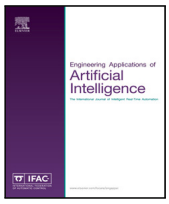
Engineering Applications of Artificial Intelligence 139 (2025) 109480

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

A dual-branch convolutional neural network with domain-informed attention for arrhythmia classification of 12-lead electrocardiograms

Rucheng Jiang [a], Bin Fu [a,*], Renfa Li [a], Rui Li [a], Danny Z. Chen [b], Yan Liu [a], Guoqi Xie [a], Keqin Li [c]

[a] Key Laboratory for Embedded and Network Computing of Hunan Province, College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, Hunan, China
[b] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[c] Department of Computer Science, State University of New York, New York, NY 12561, New Paltz, USA

ARTICLE INFO

Dataset link: http://2018.icbeb.org/Challenge.html, https://physionet.org/content/ptb-xl/1.0.3/

Keywords:
12-lead electrocardiograms
Arrhythmia classification
Dual-branch convolutional network
Domain-informed attention
Intelligent auxiliary diagnosis

ABSTRACT

The automatic classification of arrhythmia is an important task in the intelligent auxiliary diagnosis of an electrocardiogram. Its efficiency and accuracy are vital for practical deployment and applications in the medical field. For the 12-lead electrocardiogram, we know that the comprehensive utilization of lead characteristics is key to enhancing diagnostic accuracy. However, existing classification methods (1) neglect the similarities and differences between the limb lead group and the precordial lead group; (2) the commonly adopted attention mechanisms struggle to capture the domain characteristics in an electrocardiogram. To address these issues, we propose a new dual-branch convolutional neural network with domain-informed attention, which is novel in two ways. First, it adopts a dual-branch network to extract intra-group similarities and inter-group differences of limb and precordial leads. Second, it proposes a domain-informed attention mechanism to embed the critical domain knowledge of electrocardiogram, multiple RR (R wave to R wave) intervals, into coordinated attention to adaptively assign attention weights to key segments, thereby effectively capturing the characteristics of the electrocardiogram domain. Experimental results show that our method achieves an F1-score of 0.905 and a macro area under the curve of 0.936 on two widely used large-scale datasets, respectively. Compared to state-of-the-art methods, our method shows significant performance improvements with a drastic reduction in model parameters.

1. Introduction

Arrhythmia is one of the most common types of cardiovascular disease, which is the leading cause of death worldwide (Amini et al., 2021). The 12-lead electrocardiogram (ECG) is an important tool for assisting the diagnosis of Arrhythmia. Interpreting 12-lead ECG is a labor-intensive task that requires significant time and effort from medical experts with specialized knowledge to ensure accurate interpretation. With the introduction of AI, automatic ECG classification technology can help to improve the accuracy and efficiency of medical expertise in ECG analysis, arrhythmia diagnosis, and treatment planning.

However, automatic ECG classification methods are still limited by the accuracy of classification models in clinical applications. After our analysis, we found that two key factors limit the accuracy of classification models. One factor is that the similarities and differences in the 12-lead ECG have not been fully explored and jointly exploited. The other factor is the lack of domain knowledge that makes it difficult to accurately capture the dynamic characteristics of ECG.

Firstly, existing research does not effectively utilize both the similarities and differences features of the 12-lead ECG, which significantly impacts the performance of the classification model. Some studies designed a single-branch network for each lead to integrate the spatial information from multiple time points within a single lead, aiming to capture the internal characteristics of each lead (Yang et al., 2021; Zhang et al., 2021). While this approach extracts difference features between leads, it overlooks the similarities features among leads. Other studies integrate information from all 12-lead ECG data in spatial and temporal (Wang et al., 2020; Zhang et al., 2020; Yao et al., 2020), extracting global features of 12-lead ECG, which ignores the differences between leads. Secondly, existing methods treat ECG classification as a general problem similar to computer vision and natural language processing, and therefore introduce generic attention mechanisms to

* Corresponding author.
E-mail addresses: Jiangrc@hnu.edu.cn (R. Jiang), fubin@hnu.edu.cn (B. Fu), lirenfa@hnu.edu.cn (R. Li), rui@hnu.edu.cn (R. Li), dchen@nd.edu (D.Z. Chen), liuyan@hnu.edu.cn (Y. Liu), xgqman@hnu.edu.cn (G. Xie), lik@newpaltz.edu (K. Li).

https://doi.org/10.1016/j.engappai.2024.109480
Received 17 May 2024; Received in revised form 9 October 2024; Accepted 12 October 2024
Available online 25 October 2024

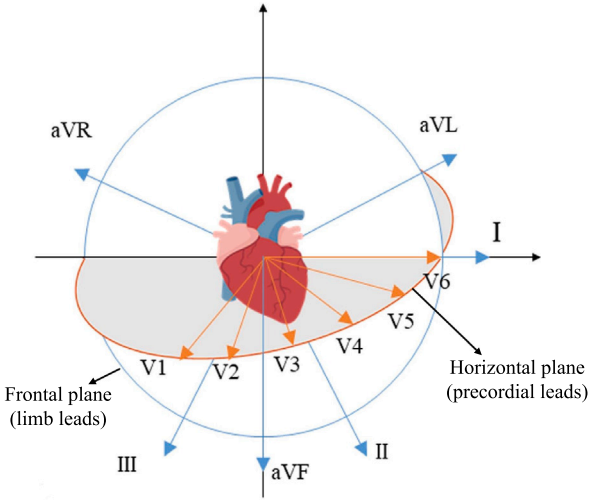


Fig. 1. Two planes of the heart. A 12-lead ECG reflects the electrical activities of the heart in the frontal plane and horizontal plane.

focus on the spatial, temporal, and channel features of ECG signals. However, domain knowledge specific to ECG, such as the RR interval, has not been fully utilized. The RR interval, which represents the time between two consecutive R-peaks in an ECG, is a key indicator for ECG diagnosis (Isler et al., 2019), prediction (Narin et al., 2018; Surucu et al., 2021), and signal segmentation (Shet et al., 2012; Peng et al., 2023; Niroshana et al., 2023). Since the current attention mechanisms do not incorporate this medical domain knowledge, the model's classification accuracy is limited.

This paper aims to consider both the similarities and differences characteristics of 12-lead ECG and introduce professional medical domain knowledge to design a high-precision ECG classification model. First, we believe that the design mechanism of the 12-lead leads to similarities and differences characteristics. In order to reflect the electrical activity of the heart in the frontal and horizontal plane (see Fig. 1), the 12-lead are divided into limb lead groups and precordial lead groups. As shown in Fig. 2, there are similarities in waveforms and signal amplitude values of ECG signals between leads in the same group (Goldberger et al., 2018). However, because the two lead groups represent different cardiac activities, there are significant differences in waveform and signal amplitude between the two groups. In order to extract the similarities and differences features of these ECG signals, we designed a dual-branch network consisting of two single-branch networks. Among them, each single-branch network takes the ECG signal of a lead group as input to extract similarities features within this lead group. There are differences between these two sets of characteristics. Second, we incorporated an attention mechanism to each single network and introduced RR interval as domain knowledge to the attention mechanism. Specifically, RR interval refers to the duration between two consecutive R waves in the ECG. If it is too long or too short, it indicates arrhythmia. Therefore, the attention mechanism can more accurately capture ECG features through domain knowledge.

In addition, we designed a group leads fusion module after the dual-branch network. This module processes the two lead group features extracted by the dual-branch network, further improving the accuracy of the model. First, the module calculates the information entropy of the lead group features to enhance important features and weaken secondary features. Second, the module fuses the feature values of the two lead groups to extract all 12-lead features. This way, the fusion module is more sensitive to important features while still paying attention to global features, thereby improving the classification accuracy of the entire model. Based on the above, we propose a new Dual-branch Convolutional Neural Network with domain-informed attention, which is named DCRR-Net, for arrhythmia classification.

Our main contributions are summarized as follows.

- We propose a dual-branch convolutional neural network to capture the intra-group similarities and inter-group differences of limb and precordial leads, hence significantly enhancing the classification performance.
- We present a domain-informed attention mechanism for effectively extract dynamic features of ECG, which employs coordinate attention with domain knowledge-RR intervals. This mechanism extensively explores the individualized characteristics of RR intervals across different individual patients.
- We conduct extensive experiments on two public datasets to verify the effectiveness of the module by controlling a single module in the network. The experimental results demonstrate the effectiveness of each module. DCRR-Net outperforms state-of-the-art methods and greatly reduces the number of model parameters.

2. Related work

Convolutional neural networks (CNN), deep belief network, recurrent neural networks (RNN), and long short-term memory network (LSTM) have been widely employed for automatic arrhythmia classification (Hong et al., 2020). Early research focused on the classification of single-lead ECG (Yildirim, 2018; Sayantan et al., 2018; Li et al., 2020; Raza et al., 2022; Hannun et al., 2019; Yıldırım et al., 2018; Zhang et al., 2017; Natesan et al., 2020; Petmezas et al., 2021; Hu et al., 2022; Park et al., 2022). Sayantan et al. proposed using a gaussian–bernoulli deep belief network combined with active learning to identify arrhythmias (Sayantan et al., 2018). Hannun et al. proposed a deep neural network to attain 12 arrhythmia classifications for 91,232 single-lead ECG, which achieved cardiologist-level arrhythmia detection performance (Hannun et al., 2019). Petmezas et al. proposed a hybrid CNN-LSTM network structure to improve the accuracy of arrhythmia detection by handling unbalanced data (Petmezas et al., 2021). The 12-lead ECG better meets the requirements of clinical applications, thus prompting more research focus on its automatic classification (Xie et al., 2021; Wang et al., 2020; Zhang et al., 2020; Yao et al., 2020; Yang et al., 2021; Zhang et al., 2021; Ribeiro et al., 2020; Yao et al., 2018; Chen et al., 2019; Ismail Fawaz et al., 2020; He et al., 2019; Wang et al., 2017; Hochreiter and Schmidhuber, 1997). Park et al. employed a 152-layer squeeze-and-excitation residual network to classify five types of ECG, and experiments demonstrated that the classification performance of 12-lead ECG is superior to that of single-lead ECG (Park et al., 2022). Subsequently, several methods have been proposed to improve the accuracy of 12-lead ECG classification. For example, Yang et al. developed a method based on cascaded convolutional neural networks and expert features(CNN_Expert Feature) for arrhythmia classification considering the spatio-temporal correlation of 12-lead ECG (Yang et al., 2021). Ribeiro et al. proposed a unidimensional residual neural network that produced superior results in recognizing six classes of abnormalities on a large dataset of over 2 million samples, outperforming the performance of cardiology residents (Ribeiro et al., 2020). Yao et al. proposed a time-incremental CNN (TI-CNN) to tackle the problem that CNN can receive only fixed-length input and thus may result in loss of critical information (Yao et al., 2018). Chen et al. proposed a CNN-LSTM network to capture the temporal relationships between latent features, considering the characteristics of multivariate time series data (Chen et al., 2019). In addition, comparative learning (Le et al., 2023) was utilized for arrhythmia classification of 12-lead ECG. However, they did not consider both the similarities and differences in the characteristics of the 12-lead ECG limb lead group and precordial lead group.

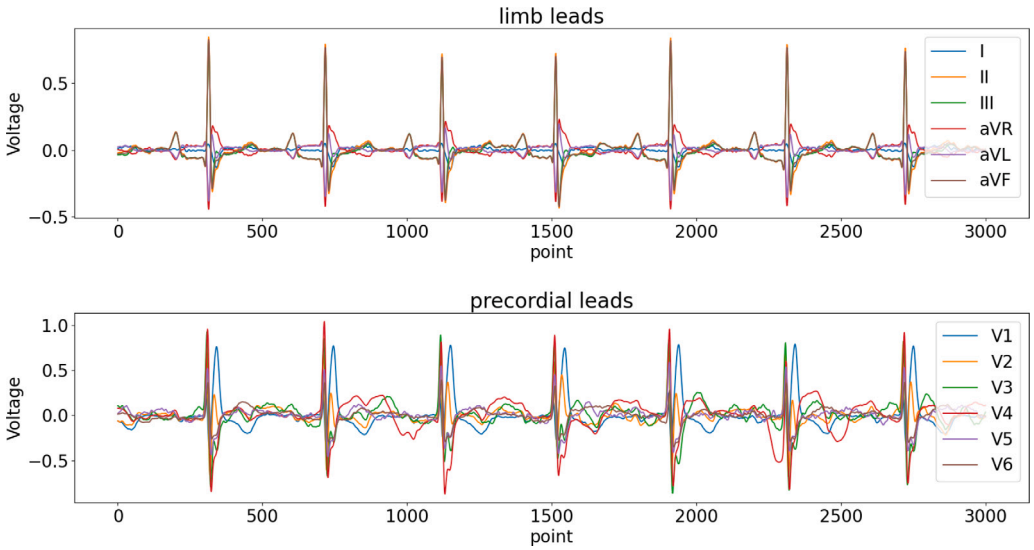


Fig. 2. A 12-lead ECG sample overlaid with group aggregation.

The attention mechanism is an effective method for improving the performance of deep learning models, and it has achieved significant success in various fields (Vaswani et al., 2017; Li et al., 2021; Liu et al., 2022). It also provides a new path for enhancing ECG classification. Some studies have explored the temporal and spatial characteristics of ECG signals (Zhang et al., 2021; Wang et al., 2020; Yao et al., 2020; Zhang et al., 2020; Wang et al., 2023; Han et al., 2023; Song et al., 2024; Li et al., 2024; Srivastava et al., 2021; Fu et al., 2020). For example, Zhang et al. proposed a multi-lead branch fusion network(MLBF-Net), introducing an attention mechanism to integrate the global features of the 12-lead within the network (Zhang et al., 2021). Wang et al. introduced a deep multi-scale fusion convolutional neural network (DMSFNet) that effectively captures disease-related abnormal patterns and performs multi-class arrhythmia detection (Wang et al., 2020). Zhang et al. proposed a spatio-temporal attention-based convolutional RNN (STA-CRNN) which sequentially embeds temporal attention and spatial attention mechanisms, emphasizing locally representative features along the spatial and temporal axes (Zhang et al., 2020). Yao et al. proposed attention-based time-incremental convolutional neural network (ATI-CNN), which uses an attention mechanism to achieve the spatial and temporal fusion of ECG signal information (Yao et al., 2020). Wang et al. proposed a multi-head self-attention mechanism capable of capturing global contextual information to extract relationships and semantic features between ECG segments (Wang et al., 2023). Han et al. used an attention-driven fusion method to fuse temporal instance features with visual instance features (Han et al., 2023). Song et al. introduced a novel attention mechanism based on an improved dynamic time warping method to analyze and control the fusion of temporal and spatial features in ECG signals (Song et al., 2024). Li et al. proposed a beat attention mechanism and a rhythm attention mechanism within the CNN-Transformer framework to enhance the spatio-temporal features of ECG (Li et al., 2024). Other studies focus on channel attention mechanisms. Srivastava et al. proposed a deep neural network based on channel self-attention mechanism to discriminate cardiac abnormalities using combinations of different numbers of leads (Srivastava et al., 2021). Fu et al. introduced a multi-lead attention mechanism in the CNN-LSTM, which automatically measures and assigns weights based on the contributions of different leads (Fu et al., 2020). Existing methods improve the accuracy of deep learning models by introducing universal temporal, spatial, and channel attention mechanisms. However, these studies fail to fully utilize the domain knowledge of ECG to effectively guide the learning and reasoning process of the attention mechanism, which may limit the improvement of model performance.

RR interval is a crucial domain knowledge in ECG (Narin et al., 2018; Isler et al., 2019; Surucu et al., 2021), and several studies utilize it to enhance the performance of classification models. Yang et al. extracted features such as the average, median, and variance of RR intervals as expert features, and combined these features using a random forest to obtain classification probabilities (Yang et al., 2021). Mathews et al. conducted classification using simple features by extracting RR interval and morphological features as input to a deep neural network (Mathews et al., 2018). Rahul et al. proposed an improved RR interval based model by extracting the main feature of RR interval and other statistical features from ECG (Rahul et al., 2021). Eltrass et al. integrated linear and nonlinear features extracted from ECG signals and RR intervals, thereby improving the diagnostic performance of ECG (Eltrass et al., 2022). Udawat et al. proposed extracting time-domain, frequency-domain, and nonlinear features from RR intervals and used statistical analysis and machine learning to detect atrial fibrillation (Udawat and Singh, 2022). Andersen et al. utilized RR intervals in the data pre-processing stage to segment ECG into RR interval segments, from which high-level features were extracted (Andersen et al., 2019). Chen et al. introduced weighted RR intervals to enhance classification performance; they conducted experiments with weight values ranging from 1 to 50 and determined the optimal weights through cluster analysis (Chen et al., 2017). However, existing methods usually use the RR interval in the preprocessing stage of ECG data, for beat segmentation or as a fixed feature value. In other words, they do not explicitly utilize the RR interval to capture the dynamic characteristics of the ECG.

3. Preliminaries

3.1. 12-lead ECG

The heart is a three-dimensional structure with autonomous electrical activity that can be recorded by a 12-lead ECG. A 12-lead ECG can be mapped to two planes (Wan, 2009): the frontal plane (limb leads) and the horizontal plane (precordial leads) (see Fig. 1). The frontal plane includes leads I, II, III, aVR, aVL, and aVF, providing information about the superior, inferior, left, and right positions of the heart. The horizontal plane includes leads V1, V2, V3, V4, V5, and V6, giving information about the left, right, anterior, and posterior positions of the heart. Fig. 2 presents a 12-lead ECG sample overlaid with group aggregation. Within each group, the leads exhibit similarities in waveforms and signal amplitude values, while differences in waveforms and voltage values are evident between the two lead groups.

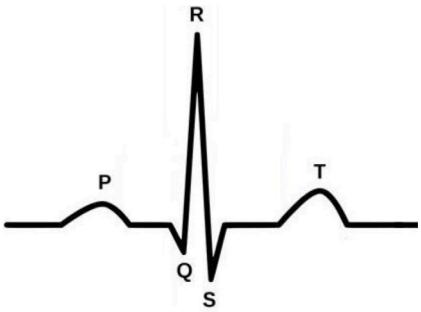


Fig. 3. The morphological characteristics of a heartbeat include the P wave, QRS complex wave, and T wave.

3.2. RR interval

RR interval represents the time elapsed between the R wave peaks of two consecutive heartbeats. A heartbeat includes morphological features such as QRS complex waves, P waves, and T waves (e.g., see Fig. 3). Fluctuations in these waveforms lead to ongoing alterations in the RR interval, which in turn reflect the dynamics of cardiac rhythm. For example, RR interval may become irregular in arrhythmia, excessively prolonged, or shortened. Fig. 4 shows RR interval variations in three aspects: within the same sample, within the same arrhythmia category, and among different arrhythmia categories. RR interval is of significant importance in ECG analysis, as it aids in the identification of heart diseases. It can be defined as follows:

$$RRI = pos(R(n+1)) - pos(R(n)), \tag{1}$$

where $pos(R(i))$ is the position of the ith R wave peak.

3.3. The 12-lead ECG classification problem

We define the 12-lead ECG classification problem as an n-classification problem with a mapping function f: $\mathbf{E} \rightarrow \mathbf{Y}$, where $\mathbf{E} = (e_1, e_2, \ldots, e_i, \ldots, e_{12}) \in \mathbb{R}^{12\times l}$ is an input 12-lead ECG, with each e_i denoting a lead and l indicating the sample length, and $\mathbf{Y} = (y_1, \ldots, y_i, \ldots, y_n)$ with $y_i \in [0, 1]$ representing the probability that $\mathbf{E}$ belongs to the ith category.

4. Method

4.1. Overview

The overall framework of our dual-branch convolutional neural network with a domain-informed attention is shown in Fig. 5. DCRR-Net contains three key components: (1) A dual-branch network is established to extract intra-group similarities and inter-group differences; (2) a domain-informed attention (RR Attention) is proposed to capture dynamic ECG features by embedding multiple RR intervals into a coordinate attention mechanism; (3) a group lead fusion module is presented to fuse global information from the 12-lead ECG.

4.2. Lead grouping and dual-branch network

As discussed in Section 1, there are similarities and differences in the 12-lead ECG. Thus, we divide a 12-lead ECG into limb leads group and precordial leads group and develop a dual-branch network to extract distinct characteristics from the two lead groups. Given the analogous nature of feature extraction from both the limb and precordial leads, we employ the same branch network structure for both groups. The limb/precordial lead branch network includes convolutional layers, depth-wise separable convolution (DSC) (Chollet, 2017), and multiple RR attention modules. It learns features of the limb lead group and precordial lead group, respectively.

A given 12-lead ECG $\mathbf{E}$ is divided into the limb lead group $\mathbf{EL} \in \mathbb{R}^{6\times l}$ and precordial lead group $\mathbf{EP} \in \mathbb{R}^{6\times l}$, as input to the branching network. We design two convolutional layers to extract the low-dimensional features of ECG. In a convolutional layer, a (3, 3) convolution kernel is typically employed. However, to preserve the temporal characteristics of ECG, we use a (1, 3) convolution kernel in each channel. Next, we utilize DSC instead of a traditional convolutional layer. A DSC has two parts: a depth-wise convolutional layer and a point-wise convolutional layer The benefit of DSC is to reduce the number of parameters used for convolutional computation by splitting the correlation between the spatial dimension and channel dimension to achieve the purpose of model lightweight.

Subsequently, the features are fed into the depth-wise convolutional layer (see Fig. 5), which performs spatial convolutions independently on feature maps in each channel. Following this, the feature maps are passed to the RR attention. We propose the RR attention to capture dynamic features between heartbeats by embedding information of multiple RR intervals into the coordinate attention, which will be described in Section 4.3. We posit that segregating spatial convolutions from channel convolutions enables the RR attention to effectively concentrate on the spatial attributes of the ECG. Finally, point-wise convolutional layers execute channel-wise convolutions to merge spatial features. In addition, we enhance the model's expressive capacity by utilizing DSC with convolutional kernels of sizes (1, 1), (1, 3), and (1, 5) at the same time.

4.3. RR attention

RR interval is a highly valuable indicator for assessing arrhythmia. Due to fluctuations in the ECG waveforms, alterations occur in RR intervals, as described in Section 3. To capture the dynamic variations in these waveforms, we design the RR attention (see Fig. 6(b)) that integrates multiple RR intervals with the coordinate attention. It includes two stages: embedding and weighting.

4.3.1. Multiple RR intervals embedding

Inspired by coordinate attention (Hou et al., 2021), we capture the dynamic features of ECG by embedding information of multiple RR intervals into coordinate attention. Coordinate attention encodes the coordinate positions within the feature map, enabling the model to comprehend the significance and relationships among different positions. As discussed in Section 1, ECG waveforms exhibit characteristics of location similarity, which align with the design intent of coordinate attention. In traditional coordinate attention, a feature map can be encoded in the horizontal and vertical directions on each channel (see Fig. 6(a)). In RR attention, we aggregate features separately with different RR intervals in the horizontal direction (Fig. 6(b)). Subsequently, we merge these features to enhance the model's capability to capture dynamic ECG characteristics. At the same time, we preserve precise positional information by vertically aggregating feature maps. The specific steps for embedding RR intervals are as follows.

First, we locate all the R peaks in $\mathbf{E}$ by using the algorithm of Pan and Tompkins (1985). We specifically employ lead II for R peak localization due to its ability to provide clearer and more easily interpretable ECG waveforms. Second, based on the detected R peaks, we calculate all the RR intervals using Eq. (1), and select the maximum RR interval ($r_{\max}$), the average RR interval (r_{avg}), and the minimum RR interval ($r_{\min}$). Due to the similarities in the locations of the 12-lead waveforms, we use RR interval information obtained from lead II for the other leads. Third, we employ multiple RR intervals to horizontally group ECG signals in each lead, facilitating a comprehensive capture of the dynamic features of the ECG. The purpose of grouping the horizontal ECG data based on RR intervals is to facilitate the aggregation of features within these intervals. The numbers of their groupings are

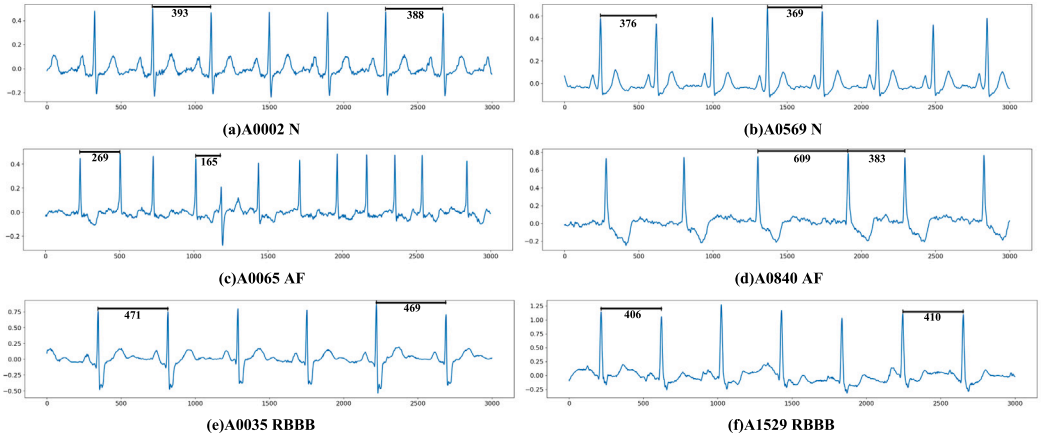


Fig. 4. Variations in the RR interval of samples within the china physiological signaling challenge 2018 (CPSC2018) dataset (Liu et al., 2018). The numbers beneath the black lines represent the lengths of RR intervals.

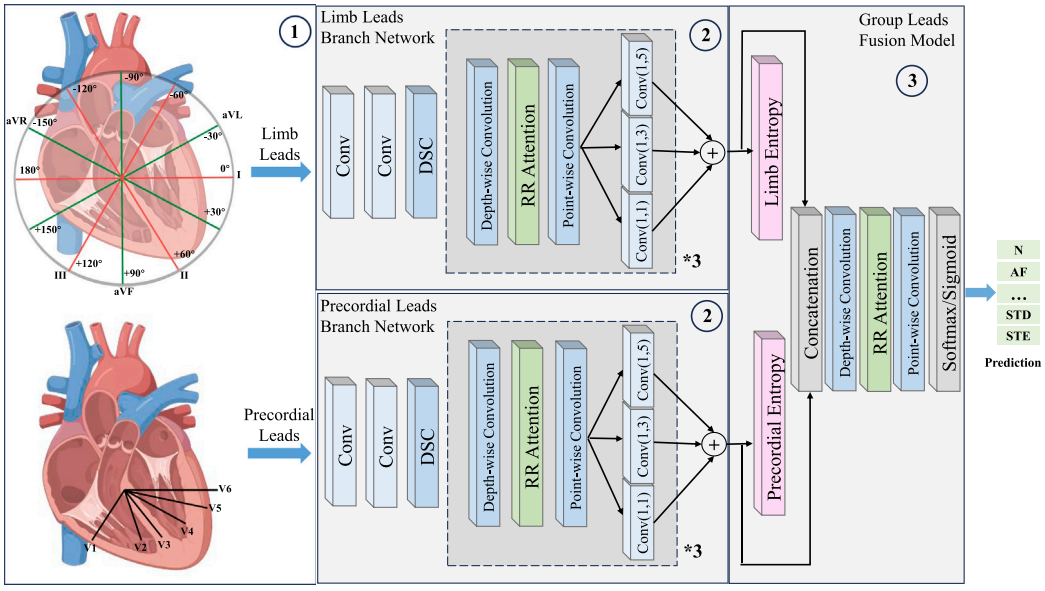


Fig. 5. The overall framework of DCRR-Net. Step 1: Separate the preprocessed ECG into limb leads and precordial leads, which are used as inputs for the two branch networks, respectively. Step 2: Fed two leads into the parallel branch networks, which consist of convolutional layers, depth-wise separable convolutions, and an RR attention module, used to learn the features of the limb leads and precordial leads separately. Step 3: Fuse the features of the limb leads and precordial leads.

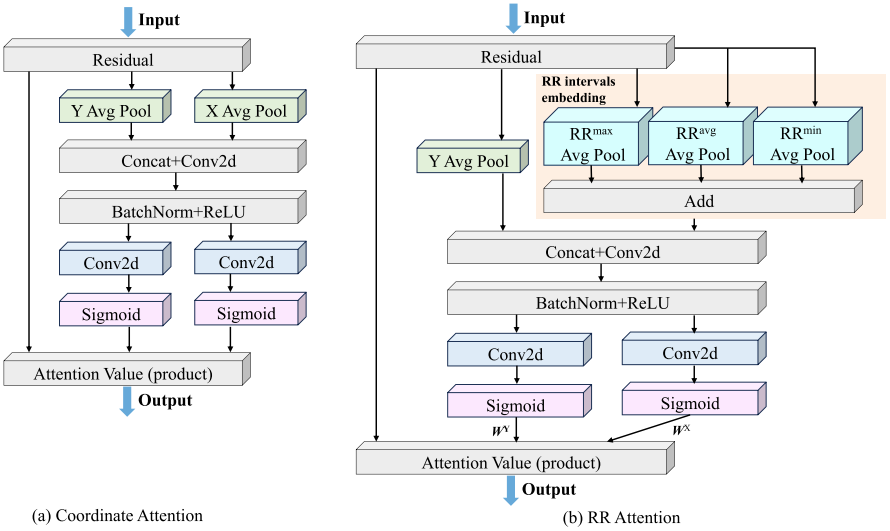


Fig. 6. Coordinate attention versus RR attention. Compared to the original coordinate attention, multiple RR intervals are embedded to capture the dynamic features of the ECG.

determined as $n_{\max} = l/r_{\max}$, $n_{\text{avg}} = l/r_{\text{avg}}$, and $n_{\min} = l/r_{\min}$, where l represents the sample length of a lead in $\mathbf{E}$. We believe that the dynamic characteristics of the ECG can be integrated through multiple RR intervals.

An input feature map $\mathbf{X} \in \mathbb{R}^{c\times h\times w}$ to RR attention undergoes encoding operations in both the horizontal and vertical directions, where c represents the number of channels, h is the height of the feature map, and w is the width. Horizontally, $\mathbf{X}$ is subject to 1D feature encoding using three pooling kernels: $(1, s_{\max})$, $(1, s_{\text{avg}})$, and $(1, s_{\min})$, where $s_{\max} = w/n_{\max}$, $s_{\text{avg}} = w/n_{\text{avg}}$, $s_{\min} = w/n_{\min}$, and w represents the width of the feature map $\mathbf{X}$. Simultaneously, encoding with the three distinct RR intervals is done by Eq. (2), denoted by $\mathbf{RR}^{max}$, $\mathbf{RR}^{avg}$, and $\mathbf{RR}^{min}$, each as:

$$\mathbf{RR}_{i,j,m} = \frac{1}{s}\sum_{k=m\cdot s}^{(m+1)\cdot s-1}\mathbf{X}_{i,j,k}. \tag{2}$$

Specifically, in Eq. (2), in the computation of $\mathbf{RR}^{max}$, the parameters are $s = s_{\max}$ and $m \in [0, n_{\max})$. In the computation of $\mathbf{RR}^{avg}$, $s = s_{\text{avg}}$ and $m \in [0, n_{\text{avg}})$. In the computation of $\mathbf{RR}^{min}$, $s = s_{\min}$ and $m \in [0, n_{\min})$.

Using the three coding schemes in Eq. (2), we embed multiple RR interval information. To more precisely capture the position information of ECG and the relationship between heartbeats, it is essential to fuse the three types of encoding information. For this, we first expand the dimensions of $\mathbf{RR}^{max}$, $\mathbf{RR}^{avg}$, and $\mathbf{RR}^{min}$ to match the dimensions of $\mathbf{X}$ using Eq. (3),

$$\mathbf{RRE}_{i,j,[m\cdot s:(m+1)\cdot s-1]} = \mathbf{RR}_{i,j,m}. \tag{3}$$

After the above expansion, we obtain $\mathbf{RRE}^{max}$, $\mathbf{RRE}^{avg}$, and $\mathbf{RRE}^{min}$, whose dimensions are the same as $\mathbf{X}$'s dimensions. Then an addition operation is performed on them using Eq. (4) to aggregate the dynamic features captured by these three RR intervals,

$$\mathbf{RX}_{i,j,k} = \mathbf{RRE}^{max}_{i,j,k} + \mathbf{RRE}^{avg}_{i,j,k} + \mathbf{RRE}^{min}_{i,j,k}, \tag{4}$$

where $\mathbf{RX}_{i,j,k}$ represents the encoded value after fusion.

Vertically, we encode $\mathbf{X}$ using Eq. (5), where h represents the height of $\mathbf{X}$:

$$\mathbf{RY}_{i,k} = \frac{1}{h}\sum_{j=0}^{h-1}\mathbf{X}_{i,j,k}. \tag{5}$$

4.3.2. RR attention weighting

After embedding through RR intervals, we perform a concatenation operation on $\mathbf{RX}$ and $\mathbf{RY}$, and then the $(1, 1)$ convolution transform function is applied to them, as shown in Eq. (6),

$$\mathbf{F} = \delta(f([\mathbf{RX}, \mathbf{RY}])), \tag{6}$$

where $\mathbf{F} \in \mathbb{R}^{(c/\lambda)\times(h+w)}$, f denotes convolution, λ represents the reduction rate (e.g., $\lambda = 32$), and δ stands for the ReLU function.

Next, we divide $\mathbf{F}$ into $\mathbf{FX} \in \mathbb{R}^{(c/\lambda)\times w}$ and $\mathbf{FY} \in \mathbb{R}^{(c/\lambda)\times h}$ respectively along the horizontal and vertical directions. Two $(1, 1)$ convolution operations fx and fy are performed to transform the $\mathbf{FX}$ and $\mathbf{FY}$ feature maps into the same number of channels as the input, using the sigmoid function σ. The transformation process is expressed as Eqs. (7) and (8):

$$\mathbf{W}^X = \sigma(fx(\mathbf{FX})), \tag{7}$$

and

$$\mathbf{W}^Y = \sigma(fy(\mathbf{FY})), \tag{8}$$

where $\mathbf{W}^X$ and $\mathbf{W}^Y$ are the attention weights in the horizontal and vertical directions, respectively. The final output of the RR attention module is expressed as Eq. (9):

$$\mathbf{A} = \mathbf{X}\cdot\mathbf{W}^Y\cdot\mathbf{W}^X. \tag{9}$$

Incorporating the RR attention into arrhythmia classification offers two significant advantages. First, coordinate attention ensures an efficient transition during the computational process, as the introduction of RR intervals incurs only a small amount of additional computation costs. Second, the combination of multiple RR intervals enables the model to effectively highlight the dynamic characteristics of ECG, thereby enhancing classification performance.

4.4. Group leads fusion module

The 12-lead contains the full range of cardiac activity, and hence the global characteristics of 12-lead should not be ignored. However, the above branch networks consider only features within either the limb or precordial lead groups. Therefore, we develop a group leads fusion module to integrate global features from the 12-lead ECG, mainly consists of information entropy and RR attention.

Information entropy can be used as a selection criterion of parameters and a quantitative indicator of information content (Shang et al., 2021). We propose to assign weights of the lead groups based on information entropy for fusing high-dimensional features of the limb lead group and precordial lead group. Information entropy is used to enhance important features and weaken secondary features. Then, RR attention not only considers all features but also focuses more on important features, thereby enhancing the model's performance. The specific steps of the group leads fusion module are as follows.

First, the output feature maps of the limb leads branch network and the precordial leads branch network are flattened for each channel. This process results in a limb leads feature map $\mathbf{X}^L$ and a precordial leads feature map $\mathbf{X}^P$. Then, the information entropy is calculated based on $\mathbf{X}^L$ and $\mathbf{X}^P$ using Eqs. (10) and (11), denoted as ie_l and ie_p, as shown below:

$$ie_l = \sum_i(-\sum_j p(\mathbf{X}^L_{i,j})\log p(\mathbf{X}^L_{i,j})), \tag{10}$$

$$ie_p = \sum_i(-\sum_j p(\mathbf{X}^P_{i,j})\log p(\mathbf{X}^P_{i,j})). \tag{11}$$

Second, we employ the softmax function to map ie_l and ie_p into [0, 1] by Eq. (12),

$$[s_l, s_p] = softmax([ie_l, ie_p]). \tag{12}$$

Third, $\mathbf{X}^L$ and $\mathbf{X}^P$ are weighted through multiplication, followed by a concatenation operation to derive the comprehensive information of the 12-lead ECG, as shown Eq. (13):

$$\mathbf{F}_{global} = [(1 + s_l)\cdot\mathbf{X}^L, (1 + s_p)\cdot\mathbf{X}^P]. \tag{13}$$

Subsequently, F_{global} is fed to RR attention, which enhances the performance of the model through capturing global features between the 12-lead ECG.

5. Experiments

5.1. Datasets and experimental settings

5.1.1. Datasets

We use two public datasets in our experiments. The china physiological signaling challenge 2018 (CPSC2018) dataset (Liu et al., 2018) was designed specifically for evaluating ECG rhythm and morphological anomaly detection algorithms. The physikalisch-technische bundesanstalt large dataset (PTB-XL) (Wagner et al., 2020) is the most extensive publicly available clinical ECG dataset to date, boasting rich ECG annotations and offering distinct data subsets for specific tasks. We mainly focus on the heart rhythm subtask.

Table 1
CPSC2018 dataset (Liu et al., 2018).

ECG class	Description	Records	Pr (%)	Time length (s)			
				Min	Max	Mean	Median
N	Normal	918	13	10	60	15.43	13.00
AF	Atrial fibrillation	1098	16	9	60	15.01	11.00
1-AVB	First-degree atrioventricular block	704	10	10	60	14.32	11.27
LBBB	Left bundle branch block	207	3	9	60	14.92	12.00
RBBB	Right bundle branch block	1695	25	10	60	14.42	11.19
PAC	Premature atrial contraction	556	8	9	60	19.46	14.00
PVC	Premature ventricular contraction	672	9	6	60	20.21	15.00
STD	ST-segment depression	825	12	8	60	15.13	12.78
STE	ST-segment elevated	202	3	10	60	17.15	11.89

Table 2
Overview of rhythm statements in PTB-XL dataset. Wagner et al. (2020).

	Records	Description
SR	16782	Sinus rhythm
AFIB	1514	Atrial fibrillation
STACH	826	Sinus tachycardia
SARRH	772	Sinus arrhythmia
SBRAD	637	Sinus bradycardia
PACE	296	Normal functioning artificial pacemaker
SVARR	157	Supraventricular arrhythmia
BIGU	82	Bigeminal pattern (unknown origin, SV or Ventricular)
AFLT	73	Atrial flutter
SVTAC	27	Supraventricular tachycardia
PSVT	24	Paroxysmal supraventricular tachycardia
TRIGU	20	Trigeminal pattern (unknown origin, SV or Ventricular)

- The CPSC2018 dataset (Liu et al., 2018) provides 6877 standard 12-lead ECG recordings. The ECG recordings were sampled from 6 s to 60 s at a sampling rate of 500 Hz. The ECG categories are shown in Table 1, described as N, AF, 1-AVB, LBBB, RBBB, PAC, PVC, STD, and STE. We use stratified sampling to divide the dataset into ten parts, 90% of which is allocated to the training set and 10% to the test set.
- The PTB-XL dataset (Wagner et al., 2020) contains 21 837 clinical 12-lead ECG of length 10 s from 18 885 patients. The dataset has a total of 12 categories of rhythm statements, as shown in Table 2. We train and test using data with at least one statement. We divided the training and test sets according to the method provided by Wagner et al. (2020).

5.1.2. Experimental configuration

Experiments were conducted on a workstation equipped with an Inter Xeon(R) E5-2673 v3 processor, NVIDIA GeForce GTX 1080Ti GPU, and 64G memory. The PyTorch framework was employed for the implementation of DL models.

5.1.3. Pre-processing

The CPSC2018 dataset exhibits sample imbalance, as shown in Table 1. Therefore, we applied oversampling to augment the data for smaller classes before training. Additionally, the CPSC2018 dataset contains ECG recordings of varying lengths. Considering that longer segments can increase the computational load and that only parts of these longer segments may contain abnormalities, we segmented the data. We divide all data into 6 s segments. For recordings longer than 6 s, we use a nested segmentation approach, where each subsequent segment overlaps with the previous one by 3 s. This is done to prevent the loss of key features due to segmentation. The records in the PTB-XL dataset are of fixed length. Hence, there was no requirement for segmentation. Furthermore, we removed baseline drift and applied normalization to each sample, which helps reduce the impact of noise and outliers on the overall sample.

5.1.4. Evaluation metrics

We employ standard classification metrics, including accuracy, precision, recall, and F1-score, which are defined as Eqs. (14)–(17):

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

$$precision = \frac{TP}{TP + FP}, \tag{15}$$

$$recall = \frac{TP}{TP + FN}, \tag{16}$$

$$F1 - score = \frac{2 \times (precision \times recall)}{precision + recall}, \tag{17}$$

where *TP*, *FP*, *FN*, and *TN* denote the numbers of true positives, false positives, false negatives, and true negatives, respectively.

To conform to the evaluation criteria established in previous studies, we employ single-labeling for the CPSC2018 dataset, computing class-specific F1-score and overall F1-score using the precision and recall metrics. On the PTB-XL dataset, in which a majority of samples possess multiple labels, we consider the area under the curve (AUC) in our evaluation. AUC represents the area under the receiver operating characteristic (ROC) curve, with a value closer to 1 indicating better model performance.

5.1.5. Training setting

The batch size is configured to 32. By evaluating the two cases of batch size = 32 and batch size = 64, 32 is more suitable in our method, which can effectively balance the convergence speed and accuracy. We set the initial learning rate to 0.001, use 1-cycle learning rate scheduling (Smith, 2018), and AdamW optimizer (Loshchilov et al., 2017). We employed stratified sampling to partition the training and testing sets of the CPSC2018 dataset. For the PTB-XL dataset, we followed the method described in Strodthoff et al. (2020) to partition the training and testing sets. We set weights for each class in PTB-XL to alleviate the data imbalance problem. SR = 0.55, AFIB = 1.16, STACH = 2.12, SARRH = 4.10, SBRAD = 2.27, PACE = 3.75, SVARR = 7.98, BIGU = 12.3, AFLT = 22.4, SVTAC = 24.0, PSVT = 64.9, and TRIGU = 87.6.

5.1.6. Compared methods

We compare our DCRR-Net with representative existing DL methods for automatic 12-lead ECG classification.

For the CPSC2018 dataset, we compare our method with CNN_Expert Feature (Yang et al., 2021),MLBF-Net (Zhang et al., 2021), DMSFNet (Wang et al., 2020), STA-CRNN (Zhang et al., 2020), and ATI-CNN (Yao et al., 2020) for three main reasons. First, these methods adopt two popular model construction strategies: one focuses on extracting internal features of a specific lead, as seen in CNN_Expert Feature and MLBF-Net, while the other treats the 12-lead ECG as a whole and extracts global features from all leads, as seen in ATI-CNN, DMSFNet, and STA-CRNN. These two strategies are currently the dominant approaches for 12-lead ECG classification. Second, ATI-CNN, DMSFNet, STA-CRNN, and MLBF-Net incorporate spatio-temporal attention mechanisms to aggregate the spatial and temporal features of

Table 3
Comparative results of our DCRR-Net and SOTA methods on the CPSC2018 dataset. Results in **bold** indicate the best results in each column.

Method	F1- score for each class									Overall F1-score
	N	AF	1-AVB	LBBB	RBBB	PAC	PVC	STD	STE	
CNN_Expert Feature (Yang et al., 2021)	0.660	0.910	0.570	**0.920**	**0.960**	0.600	0.820	0.770	0.590	0.760
ATI-CNN (Yao et al., 2020)	0.789	0.920	0.850	0.872	0.933	0.736	0.861	0.789	0.556	0.812
DMSFNet (Wang et al., 2020)	0.820	0.900	0.860	0.870	0.930	0.780	0.880	0.800	0.620	0.828
STA-CRNN (Zhang et al., 2020)	0.819	**0.936**	0.866	0.862	0.926	0.789	0.865	0.812	0.640	0.835
MLBF-Net (Zhang et al., 2021)	0.847	0.934	0.884	0.896	0.939	0.822	0.878	0.818	0.677	0.855
DCRR-Net	**0.918**	0.900	**0.921**	0.910	0.919	**0.877**	**0.887**	**0.907**	**0.904**	**0.905**

Table 4
Comparative results (class-wise AUC and macro AUC) of SOTA and our DCRR-Net on the PTB-XL dataset. Results in **bold** indicate the best results in each column.

Method	SR	AFIB	STACH	SARRH	SBRAD	PACE	SVARR	BIGU	AFLT	SVTAC	PSVT	TRIGU	macro AUC
lstm_bidir (Hochreiter and Schmidhuber, 1997)	0.755	0.839	0.953	0.593	0.954	0.846	0.794	0.776	0.870	0.997	0.997	0.804	0.848
fcn_wang (Wang et al., 2017)	0.831	0.938	0.965	0.663	0.953	0.883	0.888	0.768	0.878	0.991	0.965	0.574	0.858
inception1d (Ismail Fawaz et al., 2020)	0.894	0.989	0.982	0.716	0.966	0.884	**0.948**	0.891	**0.981**	0.986	0.986	0.736	0.913
xresnet1d101 (He et al., 2019)	**0.929**	**0.992**	0.984	0.785	**0.970**	**0.947**	0.917	**0.933**	0.904	**0.997**	0.995	0.763	0.926
DCRR-Net	0.919	0.983	**0.991**	**0.800**	0.961	0.891	0.824	0.914	0.970	0.983	**0.997**	**0.998**	**0.936**

the 12-lead ECG, and the effectiveness of attention mechanisms has been demonstrated. Third, CNN_Expert Feature integrates CNN with expert features, which aligns with our approach of leveraging domain knowledge from ECG to improve the model, making it a valuable reference for comparison.

For the PTB-XL dataset, we compare the performance with representative single models provided in Strodthoff et al. (2020), including inception1d (Ismail Fawaz et al., 2020), xresnet1d101 (He et al., 2019), fcn_wang (Wang et al., 2017) and lstm_bidir (Hochreiter and Schmidhuber, 1997). In Strodthoff et al. (2020), they used several representative existing network architectures, such as ResNet, Inception, and LSTM, to perform prediction and classification tasks on the PTB-XL dataset. They concluded that architectures based on ResNet and Inception perform better on time-series classification tasks, with inception1d and xresnet1d101 achieving the best classification performance. While fcn_wang and lstm_bidir represent standard fully convolutional networks and recurrent neural networks respectively. These methods cover the major works of CNNs and RNNs and have been validated on the PTB-XL dataset, making them valuable references for comparison. Implementations of these methods follow as closely as possible the architecture described in the original publications and references.

5.2. Experimental results

5.2.1. Comparison with state-of-the-art methods

The experiments compare the classification performance of our DCRR-Net with state-of-the-art (SOTA) methods on the two public datasets.

Table 3 shows the results of our method in comparison with the SOTA methods on the CPSC2018 dataset. One can see that DCRR-Net achieves the highest overall F1-score. Furthermore, it obtains the highest F1-score for N, 1-AVB, PAC, PVC, STD, and STE. Although its F1-score for AF, LBBB, and RBBB are not the highest, they all exceed 0.9. It is worth noting that its F1-score for STE increases by 0.227 compared to the second best result (Zhang et al., 2021). As shown in Table 3, MLBF-Net (Zhang et al., 2021) achieves the best F1-scores for STD (0.818) and STE (0.677) among the SOTA methods. In contrast, DCRR-Net outperforms (Zhang et al., 2021) with F1-scores of 0.907 for STD and 0.904 for STE, which are substantial improvements. These lower classification accuracies of the SOTA methods can be attributed to the relatively small variations in ST segments, which pose challenges to the known models in capturing such subtle changes. Our model excels by effectively capturing the dynamic features of ECG through multiple RR intervals, resulting in enhanced classification performances in these two categories.

Table 4 illustrates the AUC values for rhythm classification within the PTB-XL dataset. From Table 4, it can be observed that DCRR-Net achieved the highest macro AUC. Additionally, it also obtained the highest AUC for STACH, SARRH, PSVT, and TRIGU. Except for SVARR, the AUC for other categories also exceeded 0.9. Inception1d achieved the highest AUC in SVARR and AFLT. Xresnet1d101 obtains the highest AUC for SR, AFIB, SBRAD, PACE, BIGU, and SVTAC. However, DCRR-Net achieved comparable AUC with xresnet1d101 in AFIB, SBRAD, and SVTAC. It is worth noting that on TRIGU, DCRR-Net (0.998) is 0.233 higher than xresnet1d101 (0.765), which is a considerable improvement. The difference between results in Table 4 and those shown in Strodthoff et al. (2020) stems from different data preprocessing methods used in our implementation compared to Strodthoff et al. (2020). In Strodthoff et al. (2020), these methods are trained on fixed-length (2.5 s) random segments extracted from the complete recordings, which serves as an effective data augmentation. To maintain the same training conditions as DCRR-Net, we train these reference methods on the complete recordings without utilizing data augmentation. These results confirm the effective classification performances of DCRR-Net.

The 95% confidence interval of our model's performance is [0.894, 0.917]. First, the lower bound of the confidence interval is greater than 0.5, indicating that the model's classification performance is significantly better than random guessing. Second, the lower bound of the confidence interval is higher than those of other referenced models, suggesting that our model outperforms the comparative baselines. Therefore, we can be 95% confident that the model's actual performance falls between 0.894 and 0.917.

The above experimental results show that the performance of DCRR-Net outperforms state-of-the-art methods. We believe there are two reasons for this. First, DCRR-Net's dual-branch architecture learns the characteristics of limb leads and precordial leads respectively, thus augmenting feature discrimination when compared with single-branch network approaches. Second, compared to existing methods, the RR attention in DCRR-Net prioritizes dynamic ECG variations, consequently bolstering the model's classification proficiency. We will demonstrate those through experiments and analysis in Sections 5.2.2, 5.2.3, and 5.2.4.

5.2.2. Effect of the dual-branch network

To validate the effectiveness of our dual-branch network structure, we create a control model featuring only a single-branch network. The architecture of the single-branch network in the control model closely resembles that of DCRR-Net, with no modifications to the group leads fusion module. The key distinction lies in that 12-lead ECG is used as input to the control model as a whole.

Table 5 shows the performances of these two models on the CPSC2018 dataset. One can see that the overall F1-score of the nine categories and the F1-score of each class are better when using the

Table 5
Comparison of classification performances between single-branch and DCRR-Net(dual-branch) on the CPSC2018 dataset. Results in **bold** indicate the best results.

Method	F1-score for each class									Overall F1-score
	N	AF	1-AVB	LBBB	RBBB	PAC	PVC	STD	STE	
Single-branch	0.907	0.876	0.900	0.908	0.912	0.840	0.855	0.882	0.865	0.883
DCRR-Net(dual-branch)	**0.918**	**0.900**	**0.921**	**0.910**	**0.919**	**0.877**	**0.887**	**0.907**	**0.904**	**0.905**

Table 6
Performances of ablation experiments on DCRR-Net with the CPSC2018 dataset. Results in **bold** indicate the best results.

Model	F1-score for each class									Overall results			
	N	AF	1-AVB	LBBB	RBBB	PAC	PVC	STD	STE	Overall F1-score	Precision	Recall	Accuracy
BM	0.910	0.895	0.883	0.878	**0.921**	0.834	0.827	0.801	0.846	0.866	0.885	0.855	0.874
BM + RR Attention	**0.931**	0.881	**0.952**	0.843	0.912	0.864	**0.891**	**0.915**	0.837	0.892	0.894	0.890	0.901
DCRR-Net	0.918	**0.900**	0.921	**0.91**	0.919	**0.877**	0.887	0.907	**0.904**	**0.905**	**0.913**	**0.899**	**0.906**

group leads approach than the approach of using 12-lead as a whole. Using the group leads approach, the overall F1-score increases by 2.2%, the F1-scores for STE, RBBB, and PVC increase by over three percent, at 3.9%, 3.7%, and 3.2%, respectively, and for STD, AF, and 1-AVB increase by 2.5%, 2.4%, and 2.1%, respectively. The similarities between limb leads and precordial leads are reflected in the basic waveform structure and time step, and their differences are reflected in the amplitude and shape of the waveform. The single-branch network takes the 12-lead as a whole and does not explicitly distinguish the lead groups. Therefore, the similarity and difference characteristics may be confused in the feature extraction process, and the dual-branch network structure can make up for this shortcoming. The experiments show that by modeling from the perspective of lead groups, the performance of the classification model can be effectively improved by taking into account the similar and different characteristics of 12-lead ECG.

5.2.3. Effect of RR attention and information entropy

To validate the effectiveness of RR attention and information entropy in DCRR-Net, we applied the control variable method and incrementally added RR attention and information entropy.

- Base model(BM): The base model contains branch networks and a group leads fusion module, but excludes RR attention and information entropy. The branch networks are employed to capture similarities features within each lead group and differences features between lead groups. Moreover, we use a group leads fusion module to combine the local and global characteristics of 12-lead ECG.
- BM + RR Attention: On BM, we add RR attention within each branch network to obtain an intra-group level attention mechanism. By embedding multiple RR intervals into coordinate attention, RR attention is able to focus on the dynamic features of ECG.
- BM + RR Attention + Information entropy (DCRR-Net): Building upon the BM + RR Attention, we incorporate RR attention and information entropy into the group leads fusion module. RR attention and information entropy integrate both the local and global characteristics of 12-lead ECG, enhancing the relationships between the leads.

Table 6 reports the results of the ablation experiments on the CPSC2018 dataset. The overall results indicate that our DCRR network's RR Attention and Information entropy are effective for automatic classification of 12-lead ECG. Specifically, on the overall F1-score, BM achieves an F1-score of 0.866. Upon incorporating RR attention into BM, the F1-score increases to 0.894. Furthermore, with the application of information entropy, DCRR-Net achieves an F1-score of 0.905. These results underscore the effectiveness of RR attention and information entropy, both of which exert a favorable influence on the model's performance. The overall results all show a rising trend, underscoring the beneficial influence of DCRR-Net's key modules on the model's performance.

Fig. 7 presents the confusion matrix from the ablation study conducted on the DCRR-Net. By analyzing the results from the confusion matrix, we can observe the correspondence between the predicted and actual categories, further discussing the model's performance. The horizontal axis represents the predicted outcomes, while the vertical axis represents the true labels. Diagonal elements indicate the classification accuracy, with darker colors representing higher classification accuracy, and the off-diagonal elements indicate the probability of misclassification. Overall, from Fig. 7(a) to (c), classification performance progressively improves, indicating that the incremental addition of key modules effectively enhances the model's performance.

Comparing Fig. 7(b) to (a), after adding RR attention on top of the BM, the accuracy of STD increased from 0.72 to 0.93. In Fig. 7(a), the model misclassifies PAC as STD with a probability of 0.13, and PVC as STD with a probability of 0.1, highlighting the model's difficulty in distinguishing between PAC, PVC, and STD. In ECG, both PAC and PVC can exhibit prolonged RR intervals, and PVC is often accompanied by ST-segment changes, which might lead to the model confusing these three conditions. After incorporating RR attention, the model can more accurately capture the dynamic features of the ECG through multiple RR intervals, demonstrating the effectiveness of the RR attention mechanism. Comparing Fig. 7(c) to (b), after adding information entropy based on Fig. 7(b), the overall performance improves further, with the misclassification rate falling below 0.1. The 12-lead ECG provides comprehensive information for diagnosis, but certain arrhythmias are more prominently reflected in specific leads. For example, RBBB can observe abnormal QRS waves in leads V1, V5 or V6, and STE can determine the damaged area of the heart based on the abnormal leads. We use information entropy to enhance the importance of leads, thereby improving the performance of the model.

5.2.4. Visualization of learned features

In this section, we visually evaluate our proposed DCRR-Net model through two visualization methods, focusing on the feature distribution of the data and the areas emphasized by the model.

First, we apply the t-SNE algorithm (van der Maaten and Hinton, 2008) to evaluate the model comprehensively. t-SNE is a nonlinear dimensionality reduction algorithm suitable for reducing high-dimensional data to 2 or 3 dimensions for visualization. In Fig. 8, the horizontal and vertical axes represent the two dimensions of the data mapping. Fig. 8(a) illustrates the data distribution within the CPSC2018 dataset, displaying a substantial overlap among the nine categories and a conspicuous absence of noticeable clustering. This suggests a limited inherent correlation within the original dataset. Fig. 8(b) shows the classification results of the BM. First, the t-SNE results reveal that

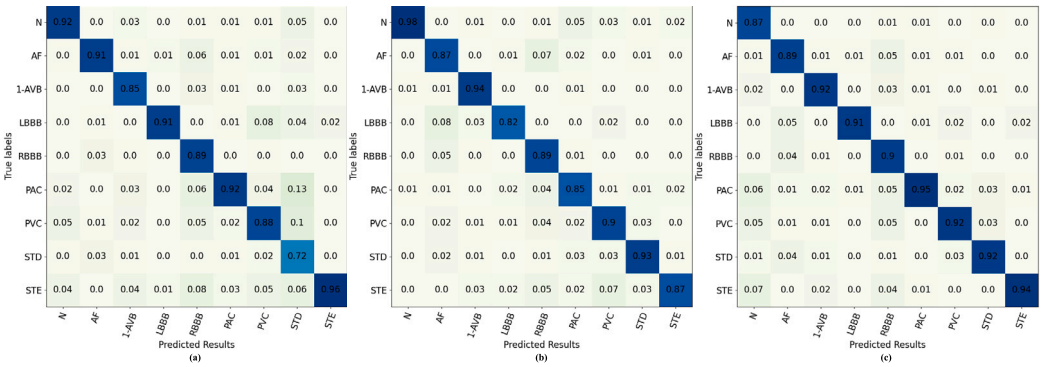


Fig. 7. Confusion matrices for ablation experiments on DCRR-Net with the CPSC2018 dataset. The horizontal axis represents predicted results and the vertical axis represents true labels. Diagonal elements indicate the classification accuracy and off-diagonal elements indicate the probability of misclassification.(a) Base model; (b) Base model and RR attention:; (c) adding information entropy to (b).

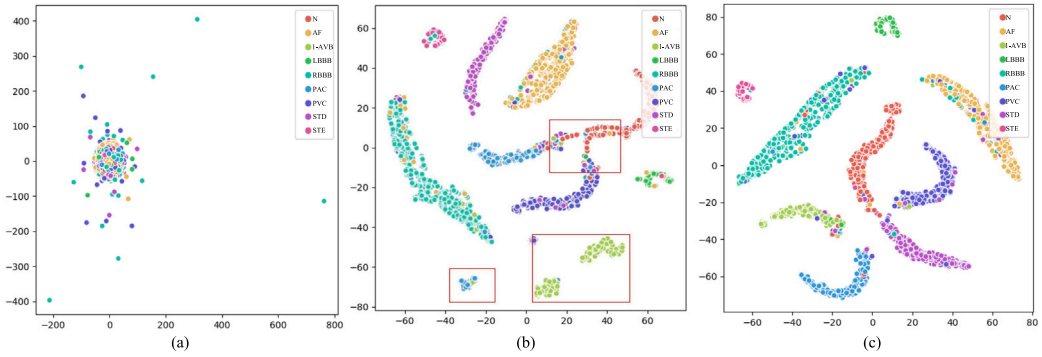


Fig. 8. Visualization results using sampled t-SNE. The horizontal and vertical axes of the graphs represent the two dimensions of the data distributions. (a) The original distribution of the CPSC2018 dataset. (b) Distribution of CPSC2018 after base model test. (c) Distribution of CPSC2018 after DCRR-Net test.

the data forms several distinct clusters, with AF, 1-AVB, LBBB, RBBB, STD, and STE each forming clear clusters, indicating a high degree of separation between different classes. However, the boundaries between N, PAC, and PVC are blurred, suggesting that these three may share similar features in high-dimensional space. Additionally, both PAC and 1-AVB exhibited two clusters, indicating that the aggregation ability of features within these categories is limited. Finally, overlapping data points across different classes suggest that the model is not entirely accurate in distinguishing between the features. Fig. 8(c) illustrates the visualization results of DCRR-Net, which incorporates RR attention and information entropy on the basis of Fig. 8(b). Compared to Fig. 8(b), Fig. 8(c) exhibits two notable changes: First, the boundaries between the nine categories are more distinct, indicating greater differentiation in the features of different classes. Second, the feature aggregation for the 1-AVB and PAC categories is tighter, suggesting a more compact internal structure within these classes. From the perspective of data distribution, this implies that the introduction of RR attention and entropy has enhanced feature aggregation, thereby improving the performance of DCRR-Net.

Second, we use gradient-weighted class activation mapping (Selvaraju et al., 2017) for feature interpretation. Fig. 9 shows the visualization results of the left bundle branch block (LBBB) and the right bundle branch block (RBBB) on the CPSC2018 dataset. The darker blue part in gradient-weighted class activation mapping means that this part of the features is more important, and a red box shows a specific feature. First, the LBBB instances and RBBB instances are from different samples, and thus their RR intervals are different. By visualizing the results, it becomes evident that the model has acquired an understanding of the RR interval features, as the blue region consistently aligns with each heartbeat. Second, Fig. 9(a) shows the LBBB features learned by the model. Diagnostic criteria for LBBB encompass the disappearance of the Q wave, broadening and splitting of the R wave, and T wave inversion, as prominently demonstrated in lead III. Furthermore, a small R wave emerges alongside a widened and deepened S wave, accompanied by an upright T wave, as depicted in lead V3. Fig. 9(b) shows that the model has learned RBBB features. The diagnostic criteria for RBBB are that the QRS waveform in lead V1 is split, r < R', and the T wave is inverted; the S wave in lead V6 is widened and the T wave is upright. Overall, the gradient-weighted class activation mapping results show that the model effectively learns the characteristics of different disease patterns by embedding the RR interval information, thus improving the model's discriminating ability.

5.2.5. Computation cost comparison

In this section, we compare the number of parameters of several models. The number of model parameters represents a crucial metric for assessing models, especially in practical application scenarios. It is imperative to evaluate a model not solely based on its accuracy but also with regard to its resource requirements, thus enhancing its practical utility. Table 7 shows the comparison of the parameter amount and overall performance of DCRR-Net and reference models. From Table 7, it is evident that DCRR-Net achieved the highest F1-score and the lowest number of parameters (0.17 million) on the CPSC2018 dataset. For PTB-XL, xresnet1d101 achieved the second highest AUC, while

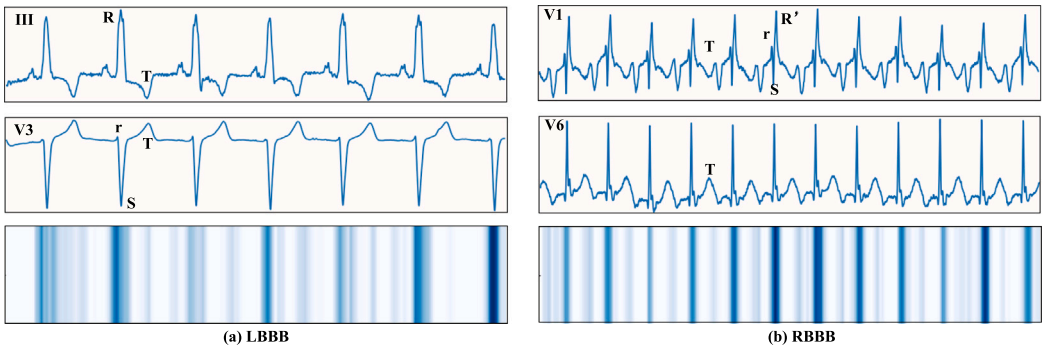


Fig. 9. Gradient-weighted class activation mapping visualized features(Third row): LBBB and RBBB on the CPSC2018 dataset. The deeper the color, the higher the attention received in that area. (a) The diagram shows the waveform characteristics of leads III and V3 of LBBB. (b) The diagram shows the waveform characteristics of leads V1 and V6 of RBBB.

Table 7
Parameter comparison between DCRR-Net and reference models. The unit of parameter count is in millions.

CPSC2018			PTB-XL		
Method	F1-score	Parameters	Method	macro AUC	Parameters
TI-CNN	77.3	5.26	inception1d	0.913	0.04
ATI-CNN	81.8	4.98	xresnet1d101	0.926	3.77
MLBF-Net	85.5	0.42	fcn_wang	0.858	0.03
DCRR-Net	90.5	0.17	lstm_bidir	0.848	2.33
			DCRR-Net	0.936	0.17

the number of parameters for xresnet1d101 (3.77 million) is significantly higher than DCRR-Net (0.17 million). Additionally, fcn_wang has the smallest number of parameters (0.03 million), but its AUC is far lower than DCRR-Net. DCRR-Net strikes a balance between model performance and the number of parameters.

5.3. Discussion

Our study demonstrates that DCRR-Net significantly improves the classification accuracy of 12-lead ECG. This improvement is mainly attributed to two factors: First, the dual-branch network effectively captures both the similarity in waveform patterns and timesteps within lead groups and the differences in waveform morphology between lead groups. Second, the introduction of RR Attention allows the model to capture the dynamic characteristics of ECG signals.

Previous studies emphasized either the spatial features across multiple time points within a single lead or the integration of global spatial and temporal features across the 12-lead. However, unlike these methods, we were inspired by the mapping relationship between ECG signals and cardiac activity. By utilizing a dual-branch network, we capture both the similarities and differences in features between the 12-lead ECG, complementing the extraction of global and local features, which enhances model accuracy. Additionally, considering domain knowledge of ECG, where the RR interval is an important feature for distinguishing between multiple categories, we combined this with an attention mechanism to more accurately capture the dynamic characteristics of ECG signals, further improving the model's classification performance.

Although our DCRR-Net has proven effective on two datasets, it still has two limitations. First, we assign different weights based on sample size to alleviate the impact of data imbalance on the model, but this method may not perform equally well on different datasets. Second, while we used visualization techniques to analyze the experimental results, we did not explore the model's interpretability.

6. Conclusion

We proposed a 12-lead ECG classification method, DCRR-Net, which combined convolutional neural networks with a domain-informed attention mechanism. First, we designed a dual-branch structure to capture similarity and difference features. Second, we introduced a domain-informed attention to highlight the dynamic characteristics of the ECG. In addition, we developed a lead fusion module to integrate the global features of the 12-lead. Experimental results showed that our proposed DCRR-Net outperformed existing methods on two public datasets. Our research establishes the foundation for developing more accurate 12-lead ECG classification models, demonstrating the potential of combining domain knowledge with deep learning, and providing new insights for the automatic analysis of 12-lead ECG.

In future work, we will consider exploring the combination of more ECG domain knowledge and deep learning models. In addition, we will further explore the potential relationship between the 12-lead and the ECG categories.

CRediT authorship contribution statement

Rucheng Jiang: Writing – original draft, Software, Methodology, Conceptualization. **Bin Fu:** Writing – original draft, Supervision. **Renfa Li:** Supervision, Funding acquisition. **Rui Li:** Writing – review & editing, Funding acquisition. **Danny Z. Chen:** Writing – review & editing. **Yan Liu:** Supervision, Funding acquisition. **Guoqi Xie:** Writing – review & editing. **Keqin Li:** Writing – review & editing.

Declaration of competing interest

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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant numbers 61932010, 62272155]; the Natural Science Foundation of Hunan Province of China [grant number 2024JJ5097].

Appendix. Notations description

Notation	Description
ECG	Electrocardiograms
DCRR-Net	Dual-branch Convolutional Neural Network with domain-informed attention
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory Network
CNN_Expert Feature	Convolutional Neural Network and Expert Features
MLBF-Net	Multi-Lead Branch Fusion Network
DMSFNet	Deep Multi-Scale Fusion convolutional neural network
STA-CRNN	Spatio-Temporal Attention-based Convolutional RNN
TI-CNN	Time Incremental Convolutional Neural Network
ATI-CNN	Attention-based Time-incremental Convolutional Neural Network
RRI	RR interval
DSC	Depth-wise Separable Convolution
CPSC2018	China Physiological Signaling Challenge 2018
PTB-XL	Physikalisch-Technische Bundesanstalt Large dataset
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
BM	Base model
SOTA	State-of-the-art

Data availability

Two datasets used in this paper are available at the following URL: http://2018.icbeb.org/Challenge.html (CPSC2018) and https://physionet.org/content/ptb-xl/1.0.3/ (PTB-XL).

References

Amini, M., Zayeri, F., Salehi, M., 2021. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. BMC Public Health 21, 1–12.

Andersen, R.S., Peimankar, A., Puthusserypady, S., 2019. A deep learning approach for real-time detection of atrial fibrillation. Expert Syst. Appl. 115, 465–473.

Chen, S., Hua, W., Li, Z., Li, J., Gao, X., 2017. Heartbeat classification using projected and dynamic features of ECG signal. Biomed. Signal Process. Control 31, 165–173.

Chen, Y.-J., Liu, C.-L., Tseng, V.S., Hu, Y.-F., Chen, S.-A., 2019. Large-scale classification of 12-lead ECG with deep learning. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics. BHI, pp. 1–4.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 1251–1258.

Eltrass, A.S., Tayel, M.B., Ammar, A.I., 2022. Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures. Neural Comput. Appl. 34 (11), 8755–8775.

Fu, L., Lu, B., Nie, B., Peng, Z., Liu, H., Pi, X., 2020. Hybrid network with attention mechanism for detection and location of myocardial infarction based on 12-lead electrocardiogram signals. Sensors 20 (4), 1020.

Goldberger, A.L., Goldberger, Z.D., Shvilkin, A., 2018. Chapter 4 - ECG leads. In: Goldberger's Clinical Electrocardiography, ninth ed. Elsevier, pp. 21–31. http://dx.doi.org/10.1016/B978-0-323-40169-2.00004-4.

Han, H., Lian, C., Zeng, Z., Xu, B., Zang, J., Xue, C., 2023. Multimodal multi-instance learning for long-term ECG classification. Knowl.-Based Syst. 270, 110555.

Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y., 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Med. 25 (1), 65–69.

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M., 2019. Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 558–567.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hong, S., Zhou, Y., Shang, J., Xiao, C., Sun, J., 2020. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Comput. Biol. Med. 122, 103801.

Hou, Q., Zhou, D., Feng, J., 2021. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 13713–13722.

Hu, R., Chen, J., Zhou, L., 2022. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. Comput. Biol. Med. 144, 105325.

Isler, Y., Narin, A., Ozer, M., Perc, M., 2019. Multi-stage classification of congestive heart failure based on short-term heart rate variability. Chaos Solitons Fractals 118, 145–151.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.-A., Petitjean, F., 2020. Inceptiontime: Finding alexnet for time series classification. Data Min. Knowl. Discov. 34 (6), 1936–1962.

Le, D., Truong, S., Brijesh, P., Adjeroh, D.A., Le, N., 2023. sCL-ST: Supervised contrastive learning with semantic transformations for multiple lead ECG arrhythmia classification. IEEE J. Biomed. Health Inf. 27 (6), 2818–2828.

Li, H., Han, J., Zhang, H., Zhang, X., Si, Y., Zhang, Y., Liu, Y., Yang, H., 2024. Clinical knowledge-based ECG abnormalities detection using dual-view CNN-transformer and external attention mechanism. Comput. Biol. Med. 178, 108751.

Li, Y., Pang, Y., Wang, K., Li, X., 2020. Toward improving ECG biometric identification using cascaded convolutional neural networks. Neurocomputing 391, 83–95.

Li, J., Wang, X., Tu, Z., Lyu, M.R., 2021. On the diversity of multi-head attention. Neurocomputing 454, 14–24.

Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., et al., 2018. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J. Med. Imag. Health Inform. 8 (7), 1368–1373.

Liu, T., Luo, R., Xu, L., Feng, D., Cao, L., Liu, S., Guo, J., 2022. Spatial channel attention for deep convolutional neural networks. Mathematics 10 (10), 1750.

Loshchilov, I., Hutter, F., et al., 2017. Fixing weight decay regularization in adam. arXiv preprint arXiv:1711.05101, 5.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (86), 2579–2605.

Mathews, S.M., Kambhamettu, C., Barner, K.E., 2018. A novel application of deep learning for single-lead ECG classification. Comput. Biol. Med. 99, 53–62.

Narin, A., Isler, Y., Ozer, M., Perc, M., 2018. Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability. Phys. A 509, 56–65.

Natesan, P., Gothai, E., et al., 2020. Classification of multi-lead ECG signals to predict myocardial infarction using CNN. In: 2020 Fourth International Conference on Computing Methodologies and Communication. ICCMC, pp. 1029–1033.

Niroshana, S.I., Kuroda, S., Tanaka, K., Chen, W., 2023. Beat-wise segmentation of electrocardiogram using adaptive windowing and deep neural network. Sci. Rep. 13 (1), 11039.

Pan, J., Tompkins, W.J., 1985. A real-time QRS detection algorithm. IEEE Trans. Biomed. Eng. (3), 230–236.

Park, J., An, J., Kim, J., Jung, S., Gil, Y., Jang, Y., Lee, K., Oh, I.-y., 2022. Study on the use of standard 12-lead ECG data for rhythm-type ecg classification problems. Comput. Methods Programs Biomed. 214, 106521.

Peng, X., Zhu, H., Zhou, X., Pan, C., Ke, Z., 2023. ECG signals segmentation using deep spatiotemporal feature fusion U-Net for QRS complexes and R-peak detection. IEEE Trans. Instrum. Meas. 72, 1–12.

Petmezas, G., Haris, K., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J.A., Katsaggelos, A.K., Maglaveras, N., 2021. Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets. Biomed. Signal Process. Control 63, 102194.

Rahul, J., Sora, M., Sharma, L.D., Bohat, V.K., 2021. An improved cardiac arrhythmia classification using an RR interval-based approach. Biocybern. Biomed. Eng. 41 (2), 656–666.

Raza, A., Tran, K.P., Koehl, L., Li, S., 2022. Designing ECG monitoring healthcare system with federated transfer learning and explainable AI. Knowl.-Based Syst. 236, 107763.

Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., Ferreira, M.P., Andersson, C.R., Macfarlane, P.W., Meira, Jr., W., et al., 2020. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Commun. 11 (1), 1760.

Sayantan, G., Kien, P., Kadambari, K., 2018. Classification of ECG beats using deep belief network and active learning. Med. Biol. Eng. Comput. 56 (10), 1887–1898.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV, pp. 618–626.

Shang, Z., Li, W., Gao, M., Liu, X., Yu, Y., 2021. An intelligent fault diagnosis method of multi-scale deep feature fusion based on information entropy. Chin. J. Mech. Eng. 34 (1), 1–16.

Shet, M.S., Patel, M., Rao, A., Kantharaj, C., Suma, K., 2012. ECG arrhythmia classification using R-peak based segmentation, binary particle swarm optimization and absolute euclidean classifier. In: Proceedings of International Conference on Advances in Computing. Springer, pp. 303–317.

Smith, L.N., 2018. A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. arXiv preprint arXiv: 1803.09820.

Song, C., Zhou, Z., Yu, Y., Shi, M., Zhang, J., 2024. An improved Bi-LSTM method based on heterogeneous features fusion and attention mechanism for ECG recognition. Comput. Biol. Med. 169, 107903.

Srivastava, A., Hari, A., Pratiher, S., Alam, S., Ghosh, N., Banerjee, N., Patra, A., 2021. Channel self-attention deep learning framework for multi-cardiac abnormality diagnosis from varied-lead ECG signals. In: 2021 Computing in Cardiology (CinC). Vol. 48, pp. 1–4.

Strodthoff, N., Wagner, P., Schaeffter, T., Samek, W., 2020. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. IEEE J. Biomed. Health Inf. 25 (5), 1519–1528.

Surucu, M., Isler, Y., Perc, M., Kara, R., 2021. Convolutional neural networks predict the onset of paroxysmal atrial fibrillation: Theory and applications. Chaos 31 (11), 113119.

Udawat, A.S., Singh, P., 2022. An automated detection of atrial fibrillation from single-lead ECG using HRV features and machine learning. J. Electrocardiol. 75, 70–81.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. Vol. 30, Curran Associates, Inc..

Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T., 2020. PTB-XL, a large publicly available electrocardiography dataset. Sci. Data 7 (1), 1–15.

Wan, H., 2009. Clinical Electrocardiography, sixth ed. Beijing: People's Medical Publishing House, pp. 4–5.

Wang, R., Fan, J., Li, Y., 2020. Deep multi-scale fusion neural network for multi-class arrhythmia detection. IEEE J. Biomed. Health Inf. 24 (9), 2461–2472.

Wang, Z., Yan, W., Oates, T., 2017. Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1578–1585.

Wang, Y., Yang, G., Li, S., Li, Y., He, L., Liu, D., 2023. Arrhythmia classification algorithm based on multi-head self-attention mechanism. Biomed. Signal Process. Control 79, 104206.

Xie, Y., Qin, L., Tan, H., Li, X., Liu, B., Wang, H., 2021. Automatic 12-leading electrocardiogram classification network with deformable convolution. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, pp. 882–885.

Yang, X., Zhang, X., Yang, M., Zhang, L., 2021. 12-Lead ECG arrhythmia classification using cascaded convolutional neural network and expert feature. J. Electrocardiol. 67, 56–62.

Yao, Q., Fan, X., Cai, Y., Wang, R., Yin, L., Li, Y., 2018. Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram. In: 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). pp. 754–761.

Yao, Q., Wang, R., Fan, X., Liu, J., Li, Y., 2020. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. Inf. Fusion 53, 174–182.

Yildirim, Ö., 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. 96, 189–202.

Yıldırım, Ö., Pławiak, P., Tan, R.-S., Acharya, U.R., 2018. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Comput. Biol. Med. 102, 411–420.

Zhang, J., Liang, D., Liu, A., Gao, M., Chen, X., Zhang, X., Chen, X., 2021. MLBF-Net: A multi-lead-branch fusion network for multi-class arrhythmia classification using 12-lead ECG. IEEE J. Transl. Eng. Health Med. 9, 1–11.

Zhang, J., Liu, A., Gao, M., Chen, X., Zhang, X., Chen, X., 2020. ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. Artif. Intell. Med. 106, 101856.

Zhang, C., Wang, G., Zhao, J., Gao, P., Lin, J., Yang, H., 2017. Patient-specific ECG classification based on recurrent neural networks and clustering technique. In: 2017 13th IASTED International Conference on Biomedical Engineering (BioMed). pp. 63–67.