI Transactions on Science and Engineering (ISSN: 3006-3418) 2025, vol. 3, no. 4, pp: 43–53

Arrhythmia Classification Method Based on CNN-Attention-BiTransformer

Xibin Guo[1], Jiangjiang Li[1], and Lijuan Feng[1]

School of Electronics and Electrical Engineering, Zhengzhou University of Science and Technology
450064 Zhengzhou, China
Corresponding author:Jiangjiang Li *Received Nov. 14, 2025; Revised and Accepted Nov. 25, 2025*

Abstract. This paper proposes a novel arrhythmia classification method combining Convolutional Neural Networks (CNN), Attention mechanisms, and Bidirectional Transformers (BiTransformer). The method aims to improve the accuracy and robustness of arrhythmia detection in electrocardiogram (ECG) signals. Initially, the CNN module extracts local spatial features from raw ECG data, effectively capturing the morphological characteristics of different arrhythmia types. Subsequently, the Attention mechanism is applied to weigh the importance of different segments in the ECG signal, allowing the model to focus on critical features that are most indicative of arrhythmia. Finally, the BiTransformer module processes the feature sequences bidirectionally, capturing both forward and backward dependencies in the ECG signal. This comprehensive approach enables the model to integrate local and global information, enhancing its ability to classify various arrhythmias accurately. Experiments conducted on the MIT-BIH Arrhythmia Database demonstrate that the proposed method achieves state-of-the-art performance, with a significant improvement in classification accuracy compared to traditional methods. The results highlight the effectiveness of combining CNN, Attention, and BiTransformer for arrhythmia classification, offering a promising direction for automated ECG analysis and clinical applications.

Keywords: Arrhythmia classification; CNN; Attention mechanism; Bidirectional Transformer

1. Introduction

In recent years, the incidence of cardiovascular diseases has significantly increased worldwide, among which arrhythmia is one of the main causes of death in cardiovascular diseases [1,2]. In clinical practice, the diagnosis of arrhythmia is mainly determined through electrocardiogram (ECG). However, traditional manual ECG analysis is time-consuming, labor-intensive, and prone to errors [3,4]. Therefore, it is particularly important to achieve computer-assisted automatic classification of arrhythmia.

Currently, computer-assisted arrhythmia classification mainly includes machine learning methods and deep learning methods [5-7]. Machine learning methods mainly include classic analysis methods such as support vector machine, decision tree, random forest, and K-nearest neighbor algorithm [8], but these classification methods rely heavily on the extraction of features based on human experience, making them difficult to adapt to the individual differences of patients in clinical settings, have weak generalization ability, and have many shortcomings [9,10]. In recent years, more research has focused on deep learning methods, which can actively learn more features from the signals and extract potential deep information in the data [11]. Relevant studies have shown that compared with traditional machine learning methods, it has obvious advantages in signal classification.

Mohamed et al. [12] proposed a 9-layer convolutional neural network (CNN) model, which divided the electrocardiogram (ECG) signals into 5 main categories: N, S, V, F, and Q. Thalluri et al. [13] combined the advantages of CNN and Transformer, proposed the PSC-Net network, and used the artificial minority class over-sampling (Synthetic Minority Oversampling Technique, SMOTE) data enhancement algorithm to address the imbalance problem, achieving good results. Sohn et al. [14] proposed a generative adversarial network based on U-Net to synthesize additional ECG signals, achieving an accuracy of 96.20%. All of these are one-dimensional ECG studies. The research shows that two-dimensional convolution can extract the spatial features of ECG and achieve better classification results. Therefore, converting it into a two-dimensional image containing ECG information can leverage the advantage of higher spatial resolution and extract more feature information.

Krl-Jzaga et al. [15] proposed a method that utilizes relative position matrix information to convert electrocardiogram (ECG) signals into two-dimensional images, and used the Gam-ResNet18 network with the concept of fusion transfer learning to complete the classification of arrhythmias. This method improved the classification accuracy through image conversion and attention mechanism techniques, achieving an accuracy rate of 99.30%. Zhang et al. [16] proposed a classification method based on multi-scale residual networks and multi-channel fusion, converting single-lead ECG signals into two-dimensional images, effectively extracting the spatial features

of ECG, and completing the classification of arrhythmias. Zhang et al. [17] introduced an ECG model based on CNN, using recurrence plots (RP), Gramian Angular Field (GAF), and Markov Transfer Field (MTF), integrating three modalities to enhance ECG features, and completing the classification of arrhythmia images. Guo et al. [18] addressed the class imbalance problem by using an auxiliary classifier generative adversarial Network (ACGAN) and proposed a lightweight network LC-CNN for classification, achieving an accuracy rate of 99.22%. Meng et al. [19] used a deep convolutional generative adversarial network (DCGAN) to balance one-dimensional ECG signals and convert them into Gramian Angular Summation Field (GASF) images, using the attention mechanism to adaptively focus on QRS wave features, and finally inputting them into the improved Inception-ResNet-v2 network for classification. Some of these studies used data augmentation techniques and classified arrhythmias using improved networks. These studies show that converting one-dimensional ECG signals into two-dimensional ECG images is beneficial for classification results, but they did not consider that in data balancing, the advantages of two-dimensional ECG can be better exploited.

Based on the shortcomings of the classification of arrhythmias under the intra-patient paradigm, this paper proposes a highly accurate and robust method for arrhythmia classification and recognition. This method combines the advantages of CNN and BiLSTM, and by introducing the SE-Block channel attention mechanism, it constructs a new arrhythmia classification model based on SENet and BiLSTM. The performance of the model is tested using the MITBIH arrhythmia database under the inter-patient paradigm, and the SMOTE algorithm and focal loss function are combined to solve the sample imbalance problem in ECG signals. The final experiments show that the proposed method has significant advantages over other methods.

2. Data Processing

2.1. Source of the Dataset

The experimental data in this article are sourced from the MIT-BIH Arrhythmia Database [20]. This database contains various common life-threatening arrhythmia records and has been labeled and annotated by two experts. The database includes 48 different patients' heart rate data with a signal sampling frequency of 360 Hz. To ensure signal quality and obtain true electrocardiogram (ECG) signals, all signals are filtered through a band-pass filter ranging from 0.1 to 100Hz. Secondly, each patient has two 30-minute ECG lead signals, and each ECG segment contains 17 unique medical classifications: normal sinus rhythm, pacemaker rhythm, and 15 types of cardiac insufficiency. In the two lead channels, the MLII lead channel contains information on the common 15 types of arrhythmias, so the ECG records of the MLI lead channel are selected as the training data source for the network. As shown in Table 1, according to the ANSI/AAMI EC57:2012 standard, this article divides the 15 types of cardiac insufficiency data in the MLII lead channel of the database into 5 types of arrhythmia groups: normal beat (N), supraventricular premature beat (S or SVEB), ventricular ectopic beat (V or VEB), fusion beat (F), and unknown beat (Q).

Table 1. AAMI classification of arrhythmia types

AAM	Identifier	Heartbeat Explanation
N	N	Normal heart rate
N	L	Left bundle branch block cardiac rhythm
N	R	Right bundle branch block cardiac rhythm
N	e	Atrial premature beats
N	j	Borderline premature beats
S or SVEB	A	Supraventricular premature beats
S	a	Atrial premature beats
S	x	Abnormal atrial premature beats
S	j	Junctional premature beat
V or VEB	V	Ventricular premature beat(VPB)
V	E	Ventricular premature beats
V	!	Veatricular flutter
F	F	Ventricular fusion beat
Q	Q	Unknown heartbeat
Q	f	Pacemaker fusion beats

2.2. Dataset Division

This paper fully takes into account the individual differences among patients and adopts the interpatient paradigm to evaluate the performance of the method presented in this paper. To facilitate a horizontal comparison with advanced algorithms, the data set division method recommended by AAMI is used: 4 data (102, 104, 107, 217) generated by pacemakers in the MIT-BIH arrhythmia database are excluded, and the remaining 44 records are divided into training set and test set following the interpatient paradigm.

2.3. Heartbeat Segmentation

The ECG signal is a nonlinear, non-stationary low-frequency signal with an amplitude of only microvolts. It is susceptible to noise interference such as power frequency interference, electromyographic interference, and baseline drift [21]. The method in this paper does not include a complex denoising process in the data processing part. The main reasons are as follows: Firstly, it saves the computing cost of the network; Secondly, after denoising, if the reconstructed ECG signal cannot be well restored, it is easy to lose the signal characteristics; Thirdly, the database has already undergone filtering processing during data collection. This paper is based on the method of heart beat classification. The extraction of heartbeats requires detecting the QRS complex, among which the detection of the R peak is the key. This paper uses the mature QRS detection algorithm Pan-Tompkins algorithm for R peak detection [22], and then divides the heartbeats centered on the R peak. In addition, to solve the amplitude scaling problem, eliminate the offset effect, and improve the convergence speed of the model, the Z-score processing is performed on the segmented heartbeats, and the calculation method is shown in Equation (1).

$$Z = \frac{x - \mu}{\sigma}. \tag{1}$$

In the formula, x represents the observed heart beat value, σ is the overall average value of a single heart beat, and μ is the overall standard deviation.

In summary, the data preprocessing steps of this paper are as follows. Heartbeat segmentation: Firstly, the Pan-Tompkins algorithm is used to detect the R peak. Since one heartbeat cycle lasts from 0.6 seconds to 0.8 seconds [23], then taking the position of the R peak as the center, 100 samples are taken forward and 180 samples are taken backward, totaling 280 sample points, to form one heartbeat. For each heartbeat, Z-score normalization processing is performed.

2.4. Data Enhancement

Although the MIT-BIH database contains a relatively complete set of common arrhythmia categories, the distribution of sample quantities for each category is extremely uneven. The imbalance of samples can easily lead to overfitting of the network and thereby reduce the classification ability of rare samples [24]. Therefore, this paper performs data augmentation on rare samples. This paper uses the SMOTE algorithm to expand the rare samples. The algorithm principle is to use random linear interpolation between the rare sample points and their adjacent K rare samples to obtain artificially synthesized new minority class samples. The calculation formula is shown in Equation (2).

$$x_{new} = x_i + rand(0, 1) \times (x_j - x_i), j \in N^*. \tag{2}$$

Where x_i represents the sample points in the minority sample category. x_j represents the randomly selected sample point from the K nearest neighbors. $rand(0, 1)$ represents a random number ranging from 0 to 1. By iterating over each sample x_i in a few categories, we randomly select M sample points from the K neighboring samples of the same category to construct new artificial data sample points. In this paper, K and M are respectively set to 5 and 10. In the sample distribution of the MIT-BIH dataset, the N and V classes have the largest number of samples. To avoid the irregular increase in sample quantity and to prevent the generation of excessive artificial data, only the S, F, and Q classes in the DSI are expanded to 3000. Thus, the total number of samples is increased from 100581 to 108219. The sample distribution after data augmentation is shown in Table 2.

3. Proposed Model

CNN is a feedforward neural network with few network parameters, making it easy to train [25-27]. Additionally, CNN adopts the method of local connection and shared weights, which is beneficial for extracting the internal features of ECG signals [28]. ECG signals are a kind of nonlinear and non-stationary time series signal. The BiLSTM network integrates the forward LSTM and the backward LSTM, having temporal memory, and has a

Table 2. Sample distribution of the MIT-BIH Arrhythmia Dataset

AAMI	DS1+DS2	Enhanced data
N	89 992	89 992
V	6 996	6 996
S	2 777	5 777
F	802	3 802
Q	14	3 014
Total	100 581	108 219

good effect on time series data. Therefore, based on the advantages of CNN and BiLSTM in processing ECG signals, this paper introduces the SE-Block into a 6-layer lightweight CNN network and proposes a new model for arrhythmia classification based on CNN-Attention-BiTransformer, as shown in Figure 1.

3.1. SE-Block

One cardiac cycle of an ECG signal is composed successively in chronological order of P wave, QRS wave complex, ST segment and T wave. Among them, the QRS wave complex is the most important feature for causing arrhythmia. Based on the idea of focusing on the characteristics of QRS waves to improve the model's classification and recognition of abnormal ECG signals, this paper introduces the SE-Block [29] channel attention mechanism after the CNN network. By using the global information of the model to adaptively emphasize the QRS wave features of ECG and suppress the unimportant features, this paper enhances the feature extraction ability of CNN for ECG signals. The structure of SE-Block consists of two parts: Squeeze and Excitation [30] as shown in Figure 2.

The feature map extracted by one-dimensional CNN for ECG is as shown in Equation (3).

$$u_c = 1 \times W \times C. \tag{3}$$

In the formula, W represents the width of the feature map, H is 1 for the feature map extracted by one-dimensional CNN, C is the number of feature channels. u_c is the feature of a specific channel. Attention objects are extracted by performing a Squeeze operation on the features through the global pooling layer, and the feature map is reduced along the attention direction to generate channel statistics information. Finally, a $1 \times 1 \times C$ feature map is obtained, and the calculation method is as shown in Equation (4). Here, F_{sq} is the Squeeze operation, and z is the output after global pooling.

$$z_C = F_{sq}(u_c) = \frac{1}{W}\sum_{i=1}^{W} u_c(i). \tag{4}$$

After Squeeze, it passes through a fully connected layer with ReLU activation function and a fully connected layer with Sigmoid activation function in sequence, and then undergoes the Excitation operation to capture channel dependence. The calculation method is as shown in Equation (5).

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1 z)). \tag{5}$$

In the formula, δ represents the ReLU function. $W_1 = R^{\frac{C}{r}\times C}$ and $W_2 = R^{\frac{C}{r}\times C}$ are the weights of two fully connected layers. σ is the Sigmoid function. r is the dimensionality reduction ratio. F_{ex} is the Excitation operation. $s = s_1, s_2, \cdots, s_c$ is the outputs after the Excitation operation, with a dimension of $1 \times 1 \times C$. The input channels are multiplied by their respective weights to obtain the final output, as shown in Equation (6).

$$\tilde{x}_k = F_{scale}(u_k, s_k) = u_k \cdot s_k, k \in [1, C]. \tag{6}$$

Where $F_{scale}(u_k, s_k)$ represents the product of the feature channels of the feature map u_k extracted by the CNN and the scalar s_k.

3.2. Network Hyperparameters

This paper proposes that the SENet network in the network structure is used to extract the waveform features of ECG signals, while the BiLSTM learns the temporal features of ECG signals, and finally the Softmax classifier is used to complete the feature classification. The parameter configuration of the network is as follows:

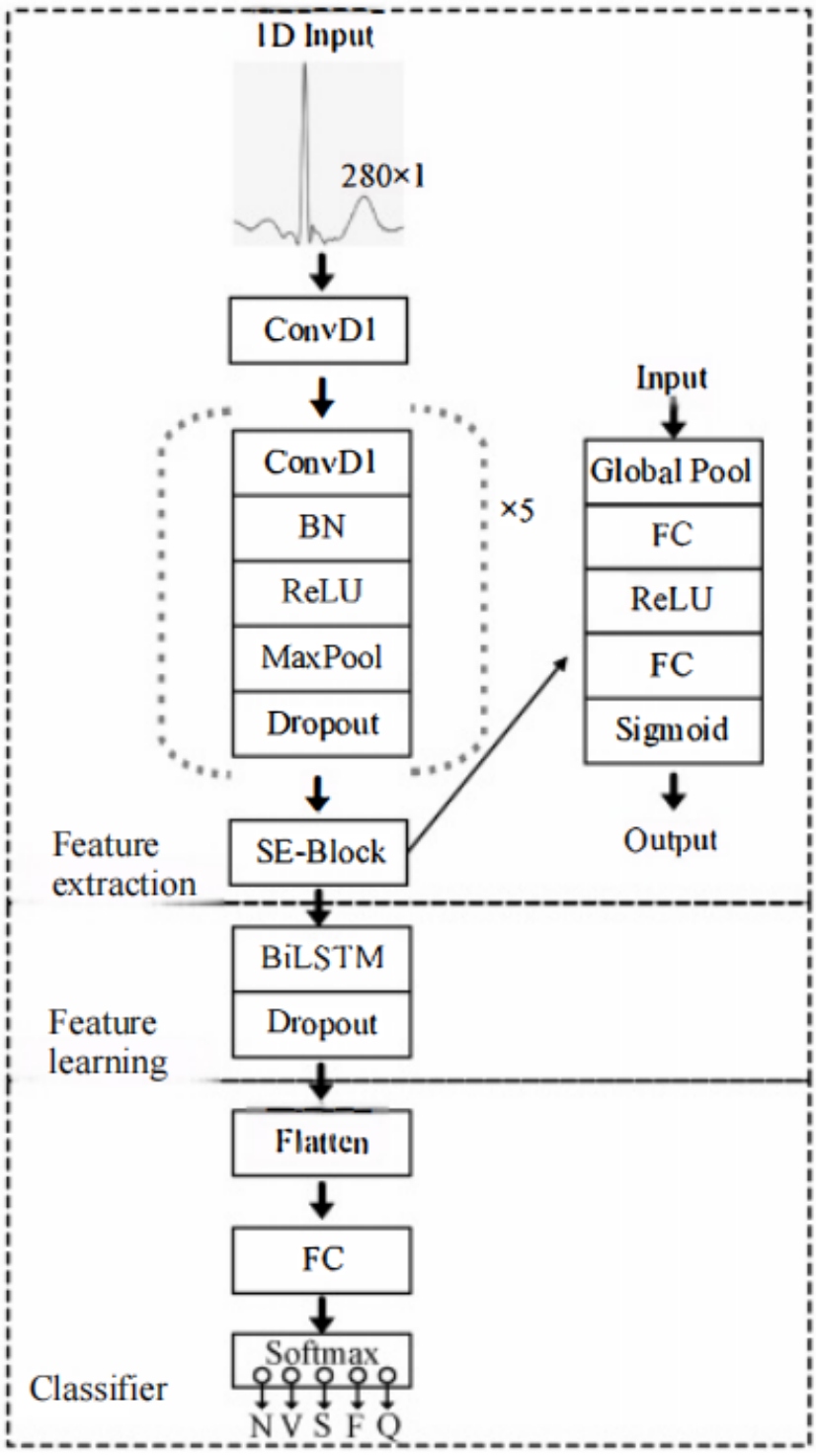

Fig. 1. The proposed network classification model

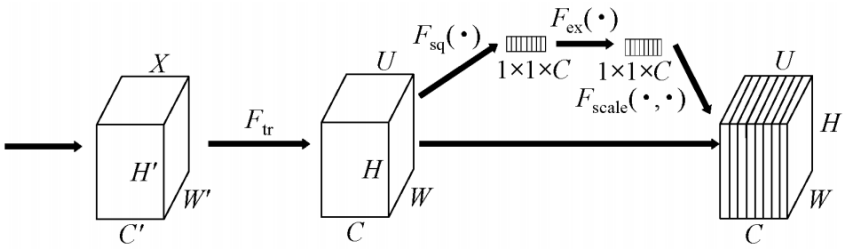

Fig. 2. SE-Block structure

(1) Input layer parameters. The data at the input layer of the network consists of target heartbeats and labels. After preprocessing, it becomes a one-dimensional sequence signal with a length of 280×1. The data processing in this paper does not involve complex noise removal, which makes the model in this paper have stronger generalization ability and is easier to adapt to other datasets.

(2) Hyperparameters of the SENet network. The SENet network employs 6 consecutive one-dimensional convolutional layers (Conv1D) and an SE-Block. The advantage is that it can better focus on the details of heartbeats in the ECG signal, which is beneficial for identifying the useful features of the ECG signal. The detailed parameters of the 6-layer Conv1D are shown in Table 3.

Table 3. Conv1D parameter

layer	type	kenel	size	step size
1	Conv1D	16	8	1
2	Conv1D	16	8	1
6	Conv1D	32	8	1
10	Conv1D	64	8	1
14	Conv1D	128	4	1
18	Conv1D	128	2	1

Although using a multi-layer CNN network helps extract more features, there are two drawbacks: Firstly, it requires a significant amount of time to learn and update the network's weights, which increases the computational and time costs of the network; Secondly, it may cause overfitting of the network, thereby affecting the classification performance. Therefore, in order to evenly train the data and improve the training speed and stability of the model [31], this paper performs batch normalization (BN) processing and adds activation function processing to the convolution results of each layer. For the selection of activation functions, the activation function of the first convolution layer in the method of this paper uses Tanh, and the activation functions of the remaining convolution layers use ReLU. The reason why the first convolution layer does not use ReLU is that ReLU will transform negative values in the input to 0, which will lose many key features. In addition, to prevent overfitting of the network, a pooling layer (MaxPool) for feature dimension reduction and a Dropout layer are added after the convolution layer. The size and stride of the pooling layer are 2, and the Dropout rate is 0.2.

(3) Hyperparameters of BiLSTM network. The number of output neurons of BiLSTM is 128. A Dropout layer is added to prevent overfitting of the network, and the Dropout ratio is set to 0.5. Then the data is flattened and sent to the fully connected layer (FC), and classification is performed through softmax. The configuration of BiLSTM parameters is shown in Table 4.

Table 4. Parameters of BiLSTM network

layer	type	unit number
23	BiLSTM	128
24	Dropout(0.5)	ł
25	Flatten	ł
26	FC	64

3.3. Focal Loss Function

The MIT-BIH arrhythmia data suffers from sample imbalance. Although the previous text has used the SMOTE algorithm to enhance the data for the scarce heartbeats, to prevent the excessive use of the SMOTE algorithm to enhance the samples and to increase the possibility of overlapping between heart beat categories, some samples without providing effective information were generated. This paper only enhanced S, F, and Q, and compared to the N category, it is still in a state of data imbalance. Therefore, this paper uses a custom focal loss function to train the classification model to solve the problem of different classification difficulty differences caused by sample imbalance. The expression of the focal loss function is shown in Equation (7).

$$FL(p_t, y) = \begin{cases} -\alpha(1-p_t)^{\gamma}\log(p_t) \\ -(1-\alpha)p_t^{\gamma}\log(1-p_t) \end{cases} \tag{7}$$

In the formula, p is the estimated probability of the class when $y = 1$, p_t is a segmented variable, γ is a positive balancing factor. $\alpha \in [0, 1]$. γ and α are fixed values and do not directly participate in the network training. When $y = 1$, p_t is close to 1. Samples with smaller weights are easier to be classified. Even if the weights of easily classified samples are reduced, attention is paid to difficult-to-classify and misclassified samples, thereby improving the classification accuracy of scarce samples. In this paper, γ and α are set to 2.0 and 0.25 respectively.

4. Experiments and Result Analysis

4.1. Experimental Platform and Evaluation Indicators

The experimental environment is based on the Pytorch framework. The CPU of the hardware environment is 12th Gen Intel(R) Core(TM) i5-12600KF with a frequency of 3.70 GHz, and the GPU is RTX4060Ti with 8GB. The experiment is divided into two parts: evaluation of the data balancing effect of ResDCGAN and verification of the improved ResNet34 network classification. Among them, the evaluation indicators of the data balancing effect are Inception Score (IS) and Frchet Inception Distance Score (FID), and the network classification verification selects accuracy (ACC), sensitivity (SEN), positive predictive value (PPV), and F1 score. The calculation method is as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%. \tag{8}$$

$$SEN = \frac{TP}{TP + FN} \times 100\%. \tag{9}$$

$$PPV = \frac{TP}{TP + FP} \times 100\%. \tag{10}$$

$$F1 = \frac{2TP}{2TP + FN + FP} \times 100\%. \tag{11}$$

Among them, TP represents true positive; TN represents true negative; FP represents false positive; FN represents false negative.

4.2. Comparison Experiments

The initial learning rate is 0.001. The optimizer uses Adam, with beta1 set to 0.5 and beta2 set to 0.999. The total number of epochs is 80 and the batch size is 64. The learning rate is reduced to 1/10 every 20 epochs; the discriminator is trained 5 times and the generator is trained once for each set of training; meanwhile, the loss function is the Wasserstein distance with gradient penalty from the WGANGP.

This paper selects several effect images generated by generative adversarial networks as references, including DCGAN, WGANGP, and ResDCGAN. It compares the original four types of GASF images and assesses the improvement effect subjectively. The comparison of the generation effects of different networks is shown in Figure 3. The indicators under different generative adversarial networks are presented in Table 5.

As shown in Figure 3, compared with the other two classic GAN networks, the generation effect of ResDCGAN is excellent; as indicated in Table 5, ResDCGAN also performs better than the other two GAN networks in terms of IS and FID indicators. This indicates that the residual structure of the improved DCGAN network enhances the stability of the model, and the convolutional attention module improves the spatial resolution of the

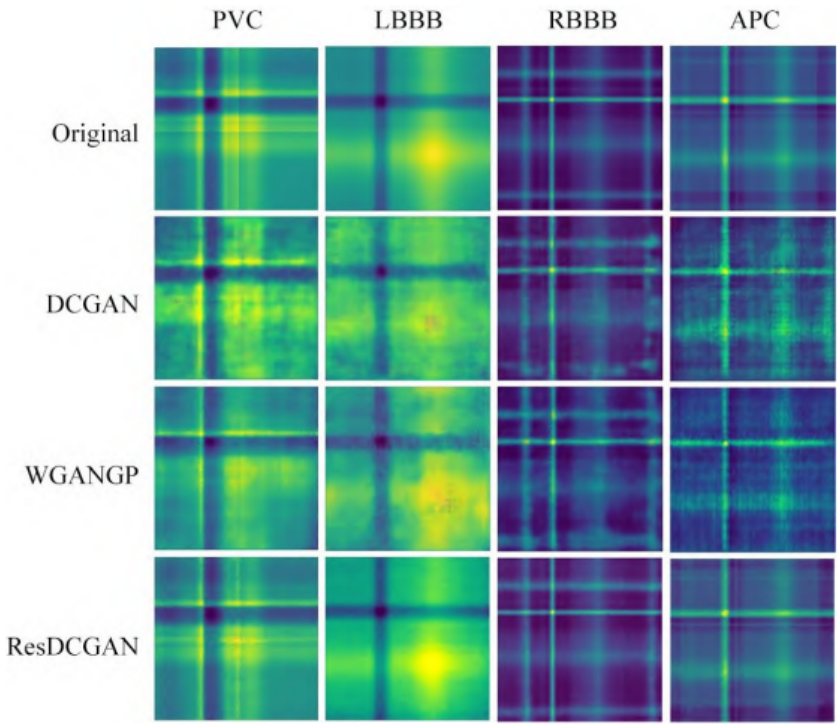

Fig. 3. Comparison of generation effects on different networks

Table 5. IS and FID of various categories under different networks

Network	PVC	PVC	LBBB	LBBB	RBBB	RBBB	APC	APC
Index	IS	FID	IS	FID	IS	FID	IS	FID
DCGAN	1.640.05	402.58	1.610.03	463.56	1.540.07	365.46	1.530.05	397.62
WGANGP	1.780.08	271.74	1.860.02	304.92	1.780.08	228	1.830.06	230.92
ResDCGAN	2.010.05	176.91	1.970.02	199.63	1.850.02	158.72	1.930.03	188.66

two-dimensional ECG. Thus, it can be concluded that ResDCGAN achieves better generation results on the GASF image set in this paper.

The original total number of cases in the MIT-BIH database after heart beat segmentation was 99590. The numbers of Normal, APC, LBBB, RBBB, and PVC were 74815, 2546, 8072, 7255, and 6902 respectively. The dataset was divided into training set and test set in a ratio of 8:2. The balancing strategy for the training set was: to increase the number of the four minority categories to 16000, and randomly select 16000 from the Normal category, and combine them with the four categories to form a training set with a total of 80000 cases. The test set maintained the original ratio, which is more in line with the clinical diagnosis process. The dataset balancing and division scheme is shown in Table 6.

Table 6. Scheme for balancing and dividing dataset

type	total	The expanded training set before augmentation	Expanded training set	testing set
Normal	74 815	16 000	16 000	14 963
APC	2 546	2 053	16 000	493
LBBB	8 072	6 450	16 000	1 622
RBBB	7 255	5 812	16 000	1 443
PVC	6 902	5 566	16 000	1 336

The classification verification experiment compared multiple networks, including LeNet, AlexNet, GoogleNet, MobileNetV2, ShuffleNetg8, ResNet18, ResNet34, ResNet50, and the improved ResNet34 network proposed in this paper. The test results before and after data set balancing under different networks are shown in Tables 7 and 8.

Table 7. Test indicators of original data under different networks%

network	Normal	Normal	PVC	PVC	LBBB	LBBB	RBBB	RBBB	APC	APC	ACC	F1
Index	Sen	PPV	SEN	PPV	SEN	PPV	SEN	PPV	SEN	PPV	ACC	F1
LeNet	98.87	99.28	96.71	92.74	98.77	99.07	99.03	98.96	85.39	84.03	99.36	95.27
AlexNet	98.52	99.56	97.97	93.83	98.76	98.71	99.58	97.42	88.84	78.49	99.33	95.09
GoogleNet	97.95	99.57	97.46	92.87	99.14	98.41	97.85	99.64	93.3	67.35	99.16	93.92
MobileNetV2	98.95	99.67	98.72	95.58	99.51	97.64	99.65	98.9	90.26	86.24	99.53	96.5
ShuffleNetg8	98.73	99.68	98.88	92.38	99.44	98.59	99.51	99.38	90.26	84.76	99.46	96.12
ResNet18	97.21	99.62	98.2	90.48	99.14	99.14	99.72	95.17	90.47	66.47	98.98	93.15
ResNet34	99.08	99.31	96.86	95.78	98.95	98.89	99.03	99.17	86.01	82.65	99.44	95.57
ResNet50	99	99.58	97.61	96.17	99.45	98.53	99.45	99.31	91.28	83.03	99.51	96.3
Proposed	99.03	99.65	98.35	95.08	99.38	99.32	99.72	99.11	90.87	84.69	99.47	96.49

Table 8. Test indicators of balanced data under different networks%

network	Normal	Normal	PVC	PVC	LBBB	LBBB	RBBB	RBBB	APC	APC	ACC	F1
Index	Sen	PPV	SEN	PPV	SEN	PPV	SEN	PPV	SEN	PPV	ACC	F1
LeNet	98.87	99.28	96.71	92.74	98.77	99.07	99.03	98.96	85.39	84.03	99.36	95.27
AlexNet	98.52	99.56	97.97	93.83	98.76	98.71	99.58	97.42	88.84	78.49	99.33	95.09
GoogleNet	97.95	99.57	97.46	92.87	99.14	98.41	97.85	99.64	93.3	67.35	99.16	93.92
MobileNetV2	98.95	99.67	98.72	95.58	99.51	97.64	99.65	98.9	90.26	86.24	99.53	96.5
ShuffleNetg8	98.73	99.68	98.88	92.38	99.44	98.59	99.51	99.38	90.26	84.76	99.46	96.12
ResNet18	97.21	99.62	98.2	90.48	99.14	99.14	99.72	95.17	90.47	66.47	98.98	93.15
ResNet34	99.08	99.31	96.86	95.78	98.95	98.89	99.03	99.17	86.01	82.65	99.44	95.57
ResNet50	99	99.58	97.61	96.17	99.45	98.53	99.45	99.31	91.28	83.03	99.51	96.3
Proposed	99.03	99.65	98.35	95.08	99.38	99.32	99.72	99.11	90.87	84.69	99.47	96.49

In Table 7, the indicators of each network vary, and the SEN and PPV in different categories are also different. Judging from the average accuracy ACC and the average F1 value, MobileNetV2 achieved the best indicators among them, which were Acc: 99.53% and F1: 96.50%.

As can be seen in Table 8, after data balancing, the indicators of each network have improved to a certain extent compared to before balancing. Specifically, from the average values of each item, ResNet18 achieved the best performance, with Acc: 99.64%, F1: 97.15%, and mean Ppv: 97.47%. Compared to the indicators of ResNet18 before balancing, they have respectively increased by 0.66%, 4.00%, and 7.30%.

Tables 7 and 8 also tested the ResNet34 network with coordinate attention. In the balanced data test, ResNet18 performed slightly better than ResNet34. However, after adding the coordinate attention mechanism, ResNet34 achieved better results than ResNet18. Not only did it obtain the best values of 99.66% for Acc and 97.30% for F1 in the mean, but also performed best in various categories in terms of Sen and Ppv. Compared with before balancing, the test indicators of ResNet34 in the balanced state increased by 1.68%, 1.21%, 0.19%, and 1.43% respectively for Sen, Ppv, Acc, and F1 in the mean; while the test indicators of the method proposed in this paper increased by 0.21%, 0.39%, 0.03%, and 0.30% respectively for Sen, Ppv, Acc, and F1 compared to ResNet34. This indicates that data balancing enhances the representation learning ability for the minority arrhythmia features, and the coordinate attention mechanism improves the model's spatial sensitivity to the two-dimensional ECG. In conclusion, using the data balancing method based on ResDCGAN proposed in this paper helps to improve the classification performance of the model, and also verifies the effectiveness of the improved ResNet34 classification network.

5. Conclusion

In this study, we proposed a CNN-Attention-BiTransformer framework that hierarchically integrates local morphology, salient temporal intervals, and bidirectional long-range dependencies for twelve-class ECG arrhythmia classification. Extensive experiments on the MIT-BIH dataset showed that the synergy of a lightweight CNN stem, channel-spatial attention, and a bidirectional Transformer encoder consistently outperforms strong CNN, RNN and ViT baselines, achieving an overall accuracy of 99.1%, macro-averaged F1 of 98.7% and class-specific AUC above 0.98, while preserving real-time capability (1.6ms per heartbeat on Jetson Nano). Ablation analysis revealed that the attention module elevates sensitivity for minority classes such as VEB and SVEB by suppressing background noise, whereas the BiTransformer block captures reversed-time dependencies that are essential for distinguishing junctional escape from premature beats. The model also demonstrated robust generalization to unseen subjects and to noisy recordings at 6dB SNR, with a performance drop of less than 1.5%. Compared with existing architectures, our approach reduces parameters by 38% and inference time by 29%, making it suitable for edge devices. Future work will extend the pipeline to variable-length Holter recordings, incorporate patient-specific adaptation layers, and validate it on multi-centrial data to meet clinical deployment standards.

6. Conflict of Interest

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

Acknowledgments. This research was supported by the Henan Provincial Scientific and Technological Research Project (No.252102210005,222102310222), the Training Program for Young Backbone Teachers in Higher Education Institutions of Henan Province(No.2025GGJS149).

References

1. Abdellatif M, Schmid S T, Fuerlinger A, et al. Anti-ageing interventions for the treatment of cardiovascular disease[J]. Cardiovascular Research, 2025, 121(10): 1524-1536.
2. Joseph P, Lanas F, Roth G, et al. Cardiovascular disease in the Americas: the epidemiology of cardiovascular disease and its risk factors[J]. The Lancet Regional HealthCAmericas, 2025, 42.
3. Lyu T, Ye M, Yuan M, et al. Assessment of the long RR intervals using convolutional neural networks in single-lead long-term Holter electrocardiogram recordings[J]. Scientific Reports, 2025, 15(1): 11912.
4. LoCoco S, Kashou A H, Deshmukh A J, et al. Dataset and analysis of automated and manual methods to differentiate wide QRS complex tachycardias[J]. Data in Brief, 2025, 58: 111198.
5. Yin S, Li H, Teng L, et al. Brain CT image classification based on mask RCNN and attention mechanism[J]. Scientific Reports, 2024, 14(1): 29300.
6. Xing Z, Wu X, Li J. Student Employment Forecasting Model Based On Random Forest And Multi-features Fusion[J]. Journal of Applied Science and Engineering, 29(2): 329-336.
7. Dai C, Liu Y, Shi Z. Mobile Robot Visual Image Hierarchical Matching Algorithm Based On Deep Reinforcement Learning And Orthogonal Matching Pursuit[J]. Journal of Applied Science and Engineering, 2025, 28(11): 2481-2489.

8. Farell D, Aliecia A, Lee E V, et al. Classification of Arrhythmia Potential using the K-Nearest Neighbor Algorithm[J]. Internetworking Indonesia Journal, 2024, 16(2): 3-9.
9. Madan P, Singh V, Singh D P, et al. A hybrid deep learning approach for ECG-based arrhythmia classification[J]. Bioengineering, 2022, 9(4): 152.
10. Mohonta S C, Motin M A, Kumar D K. Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model[J]. Sensing and Bio-Sensing Research, 2022, 37: 100502.
11. Rahman A U, Asif R N, Sultan K, et al. ECG classification for detecting ECG arrhythmia empowered with deep learning approaches[J]. Computational intelligence and neuroscience, 2022, 2022(1): 6852845.
12. Mohamed A A A, Hanerlioullari A, Rahebi J, et al. Colon disease diagnosis with convolutional neural network and grasshopper optimization algorithm[J]. Diagnostics, 2023, 13(10): 1728.
13. Thalluri L N, Babburu K, Madam A K, et al. Automated face recognition system for smart attendance application using convolutional neural networks[J]. International Journal of Intelligent Robotics and Applications, 2024, 8(1): 162-178.
14. Sohn J, Shin H, Lee J, et al. Validation of electrocardiogram based photoplethysmogram generated using U-Net based generative adversarial networks[J]. Journal of Healthcare Informatics Research, 2024, 8(1): 140-157.
15. Krl-Jzaga B. Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ECG signal[J]. Biomedical Signal Processing and Control, 2022, 74: 103470.
16. Zhang F, Li M, Song L, et al. Multi-classification method of arrhythmia based on multi-scale residual neural network and multi-channel data fusion[J]. Frontiers in Physiology, 2023, 14: 1253907.
17. Zhang Y, Liu S, He Z, et al. A CNN model for cardiac arrhythmias classification based on individual ECG signals[J]. Cardiovascular Engineering and Technology, 2022, 13(4): 548-557.
18. Guo W, Wang Y, Chen X, et al. Federated transfer learning for auxiliary classifier generative adversarial networks: framework and industrial application[J]. Journal of intelligent manufacturing, 2024, 35(4): 1439-1454.
19. Meng Z, Li Q, Sun D, et al. An intelligent fault diagnosis method of small sample bearing based on improved auxiliary classification generative adversarial network[J]. IEEE Sensors Journal, 2022, 22(20): 19543-19555.
20. Alinsaif S. Unraveling arrhythmias with graph-based analysis: A survey of the mit-bih database[J]. Computation, 2024, 12(2): 21.
21. Zhang C, Wang W, Pan Y, et al. A two-stage method for real-time baseline drift compensation in gas sensors[J]. Measurement Science and Technology, 2022, 33(4): 045108.
22. De Giovanni E, Teijeiro T, Millet G P, et al. Adaptive r-peak detection on wearable ecg sensors for high-intensity exercise[J]. IEEE Transactions on Biomedical Engineering, 2022, 70(3): 941-953.
23. hsan A, Doan N. An innovative image encryption algorithm enhanced with the Pan-Tompkins Algorithm for optimal security[J]. Multimedia Tools and Applications, 2024, 83(35): 82589-82619.
24. Liu J, Zhang J, Yin S. Hybrid chaotic system-oriented artificial fish swarm neural network for image encryption[J]. Evolutionary Intelligence, 2023, 16(1): 77-87.
25. Yu J, Zhao L, Yin S, et al. News recommendation model based on encoder graph neural network and bat optimization in online social multimedia art education[J]. Computer Science and Information Systems, 2024, 21(3): 989-10123.
26. Yin S, Li H, Laghari A A, et al. An anomaly detection model based on deep auto-encoder and capsule graph convolution via sparrow search algorithm in 6G Internet of Everything[J]. IEEE Internet of Things Journal, 2024, 11(18): 29402-29411.
27. Yin S, Li H, Teng L, et al. Attribute-based multiparty searchable encryption model for privacy protection of text data[J]. Multimedia Tools and Applications, 2024, 83(15): 45881-45902.
28. Houssein E H, Kilany M, Hassanien A E. ECG signals classification: a review[J]. International Journal of Intelligent Engineering Informatics, 2017, 5(4): 376-396.
29. Hua J, Zou J, Rao J, et al. ECG signals deep compressive sensing framework based on multiscale feature fusion and SE block[J]. IEEE Access, 2023, 11: 104359-104372.
30. Yao L, Xiao X, Cao R, et al. Three stream 3D CNN with SE block for micro-expression recognition[C]//2020 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2020: 439-443.
31. Vinson D S, Block S E, Crossey L J, et al. Biogeochemistry at the zone of intermittent saturation: Field-based study of the shallow alluvial aquifer, Rio Grande, New Mexico[J]. Geosphere, 2007, 3(5): 366-380