Mathematical Biosciences and Engineering

Research article

Arrhythmia classification based on multi-feature multi-path parallel deep convolutional neural networks and improved focal loss

Zhongnan Ran[1], Mingfeng Jiang[2,*], Yang Li[2], Zhefeng Wang[3], Yongquan Wu[3], Wei Ke[4] and Ling Xia[5]

[1] School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China
[2] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China
[3] Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, No. 2 Anzhen Road, Chaoyang District, Beijing 100029, China
[4] School of Applied Sciences, Macao Polytechnic Institute, Macao SAR, China
[5] Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China

* **Correspondence:** Email: m.jian@zstu.edu.cn; Tel: +8657186843312; Fax: +8657186843312.

Abstract: Early diagnosis of abnormal electrocardiogram (ECG) signals can provide useful information for the prevention and detection of arrhythmia diseases. Due to the similarities in Normal beat (*N*) and Supraventricular Premature Beat (*S*) categories and imbalance of ECG categories, arrhythmia classification cannot achieve satisfactory classification results under the inter-patient assessment paradigm. In this paper, a multi-path parallel deep convolutional neural network was proposed for arrhythmia classification. Furthermore, a global average RR interval was introduced to address the issue of similarities between *N* vs. *S* categories, and a weighted loss function was developed to solve the imbalance problem using the dynamically adjusted weights based on the proportion of each class in the input batch. The MIT-BIH arrhythmia dataset was used to validate the classification performances of the proposed method. Experimental results under the intra-patient evaluation paradigm and inter-patient evaluation paradigm showed that the proposed method could achieve better classification results than other methods. Among them, the accuracy, average sensitivity, average precision, and average specificity under the intra-patient paradigm were 98.73%, 94.89%, 89.38%, and 98.24%, respectively. The accuracy, average sensitivity, average precision, and average specificity under the inter-patient paradigm were 91.22%, 89.91%, 68.23%, and 95.23%, respectively.

1. Introduction

According to statistics from the World Health Organization (WHO), approximately 17.9 million people die from cardiovascular diseases every year [1]. Arrhythmia is a common clinical manifestation of cardiovascular diseases, and electrocardiogram (ECG) monitoring is an effective method for detecting and recording arrhythmias. Therefore, real-time and high-precision automatic ECG monitoring can effectively predict arrhythmias.

Nowadays, traditional machine learning methods and deep learning-based methods are the most widely used for arrhythmia classification methods. Traditional machine learning methods are usually implemented for ECG classification using sample feature-based learning methods, such as support vector machine (SVM) [2], random forest [3], k-nearest neighbor classifier (k-NN) [4], decision tree [5], Bayesian classifier [6], etc. For example, Faziludeen and Sabiq [7] proposed a new method that utilizes Daubechies 4 wavelet decomposition for heartbeat feature selection, and 25 features were extracted from each heartbeat through wavelet analysis, and then SVM method was used to classify them. In addition, Li and Zhou [8] used the entropy of wavelet packet decomposition (WPD) coefficients as representative features and subsequently employed random forest (RF) for ECG classification. Venkatesan et al. [9] proposed an adaptive filter with the delay error normalized least mean squares (LMS) algorithm to address the issue of prolonged processing time for ECG signals, which achieved high speed and good classification performances. However, in these traditional machine learning methods, the classification performance often heavily depends on the quality of feature engineering, and the normal beat (N) and supraventricular premature beat (S) classes often represents similar wave form with similar features. Therefore, it is difficult to classify them accurately using conventional approaches.

Among deep learning-based methods, one-dimensional convolutional neural network (1-D CNN) is widely used for the classification of cardiac arrhythmias. Acharya et al. [10] proposed a 9-layer deep convolutional neural network (CNN) which can successfully classified 5 different categories of ECG signals by eliminating high-frequency noise and data augmentation. Hannun et al. [11] conducted a comprehensive evaluation of end-to-end deep learning methods for ECG classification, and achieved high diagnostic performance similar to that of cardiologists. Mousavi and Afghah [12] proposed a novel and effective ECG-based automatic heartbeat classification method, which addresses the challenge of ECG data imbalance by utilizing a sequence-to-sequence deep learning method and an oversampling technique algorithm. Sabut et al. [13] proposed a deep neural network (DNN) approach based on mixed time-frequency features, which can enhance the diagnostic efficiency of ventricular tachycardia (VT) and ventricular fibrillation (VF). In addition, recurrent neural networks (RNN) [14] was a promising way to capture the forward and backward dependencies of time series. Lynn et al. [15] proposed a bidirectional gated recurrent unit (GRU) method to implement ECG classification effectively. Yildirim [16] proposed deep bidirectional long short-term memory (LSTM) networks and wavelet sequences for ECG classification, achieving good classification performances.

Most ECG classification methods focused on the evaluation under an intra-patient assessment paradigm. This intra-patient evaluation paradigm often neglected the differences of ECG signals from different patients, thereby limiting its generalization ability. The inter-patient assessment, as used in

the clinical application scenarios, often meets these problems such as data imbalance and inter-individual variability, which can deteriorate the classification accuracy and generalization capability of the model. To address the above issues, we propose a novelty method for ECG classification as following:

1) A multi-path parallel deep convolutional neural network is proposed for arrhythmia classification by extracting the features of *P* wave, *QRS* wave group, and *T* wave simultaneously, and a global average RR interval is introduced to address the issue of data similarity between normal beat (*N*) and supraventricular premature beat (*S*).

2) A weighted loss function has been developed to dynamically adjust weights based on the proportion of each class in the input batch, so as to overcome the data imbalance problem.

3) Both intra-patient and inter-patient evaluation paradigms are used to validate the classification performances of the proposed method, and the experimental results showed that the proposed method can achieve the most accurate classification performances among these state-of-the-art methods.

The rest of this paper is organized as follows: Section 2 introduces the ECG datasets and the proposed method. Section 3 outlines the experimental setup and comparison between the proposed method and state-of-the-art methods. Section 4 discusses the performance of the proposed model and conducts ablation experiments to verify the effectiveness of the model. Section 5 summarizes and looks forward to this paper.

2. Materials and methods

2.1. ECG database

In this paper, the MIT-BIH arrhythmia database [17] is used to validate the proposed method for ECG classification. The MIT-BIH Arrhythmia Database is an open-source dataset provided by the Massachusetts Institute of Technology (MIT) and recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [18]. It contains 48 dual-channel ECG records collected from 47 patients, with a duration of about half an hour. Each recording consisted of two leads (i.e., Leads I and II), sampled in the range of 10 mV per channel at a sampling frequency of 360 Hz, and each recording was individually annotated by two cardiologists (total 110,000 notes). In this paper, Lead II (MLII) ECG datasets was used for training and testing the proposed method.

According to the standards from AAMI, the beats of ECG signals can be categorized into five categories for the classification of arrhythmias, as shown in Table 1. Because AAMI defines the Q-type cardiac beat as an unknown category, which is for reference only and has no specific application value, in this paper, we primarily utilize the first four categories as experimental data.

In this paper, intra-patient and inter-patient experiments are used to validate the classification performances of the proposed method. In the intra-patient evaluation approach, the training and testing sets are collected together with a total of 100,501 heartbeats from different patients, and the training and testing samples are partitioned based on heartbeats. Additionally, we utilize a 10-fold cross-validation method to validate the model. We divide all heartbeats into 10 equal parts, sequentially selecting 9 parts as the training set and the rest 1 part as the testing set. This process is repeated 10 times. However, there are some limitations to this evaluation method, because some heartbeat samples may come from the same patient in both the training and testing sets. In the inter-patient evaluation approach, we followed De Chazal et al. [19] definition of dataset partitions and divided it into DS1 and DS2 datasets, as shown in Table 2. Furthermore, DS1 was used for model training, while DS2 was

used for model testing. In this scheme, different numerical identifiers represent different patients' ECG records. It fully accounts for individual diversity and can effectively evaluate the universality and robustness of the proposed method.

Table 1. AAMI classification criteria for cardiac arrhythmias.

AAMI heartbeat class	N	S	V	F	Q
MIT-BIH heartbeat types	Normal beat	Atrial premature beat	Premature ventricular contraction	Fusion of ventricular and normal beat	Paced beat
	Left bundle branch block beat	Aberrated atrial premature beat	Ventricular escape beat		Fusion of paced and normal beat
	Right bundle branch block beat	Nodal (junctional) premature beat			Unclassifiable beat
	Atrial escape beat	Supraventricular premature beat			
	Nodal (junctional) escape beat				

Table 2. ECG data sample division.

Dataset	ECG recording
DS1	'101', '106', '108', '109', '112', '114', '115', '116', '118', '119','122', '124', '201', '203', '205', '207', '208', '209', '215', '220','223', '230'
DS2	'100', '103', '105', '111', '113', '117', '121', '123', '200', '202','210', '212', '213', '214', '219', '221', '222', '228', '231', '232','233', '234'

2.2. ECG data preprocessing

For a given ECG signal, the ECG preprocessing procedure includes as follows:

1) ECG heartbeat segmentation. According to the annotations made by experts on the R-peak and heartbeat categories of the ECG signal, the R-peak position of each heartbeat in the data is obtained. Since the sampling frequency of the ECG acquisition equipment is 360 Hz, and the cardiac cycle is about 0.8 s, we use a fixed window of length 255 to capture the heartbeat waveform. For each heartbeat, we take the preceding 128 points and the following 127 points from the R peak as a sample.

2) Eliminating data instability. Considering that wearing and taking off the instrument will cause severe heartbeat fluctuations, we removed the first 10 and last 10 heartbeats to ensure the stability of the heartbeat.

2.3. Dynamic weighted focal loss

The distribution of beats in different categories of the MIT-BIH arrhythmia database is presented in Table 3. It can be found that the heartbeat distribution of different categories is seriously imbalanced. Moreover, unbalanced data often leads to overfitting in categories with a large number of samples.

Therefore, a dynamic weighted focal loss is proposed to address the data imbalance problem, which can adjust the weight dynamically using the number of heartbeats of different categories in the input batch.

Table 3. Beat distribution of different categories in the MIT-BIH arrhythmia database.

Heart class	DS1	DS2	Total
N	45,490	43,885	89,375
S	929	1822	2751
V	3770	3196	6966
F	412	388	800

First, we define a label set $Batch\_label_i$ as show in Eq (1), representing X heartbeats contained in the i-th training batch, where y_j denote the j-th beat label in $Batch\_label_i$, $y_j \in \{N, S, V, F\}$, and each category corresponds to a different one-hot encoding, such as $N = [1, 0, 0, 0]$, $S = [0, 1, 0, 0]$.

$$Batch\_label_i = \{y_1, y_2, y_3, \ldots, y_X\} \tag{1}$$

Therefore, we define the weight of the i-th batch as α_i, as shown in Eq (2), assigning different weights based on the number of different categories in each batch. The larger the proportion of a certain category in each batch, the lower the weight obtained. In addition, the smaller the proportion, the greater the weight obtained, of which 0.05 is to prevent the weight from reaching 0.

$$\alpha_i = \frac{X - \sum_{i=1}^{X} y_i}{2} + 0.05 \tag{2}$$

In addition, the i-th batch weighted focal loss is defined as L_i, as shown in Eq (3), where $\hat{y}$ represents the predicted probability of each class and y is the one-hot encoding of the ground truth label.

$$L_i = \alpha_i (1 - \hat{y}) y log \hat{y} \tag{3}$$

2.4. Global average RR interval feature extraction

When performing classification tasks, convolutional networks often struggle to learn the differences between N-class and S-class waveforms due to their wave similarities. Additionally, N-class data is much more abundant than S-class data, which can further degrade the generalization of model. However, compared to N-class data, S-class data typically exhibit incomplete compensatory intervals, indicating shorter forward RR intervals. The forward RR interval represents the time interval between the current heartbeat's R-wave and the R-wave of the previous heartbeat. Therefore, in this paper, we introduce RR intervals to assist the network in classification. The definition of the forward RR interval is illustrated in Eq (4), where i represents the i-th RR interval and j represents the j-th patient.

$$RR_j(i) = R(i) - R(i - 1) \tag{4}$$

In this paper, we adopted an overall average normalization method, given that the ventricular beat rate varies among different patients, resulting in variable RR intervals. First, the average of all RR intervals for j-th patient is defined as $RR_{aver,j}$, as shown in Eq (5). With j-th patient's $RR_{aver,j}$ as

the standard for this patient, the forward RR interval divided by the patient's $RR_{aver,j}$ is defined as $RR_{global,j}(i)$. The $RR_{global,j}(i)$ represents the global average RR interval, as shown in Eq (6).

$$RR_{aver,j} = \frac{1}{N}\sum_{i=0}^{N} RR_j(i) \tag{5}$$

$$RR_{global,j} = \frac{RR_j(i)}{RR_{aver,j}} \tag{6}$$

The global average normalization method standardizes RR intervals across different patients, facilitating data comparison. Additionally, due to the limited number of features, four transformations are used to expand the global average RR interval feature space, as defined in Eqs (7)–(10). In this study, five features ($RR_{global,j}(i)$, RR_1, RR_2, RR_3, RR_4) are used to train the network for ECG classification tasks.

$$RR_1 = RR_{global,j}(i)^2 \tag{7}$$

$$RR_2 = e^{RR_{global,j}(i)} \tag{8}$$

$$RR_3 = e^{RR_{global,j}(i)-1} \tag{9}$$

$$RR_4 = ln(RR_{global,j}(i)) \tag{10}$$

In order to present the data distribution more intuitively, we use Gaussian kernel density estimation to analyze the relationship between the global average RR interval and variants, as shown in Figure 1. In the figure, the horizontal axis represents the global average RR interval. The vertical axis represents the probability density of occurrence. Figure 1 illustrates that in both the training set and the testing set, the global average RR interval can make the N class and the S class be linearly separable.
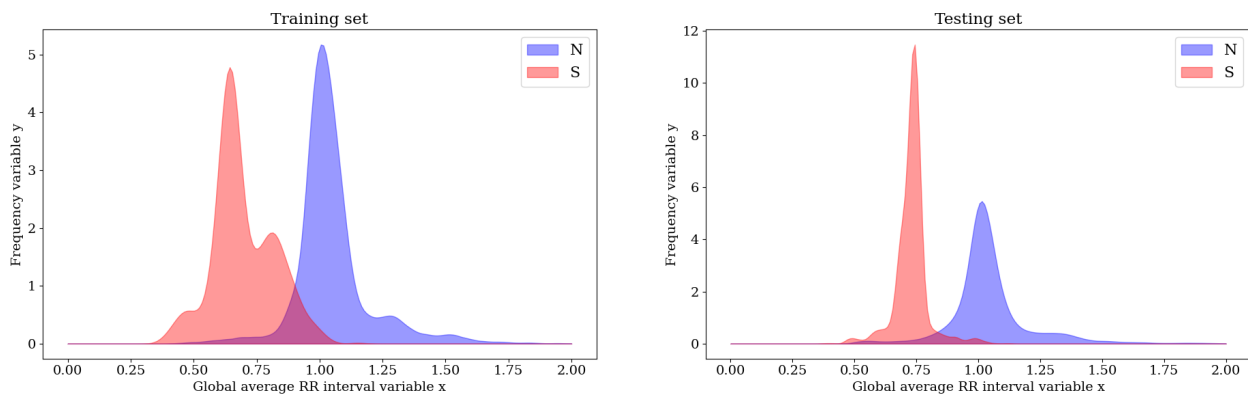

Figure 1. Probability density distribution of the global mean RR interval for the training set and testing set.

2.5. The multi-path parallel deep convolutional neural network

We propose a multi-path parallel deep convolutional neural network with strong generalization ability, as shown in Figure 2. The network is mainly composed of 20 convolutional layers, 20 activation function layers, 20 batch normalization layers [20], 8 dropout layers, 3 MaxPool layers, 2 fully connected layers, 1 LSTM layer, and 1 softmax layer. The following are details about the model structure.
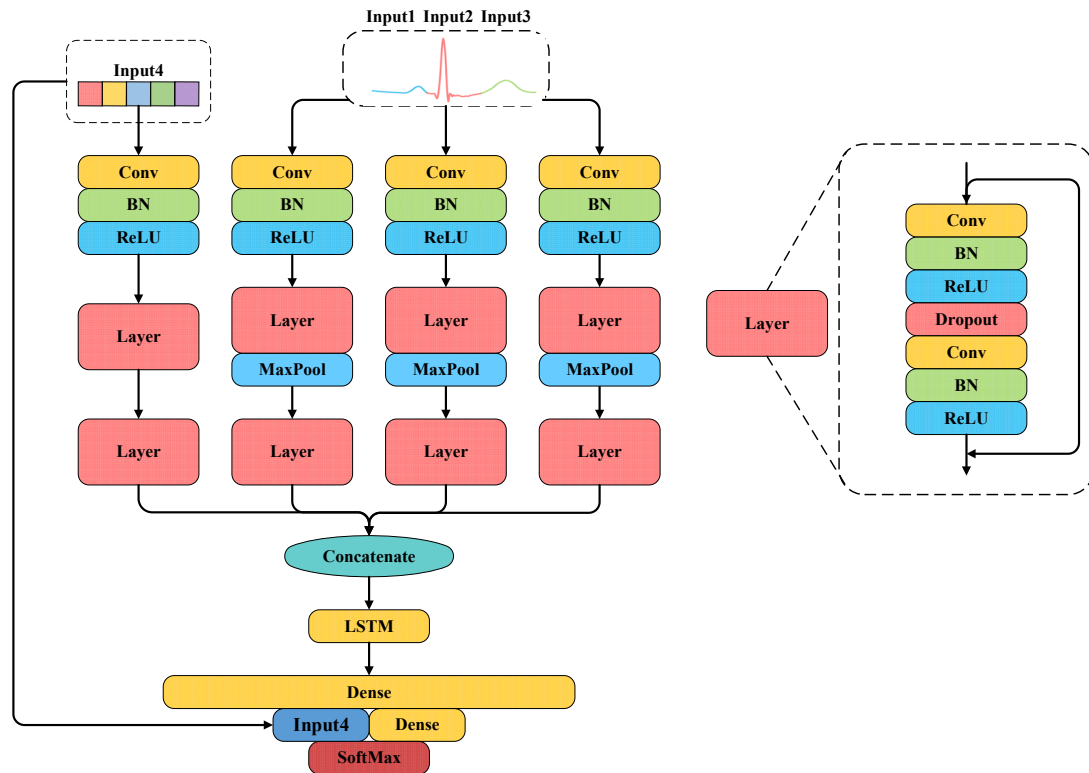
Figure 2. Multi-path parallel deep convolutional neural network.

Input: In this paper, we divided the ECG signal into 3 segments, each with a length of 85, and fed them into three identical parallel networks to extract the features of the *P* wave, *QRS* wave group, and *T* wave, respectively. Therefore, the network can focus on features extraction from different periods simultaneously. In addition, we adopted the global average RR interval feature and its four variants as the fourth input of the network to help the network perform classification. We use a convolutional network for feature extraction on this feature, mainly to increase the dimension of the feature, thereby increasing its weight in network decision-making and classification.

Convolutional Blocks: In our network, we define the initial three layers as convolutional blocks for extracting initial features at different time periods. This module consists of a convolutional layer, a batch normalization layer, and a ReLU activation function layer. The convolutional layer consists of 16 convolutions with a kernel size of 5×1.

Layer structure: Compared to the traditional convolutional neural network (CNN), the residual network achieves higher classification accuracy and overcomes the issue of information loss [21]. In this paper, we design a residual structure. The structure initially conducts feature extraction through 32 1-D convolutions with a convolution length of 5×1 and a stride of 1. Then, batch normalization is applied to the extracted features. The ReLU activation function is used to prevent gradient vanishing and saturation. Next, a dropout layer [22] randomly deactivates 20% of the neurons to prevent overfitting. Finally, feature extraction is performed again through 32 1-D convolutions with a convolution length of 7×1 and a stride of 1. The extracted features are then normalized and activated. It is worth noting that the residual block used to extract RR interval features, the convolution length as 1×1.

LSTM layer: LSTM boasts a distinctive gate mechanism internally, endowing it with long-term memory capabilities and addressing concerns related to gradient vanishing and exploding [23]. Hence,

following the concatenation of the extracted features, we employed LSTM to discern critical information among features, with the LSTM units set to 64.

Fully connected layer: After learning the dependencies between sequences, we utilized a fully connected layer with 64 units and ReLU activation function for feature mapping. Subsequently, we concatenate the RR interval features with another fully connected layer containing 5 units to enhance its classification decision-making capability. Finally, we mapped the obtained 10 features to a fully connected layer with 4 units and utilized the softmax activation function to compute the probability of each class.

3. Experimental results

3.1. Experiment setting

In this paper, the experiment was conducted on a server with an NVIDIA GeForce RTX 3090 graphics card and Ubuntu 20.04 operating system. The loss function adopts dynamic weighted focal loss.

The gradient descent optimizer uses the Adam optimizer [24] with a learning rate of 0.001. The batch size of ECG data is set to 16 in the multi-path parallel deep convolutional neural network.

3.2. Metrics

To evaluate the performance of the proposed method, we applied several widely used metrics. The detailed definitions are given as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{11}$$

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

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

$$specificity = \frac{TN}{FP+TN}, \tag{14}$$

where TP is the number of positive samples predicted to be positive. TN is the number of negative samples predicted as negative. FN is the number of positive samples predicted as negative. FP is the number of negative samples predicted to be positive.

3.3. Experiment results

Most of arrhythmia classification methods use the MIT-BIH ECG database as the research object. However, some researchers use only an intra-patient evaluation paradigm, neglecting the differences in ECG signals among individuals. In this paper, both intra-patient and inter-patient methods are used to evaluate the proposed method for ECG classification.

3.3.1. Intra-patient based model evaluation

In the intra-patient evaluation method, the training and testing sets involve non-repetitive random sampling of heartbeats from all patients. In the training and testing sets, there may be heartbeats of the same category from the same patient, and a 10-fold cross-validation method is used to evaluate the model's generalization ability. The 100,501 heartbeats in the DS1 + DS2 dataset were randomly shuffled and divided into 10 parts, with 9 parts ware used for model training, and 1 part was used for model testing. After repeating this process 10 times, the results of all model classification performances were aggregated. The resulting confusion matrix is presented in Table 4.

Table 4. Confusion matrix under the intra-patient assessment paradigm.

Predicted label / True label	N	S	V	F
N	**88,464**	555	238	118
S	135	**2603**	13	0
V	38	23	**6850**	55
F	49	3	47	**701**

Table 5. Comparison of methods using the intra-patient assessment paradigm.

Methods	ACC	N			S			V			F			Q		
	%	SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC
Acharya et al. [10]	94.03	91.54	87.43	96.71	90.59	94.30	98.63	94.22	95.30	98.84	96.06	94.76	98.67	97.75	98.73	99.69
Kachuee et al. [25]	95.90	97.00	-	-	89.00	-		96.00	-	-	86.00	-	-	98.00	-	-
Romdhane et al. [26]	98.41	99.49	98.96	-	77.88	87.65		94.54	95.73	-	82.10	83.65	-	98.51	99.31	-
Pandey et al. [27]	98.58	99.54	99.44	-	92.00	91.02	-	95.81	96.80	-	80.55	85.22	-	-	-	-
Shoughi et al. [28]	98.71	99.01	99.60	98.01	92.74	82.82	99.58	97.35	94.57	99.63	83.33	87.50	99.91	99.75	98.54	99.88
Proposed method	**98.73**	**98.98**	**99.75**	**96.55**	**94.62**	**81.75**	**98.84**	**98.33**	**95.83**	**98.75**	**87.63**	**80.21**	**98.81**	-	-	-

From Table 5, compared with state-of-the-art methods using the intra-patient assessment paradigm, it can be found that the proposed method outperforms other methods in terms of classification performances, especially in the minority categories (*S, V, F*). Acharya et al. [10] achieved the best performance in the F category, because data augmentation method was used to generate the F samples. In the methods of Romdhane and Pr [26], Pandey and Janghel [27], and Shoughi and Dowlatshahi [28], these methods can get better performances in the *N* category than the proposed method. However, the sensitivities of the proposed method in the *S, V*, and F categories were larger than those of the other three methods, which has more important value in the clinical Arrhythmia Classification. According to the experimental results, it can be found that the weighted focal loss method can address the data imbalance problem effectively and improve the classification performances of the minority.

3.3.2. Inter-patient based model evaluation

In the inter-patient evaluation method, the training set and the testing set are divided based on

different patients. The training set and testing set extracted by this method will be completely independent from different patients. Due to data imbalance and wave similarities between the *N* class and *S* class, the ECG classification becomes more challenging using inter-patient than using intra-patient assessment.

Due to the difficulty of classification in the inter-patient evaluation method, most approaches only focus on the classification of normal beats (*N*), supraventricular ectopic beats (*S*), and ventricular ectopic beats (*V*), without considering the minority class of fusion beats (*F*). To compare other methods using this evaluation paradigm, we just consider for implementing classification experiments of *N*, *S*, and *V*. The confusion matrix resulting from the experiments is shown in Table 6.

Table 6. Confusion matrix under the inter-patient assessment paradigm.

Predicted label True label	N	S	V
N	**40,042**	2557	1286
S	219	**1510**	93
V	68	72	**3056**

Table 7. Comparison of methods using the inter-patient assessment paradigm.

Methods	ACC %	N			S			V		
		SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC
Garcia et al. [29]	92.40	94.00	98.00	-	62.0	53.0	-	87.3	59.4	-
Takalo-Mattila et al. [30]	89.92	91.89	97.00	-	62.49	55.86	-	89.23	50.89	-
Li et al. [31]	88.99	94.54	93.33	80.08	35.22	65.88	98.83	88.35	79.86	94.92
Siouda et al. [32]	93.11	99.14	93.78	-	13.27	69.41	-	56.13	82.57	-
Proposed method	**91.22**	**91.24**	**99.29**	**94.28**	**82.88**	**36.48**	**94.42**	**95.62**	**68.91**	**96.98**

As shown in Table 7, compared with other methods, the proposed method achieves the best performances in the minority *S* and *V* classes, which can detect the abnormal ECG signals accurately. In terms of the overall accuracy, the methods proposed by Garcia et al. [29] and Siouda et al. [32] get slightly better results than the proposed method, because these methods achieve high sensitivity in the *N* category, resulting in overall good accuracy. However, these methods exhibit a high misdiagnosis rate for the *S* and *V* categories. In all, the proposed method can address the data imbalance and wave similarities between the *N* class and *S* class problem effectively, so as to improve the classification performances in the minorities, such as the *S* and *V* categories.

4. Discussion

To validate the effectiveness of the global average RR interval feature and the dynamic weighted Focal Loss function, we designed relevant ablation experiments in the inter-patient paradigm. The experiments utilized the global average RR interval feature, Focal Loss (FL), dynamic weighted Focal Loss function (DFL), and Cross-Entropy Loss function (CE) as experimental variables to investigate the ablation study. We conducted a total of six comparison experiments, denoted as: 1)

Using CE loss function, 2) using global average RR interval feature with CE loss function, 3) using FL loss function, 4) using global average RR interval feature with FL loss function, 5) using FL loss function, and 6) using global average RR interval feature with DFL loss function.

As shown in Table 8, when combining DFL and the global average RR interval features, the proposed method can solve the overfitting problem effectively caused by the imbalance between data and the similarity between N and S classes. Although in terms of overall accuracy, the proposed method is not the best one with 91.22%, smaller than the method with CE loss with 94.35%. Because N class accounts for 90.63% of the total training data, the network tends to overfit to N class, leading to a high accuracy rate for N class, which directly impacts the overall accuracy. However, it is not useful to help detect the abnormal ECG signals, resulting in a lower classification accuracy for S and V class.

Table 8. Evaluation indicators under different methods.

CE	FL	DFL	RR	Class	AVG SEN	SEN	PPV	SPEC	ACC
√				N	59.45	99.40	94.63	50.64	94.35
				S		0.55	17.24	99.90	
				V		78.41	91.76	99.51	
√			√	N	72.29	96.54	97.01	73.99	93.43
				S		37.87	46.56	98.32	
				V		82.45	71.37	97.69	
	√			N	60.32	96.04	94.80	53.95	91.61
				S		4.28	4.84	96.74	
				V		80.63	90.93	99.44	
	√		√	N	89.79	91.08	99.16	93.22	91.02
				S		84.19	37.78	94.63	
				V		94.09	66.34	96.66	
		√		N	65.00	93.65	95.64	61.65	90.37
				S		10.54	8.01	93.46	
				V		90.80	82.14	90.34	
		√	√	N	89.91	91.24	99.29	94.28	91.22
				S		82.88	36.48	94.42	
				V		95.62	68.91	96.98	

From Table 8, it can be found that the classification performances of S classes improve obviously using the global average RR interval features, which demonstrates that the global average RR interval features aid to distinguish between classes N and S effectively. When compared with CE loss function, it can be found that the FL and DFL can address data imbalance problem effectively with better classification accuracy. However, the parameter selection of FL method is a tricky problem by parameter sweeping. In contrast, the DFL can adjust the weight factors dynamically based on the number of different classes in the input without multiple experiments. When the FL method achieved the optimal wight factors, as [0.25, 8, 1], it can get approximate classification accuracy as DFL method. In contrast, the dynamic weighted focal loss (DFL) can effectively suppress the data imbalance problem and pay a key role to a few categories with larger weight factors.

To better illustrate the advantages of DFL and RR intervals, we plotted the corresponding electrocardiograms as shown in Figure 3. In Figure 3(a), we can clearly observe that the waveforms of the *N* class and *S* class are highly similar, and the difference between the *N* class and *S* class is small. Therefore, during the classification task, without RR intervals, the network often finds it challenging to learn the differences between them. As shown in Figure 3(b), compared to the *N* class, the waveform of the *V* class tends to exhibit morphological abnormalities. However, due to data imbalance, the classification network is predominated by the *N* class. After using the DFL method, the proposed method can address the data imbalance effectively, so as to detect the *V* class.

(a)   (b)
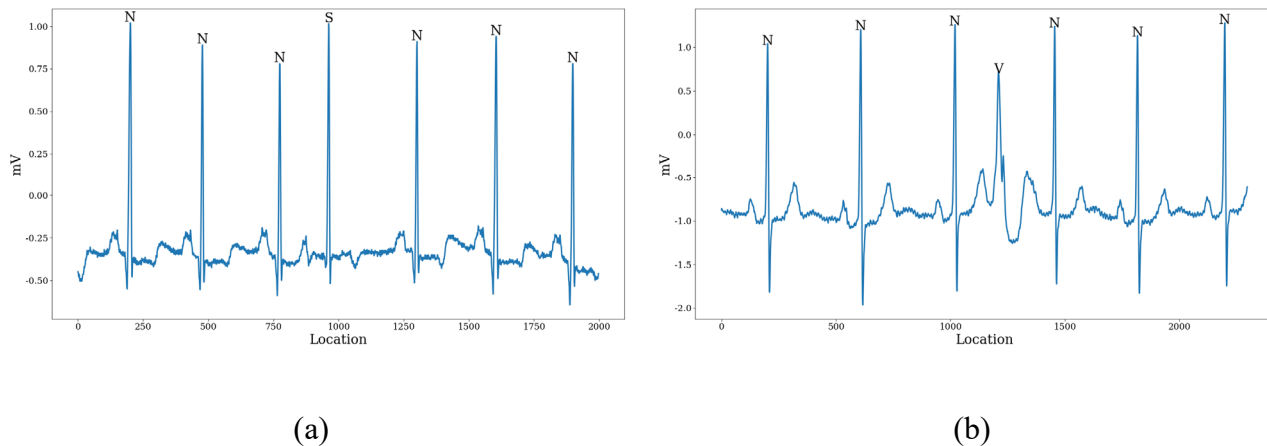
Figure 3. Heartbeat display of different categories.

To evaluate the impact of global average RR interval features on classification performance under different input conditions, we conducted relevant ablation experiments. As shown in Figure 2, we utilized two distinct input methods simultaneously for the global average RR interval features: 1) Extracting global average RR interval features using convolutional networks, and 2) concatenating global average RR interval features directly at the final fully connected layer. Consequently, we defined three experimental conditions: A1 without global average RR interval features; A2 with the global average RR interval features using convolutional networks; A3 with global average RR interval features directly at the final fully connected layer; and the proposed method used both the global average RR interval features.

Table 9. Evaluation metrics under different input methods.

Methods	ACC	N			S			V		
	%	SEN	PPV	SPEC	SEN	PPV	SPEC	SEN	PPV	SPEC
A1	90.37	93.65	95.64	61.65	10.54	8.01	93.46	90.80	82.14	90.34
A2	89.31	91.02	97.17	76.56	39.74	19.67	93.72	93.96	73.23	97.60
A3	84.46	86.41	96.10	69.37	24.97	7.58	88.22	91.58	84.96	98.87
Proposed method	91.22	91.24	99.29	94.28	82.88	36.48	94.42	95.62	68.91	96.98

It can be seen from Table 9 that under different input methods, using the global average RR interval feature can improve the decision-making ability of the network. Using two input methods at

the same time can significantly improve the sensitivity, accuracy, and specificity of the S-class classification. Therefore, using global average RR interval features can solve the problem of similarity between N-class and S-class data effectively.

5. Conclusions

In this paper, we propose a multi-path parallel deep convolutional neural network (MPP-CNN) and a dynamically weighted focal loss function for arrhythmia classification. In addition, the global average RR interval features are proposed to solve the problem of data similarity through two input methods. Furthermore, a dynamic weighted loss function has been developed to solve the imbalance problem by adjusting weights dynamically based on the proportion of each class in the input batch. Intra-patient and inter-patient experimental results show that the proposed method can achieve good classification performance, especially for the minority. In future work, the proposed method will be put into clinical diagnosis for arrhythmia classification.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

Ling Xia is a guest editor for [Mathematical Biosciences and Engineering] and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. *World Health Organization*, Cardiovascular diseases (CVDs), WHO Cardiovascular Diseases Factsheet 2021, 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

2. O. M. A. Ali, S. W. Kareem, A. S. Mohammed, Evaluation of electrocardiogram signals classifycation using CNN, SVM, and LSTM algorithm: A review, in *2022 8th International Engineering Conference on Sustainable Technology and Development (IEC)*, IEEE, (2022), 185–191. https://doi.org/10.1109/IEC54822.2022.9807511

3. P. S. Ignacio, J. A. Bulauan, J. R. Manzanares, A topology informed random forest classifier for ECG classification, in *2020 Computing in Cardiology*, IEEE, (2020), 1–4. https://doi.org/10.22489/CinC.2020.297

4. B. Venkataramanaiah, J. Kamala, ECG signal processing and KNN classifier-based abnormality detection by VH-doctor for remote cardiac healthcare monitoring, *Soft Comput.*, **24** (2020), 17457–17466. https://doi.org/10.1007/s00500-020-05191-1

5. S. Sahoo, A. Subudhi, M. Dash, S. Sabut, Automatic classification of cardiac arrhythmias based on hybrid features and decision tree algorithm, *Int. J. Autom. Comput.*, **17** (2020), 551–561. https://doi.org/10.1007/s11633-019-1219-2

6. S. T. Aarthy, J. L. M. Iqbal, Retracted article: Time series real time naive bayes electrocardiogram signal classification for efficient disease prediction using fuzzy rules, *J. Ambient Intell. Hum. Comput.*, **12** (2021), 5257–5267. https://doi.org/10.1007/s12652-020-02003-0

7. S. Faziludeen, P. V. Sabiq, ECG beat classification using wavelets and SVM, in *2013 IEEE Conference on Information & Communication Technologies*, IEEE, (2013), 815–818. https://doi.org/10.1109/CICT.2013.6558206

8. T. Li, M. Zhou, ECG classification using wavelet packet entropy and random forests, *Entropy*, **18** (2016), 285. https://doi.org/10.3390/e18080285

9. C. Venkatesan, P. Karthigaikumar, R. Varatharajan, A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection, *Multimedia Tools Appl.*, **77** (2018), 10365–10374. https://doi.org/10.1007/s11042-018-5762-6

10. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, et al., A deep convolutional neural network model to classify heartbeats, *Comput. Biol. Med.*, **89** (2017), 389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022

11. A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, et al., Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.*, **25** (2019), 65–69. https://doi.org/10.1038/s41591-018-0268-3

12. S. Mousavi, F. Afghah, Inter-and intra-patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach, in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, (2019), 1308–1312. https://doi.org/10.1109/ICASSP.2019.8683140

13. S. Sabut, O. Pandey, B. S. P. Mishra, M. Mohanty, Detection of ventricular arrhythmia using hybrid time–frequency-based features and deep neural network, *Phys. Eng. Sci. Med.*, **44** (2021), 135–145. https://doi.org/10.1007/s13246-020-00964-2

14. B. Lindemann, T. Müller, H. Vietz, N. Jazdi, M. Weyrich, A survey on long short-term memory networks for time series prediction, *Procedia CIRP*, **99** (2021), 650–655. https://doi.org/10.1016/j.procir.2021.03.088

15. H. M. Lynn, S. B. Pan, P. Kim, A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks, *IEEE Access*, **7** (2019), 145395–145405. https://doi.org/10.1109/ACCESS.2019.2939947

16. Ö. Yildirim, A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification, *Comput. Biol. Med.*, **96** (2018), 189–202. https://doi.org/10.1016/j.compbiomed.2018.03.016

17. G. B. Moody, R. G. Mark, The impact of the MIT-BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.*, **20** (2001), 45–50. https://doi.org/10.1109/51.932724

18. American National Standard, *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, AAMI, 2013. https://doi.org/10.2345/9781570204784.ch1

19. P. De Chazal, M. O'Dwyer, R. B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.*, **51** (2004), 1196–1206. https://doi.org/10.1109/TBME.2004.827359

20. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, (2015), 448–456.

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

22. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.*, **15** (2014), 1929–1958.

23. K. Smagulova, A. P. James, Overview of long short-term memory neural networks, in *Deep Learning Classifiers with Memristive Networks: Theory and Applications*, Springer, (2020), 139–153. https://doi.org/10.1007/978-3-030-14524-8_11

24. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, arXiv:1412.6980.

25. M. Kachuee, S. Fazeli, M. Sarrafzadeh, ECG heartbeat classification: A deep transferable representation, in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, IEEE, (2018), 443–444. https://doi.org/10.1109/ICHI.2018.00092

26. T. F. Romdhane, M. A. Pr, Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss, *Comput. Biol. Med.*, **123** (2020), 103866. https://doi.org/10.1016/j.compbiomed.2020.103866

27. S. K. Pandey, R. R. Janghel, Classification of electrocardiogram signal using an ensemble of deep learning models, *Data Technol. Appl.*, **55** (2021), 446–460. https://doi.org/10.1108/DTA-05-2020-0108

28. A. Shoughi, M. B. Dowlatshahi, A practical system based on CNN-BLSTM network for accurate classification of ECG heartbeats of MIT-BIH imbalanced dataset, in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, IEEE, (2021), 1–6. https://doi.org/10.1109/CSICC52343.2021.9420620

29. G. Garcia, G. Moreira, D. Menotti, E. Luz, Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO, *Sci. Rep.*, **7** (2017), 10543. https://doi.org/10.1038/s41598-017-09837-3

30. J. Takalo-Mattila, J. Kiljander, J. P. Soininen, Inter-patient ECG classification using deep convolutional neural networks in *2018 21st Euromicro Conference on Digital System Design (DSD)*, IEEE, (2018), 421–425. https://doi.org/10.1109/DSD.2018.00077

31. Y. Li, R. Qian, K. Li, Inter-patient arrhythmia classification with improved deep residual convolutional neural network, *Comput. Methods Programs Biomed.*, **214** (2022), 106582. https://doi.org/10.1016/j.cmpb.2021.106582

32. R. Siouda, M. Nemissi, H. Seridi, A random deep neural system for heartbeat classification, *Evol. Syst.*, **14** (2023), 37–48. https://doi.org/10.1007/s12530-022-09429-1