



ECG recognition based on deep convolutional neural network with dual attention mechanism

Xiaoshan Zhang^a, Xiaolong Chen^a, Yinwei Li^{a,c,*}, Yonggang Luo^b, Yiming Zhu^{a,c}

^a School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

^b Neurosurgical Intensive Care Unit, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China

^c Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China

ARTICLE INFO

Keywords:

ECG signal

Convolutional neural network

Attention mechanism

ABSTRACT

Electrocardiogram (ECG) signal classification based on deep learning has made great progress in recent years. However, the heartbeat signal has strong specificity, insufficient feature richness, and weak recognition ability. The conventional deep neural network adopts the structure of convolution layer and lacks the ability to pertinently learn important features. Therefore, a deep convolution neural network based on dual attention mechanism (DA-DCNN) is proposed in this paper to classify the ECG signals of arrhythmia. First, ECG signal is preprocessed by time–frequency analysis to enrich its characteristics. Then, by adding the attention mechanism to enhance the weight and influence of features that contribute more, the DA-DCNN pays more attention to important features and reduces the learning of general features. These two improvements reduce the dependence of the network on the number of convolutional layers, making the ECG signal classification more robust and generalized. Experiments on MIT-BIH data sets are conducted and the classification accuracy of DA-DCNN network in N, S, V, F and Q beats are 99.13%, 99.52%, 99.57%, 99.75% and 99.98% respectively, which proves that the performance of DA-DCNN is superior to the existing methods.

1. Introduction

With the improvement of medical care, the average life expectancy of human beings has increased from 66.8 years in 2000 to 73.3 years in 2019 [1]. Therefore, people pay more and more attention to cardiovascular diseases, especially cardiovascular diseases, such as heart rate imbalance. As an intuitive embodiment of people's physical function, the beating law of the heart plays an important role in physical detection. So how to accurately identify the beating signal of the heart has become an urgent need to monitor the condition of the heart. Electrocardiogram (ECG) is an important method to detect human heart beat. How to accurately distinguish the type of ECG by patients' ECG signals plays a key role in the treatment and prevention of cardiovascular diseases in clinical medicine, which is also a hot topic for scholars.

The traditional ECG signal classification mainly adopts the method that requires signal analysis experts or experienced personnel to manually add features, such as statistical methods based on fuzzy mathematics [2], random forest method [3], morphological features [4,5], time information [6,7], Hermite transform [8], wavelet transform

[9], etc. However, due to the diversity and uniqueness of ECG signals, the characteristics of ECG signals in different diseases of the same person are different, and the ECG signals in different people with the same disease are also different. Therefore, the traditional method is difficult to extract the depth characteristics of ECG signals, and its accuracy is low, which cannot reach the diagnosis level of doctors. There is still a long way to go in the application of traditional ECG automatic diagnosis in people's medical treatment.

Deep convolution neural network is a popular algorithm in recent years. Compared with traditional algorithms, it uses convolution, pooling, feedback learning and other methods for image processing, and can realize accurate recognition and classification of the signal without adding signal features. In [10,11], one-dimensional signals are directly input into the convolutional neural network, resulting in the loss of two-dimensional feature information of ECG signals, such as feature peaks. Since it is difficult for networks to extract useful features from discrete-time signals, classification results are prone to error. Recently, two-dimensional signal input convolution neural network has been used to improve the resolution of the network. In 2019, Hannun et al. [12] used

* Corresponding author at: School of Optoelectronic Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China.

E-mail address: liyw@usst.edu.cn (Y. Li).

<https://doi.org/10.1016/j.bspc.2025.109142>

Received 11 September 2023; Received in revised form 13 October 2025; Accepted 9 November 2025

Available online 18 November 2025

1746-8094/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

91,232 ECG images from 53,877 patients to successfully classify these data into 12 heartbeat types with a complex network of 33 layers. In 2020, Dokur et al. [13] proposed a convolution neural network (CNN) heartbeat classification method based on Walsh function training, which reduces the number of convolutional. Jeong et al. [14] and Xie et al. [15] improved the classification accuracy by using the similar method of feature-enrich CNN. The F1 values of the former reached to 0.89 and 0.52 for the highest classification LBBB beat and the lowest classification STE beat, respectively. And the F1 values of the latter are 0.94 and 0.82 for V beat and S beat respectively, which reaches a higher accuracy. However, although the features of ECG signals are more abundant, it also increases the complexity of convolution, and makes the network pay relatively less attention to important features and be interfered by other general features. In addition, different areas of the image may have different weight. Such as the P-wave of the heartbeat signal: this is the first positive spike caused by a change in the potential generated by the sinoatrial node, which usually indicates atrial contraction. How to make neural network pay more attention to these important features and deepen the learning intensity and depth of these features is a problem that must be solved in ECG signal classification.

Recently, attention mechanism [16] has been widely and successfully applied in artificial intelligence fields such as natural language processing [17], image recognition [18–20], recommendation system [21,22]. The attention mechanism can adaptively weight the features in the input sample, making the network pay more attention to important information, ignore irrelevant or redundant information, and reduce the risk of overfitting. By adding attention mechanism to convolution neural network, the network can learn more comprehensive and in-depth details, so as to complete the classification task better.

Therefore, in this paper, we use time–frequency analysis method to preprocess ECG signals to increase the richness of signal features. And we propose a deep convolutional neural network with dual attention mechanism (DA-DCNN) to solve the problem of effective feature learning for feature-enrich ECG signals and realize the recognition of multi-beat heartbeat signals. The dual attention mechanism uses two different modules to calculate the weight of features and synthetically feedback the results back to the network. Because of the variety and complexity of ECG signals, different modules can pay more attention to different important characteristics of ECG signals. The dual attention mechanism demonstrates superior classification performance compared to its single-module counterpart by selectively focusing on pivotal feature representations through coordinated channel-spatial interactions. Furthermore, the synergistic integration of complementary attention modules not only mitigates model overfitting but also enhances learning efficiency through diversified feature optimization pathways. Using the attention mechanism, the model can adaptively adjust the weights of different regions, focus on the regions containing important information, extract important features, and finally classify the ECG signals. The attention mechanism can also enhance the model's ability to recognize local details and better focus on the most relevant parts of the input data, thereby improving the speed of network convergence and the accuracy of the recognition results. Two problems of ECG signal classification are solved effectively: (1) The increase of network layer makes the system more complex, and the network operation parameters are too large, resulting in resource waste and performance redundancy; (2) With the increase of signal features, the traditional network cannot focus on learning the important features, and the learning efficiency decreases.

2. Dataset

Data sets are the basis for training DA-CNN. Only a sufficient and effective data set can ensure that the prediction model is iterated on the network, and finally an excellent and universally applicable model is obtained. In this paper, MIT-HIB dataset, provided by the Massachusetts Institute of Technology [23], is used. The dataset collects relevant ECG

data from 47 arrhythmia laboratories, providing an open research platform for arrhythmia researchers to promote the development of the field. The “. Atr” annotation file in the MIT-BIH dataset contains 19 data types. According to the guidelines from the Association for the Advancement of Medical Instrumentation (AAMI), the MIT-BIH dataset can be divided into the following five categories: normal heartbeat (N), supraventricular abnormality (S), ventricular extrasystole (V), ventricular fusion heartbeat (F), and unclassified heartbeat (Q), with type N being the most common ECG signal and the other four categories being abnormal or arrhythmia signals. These classification criteria help medical staff to classify and analyze ECG data, and help doctors to make better and faster diagnosis, so as to better understand and diagnose heart disease.

The results of extracting and plotting two-dimensional signals from the MIT-BIH dataset are shown in Fig. 1. The signal sampling number is 360 Hz. Fig. 1-a indicates normal heartbeat, and Fig. 1-b to 1-e indicates abnormal ECG signals.

3. Network

3.1. System framework

In this paper, we built a deep convolution neural network with dual attention mechanism, and use this network to classify and process heartbeat signals of MIT-BIH dataset. The preprocessing of the network uses time–frequency analysis to process the ECG signal to enrich the feature information of the ECG signal and provide more features for the network learning. The DA-CNN structure can perform deep learning on the pre-processed data. Meanwhile, the attention mechanism can make the network pay more attention to the important features among many features, and concentrate the learning, which is conducive to the network to learn the local details of ECG signals. The above improvements improve the classification accuracy of the network enhance the robustness of the network.

The system framework is shown in Fig. 2. First, the ECG signal is preprocessed to enrich the features of the signal. Then the deep convolution neural network with dual attention mechanism is used to extract the features of the pre-processed data. Finally, the full connected (FC) layer receives and identifies these features to realize the classification of ECG signals.

3.2. Preprocessing

The two-dimensional image of ECG signal is shown in Fig. 1, but it does not contain insufficient feature information. In order to enrich the characteristics of ECG signal, we can carry out time–frequency analysis of the signal, and convert the signal from the time domain to the time–frequency domain. The time–frequency signals can show the different frequency components of the signal at different times, so that the network can focus on more frequency characteristics of the ECG signal. The time–frequency signals can also be used to extract features, with intuitive graphic expression, facilitate the visualization of data, and enhance the richness of data sets. Therefore, converting the signal from time domain to time–frequency domain can help us to understand the composition and characteristics of the signal more deeply, and better analyze, process, and apply it.

Short-Time Fourier Transform (STFT) is a commonly used time–frequency analysis method, which can decompose a signal into multiple components in the time–frequency domain. Compared with the wavelet transform, it can analyze the transient characteristics and frequency changes of signals, and the calculation is faster, which is conducive to the preprocessing of MIT-BIH data set. Different from the traditional Fourier transform, STFT divides the signal into several time periods, transforms the signal of each time period to obtain the frequency domain information, and finally get the time–frequency distribution of the signal. Meanwhile, the characteristic peak of ECG signal is

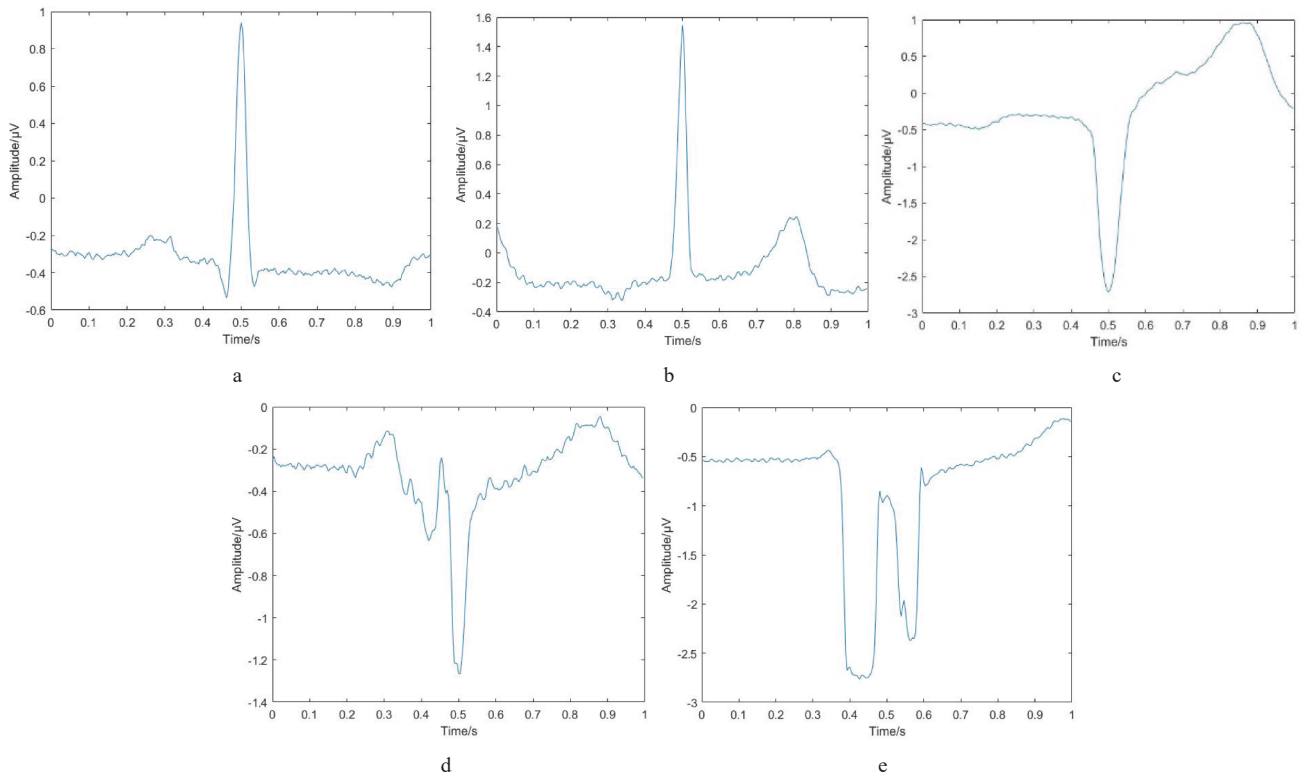


Fig. 1. 1-a to 1-e are ECG signals of N, S, V, F and Q beats.

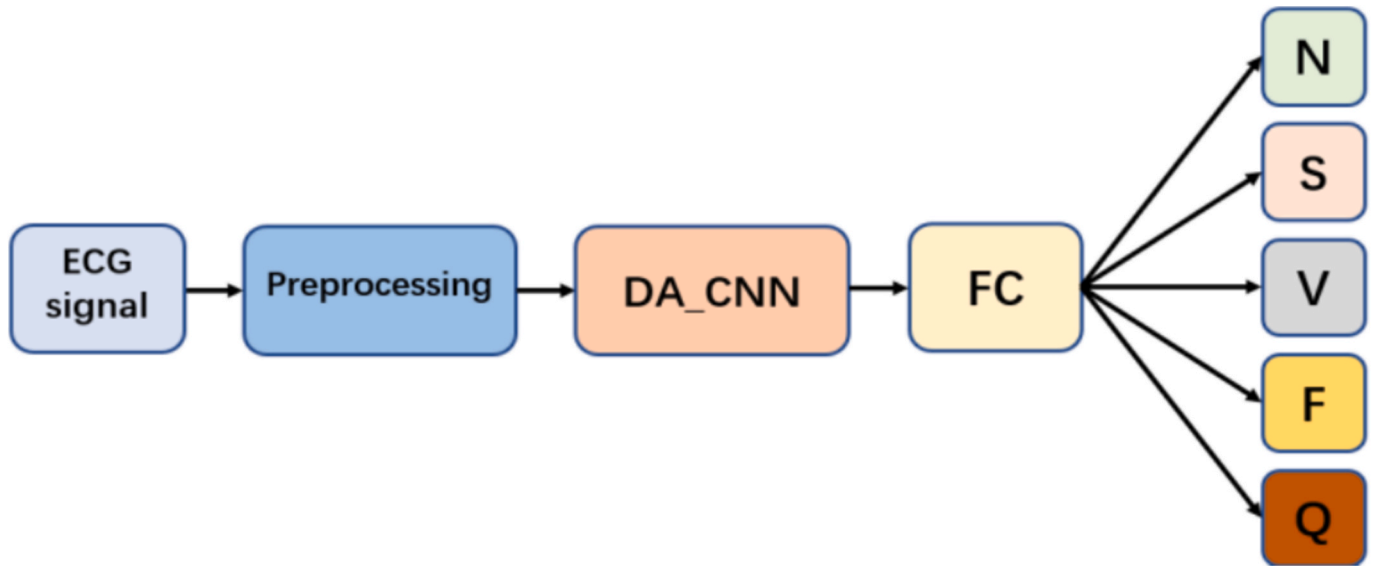


Fig. 2. The structure of the classification system.

concentrated in a small period of time, which is consistent with the frequency information of the segmented signal in STFT. Therefore, STFT can satisfy our pre-processing of ECG signals, enrich the characteristics of signals, and make the network learn better.

When the length and shape of window function are different, STFT can get the results with different time resolution and frequency resolution. The mathematical expression of STFT is shown in

$$STFT(\tau, \omega) = \int_{-\infty}^{+\infty} [f(t) * g(t - \tau)] * e^{-j\omega t} dt \quad (1)$$

where $f(t)$ is a time-domain signal, $g(t)$ is a window function, τ is

step size. $g(t)$ and $e^{-j\omega t}$ play the role of time limit and frequency limit, respectively. Together, they can achieve the double limit of time and frequency.

STFT is performed on the five ECG beats shown in Fig. 1, where the window function is hamming window and the moving step of hamming window is 3. The time-frequency diagram of the ECG signal is shown in Fig. 3. The experimental results show that the images produced by different types of ECG after STFT are different. N and S beats have smaller amplitudes, F and V beats have slightly larger amplitudes, and Q beat have the largest amplitudes. These subtle differences are difficult to distinguish in the eyes of ordinary people, but after the feature

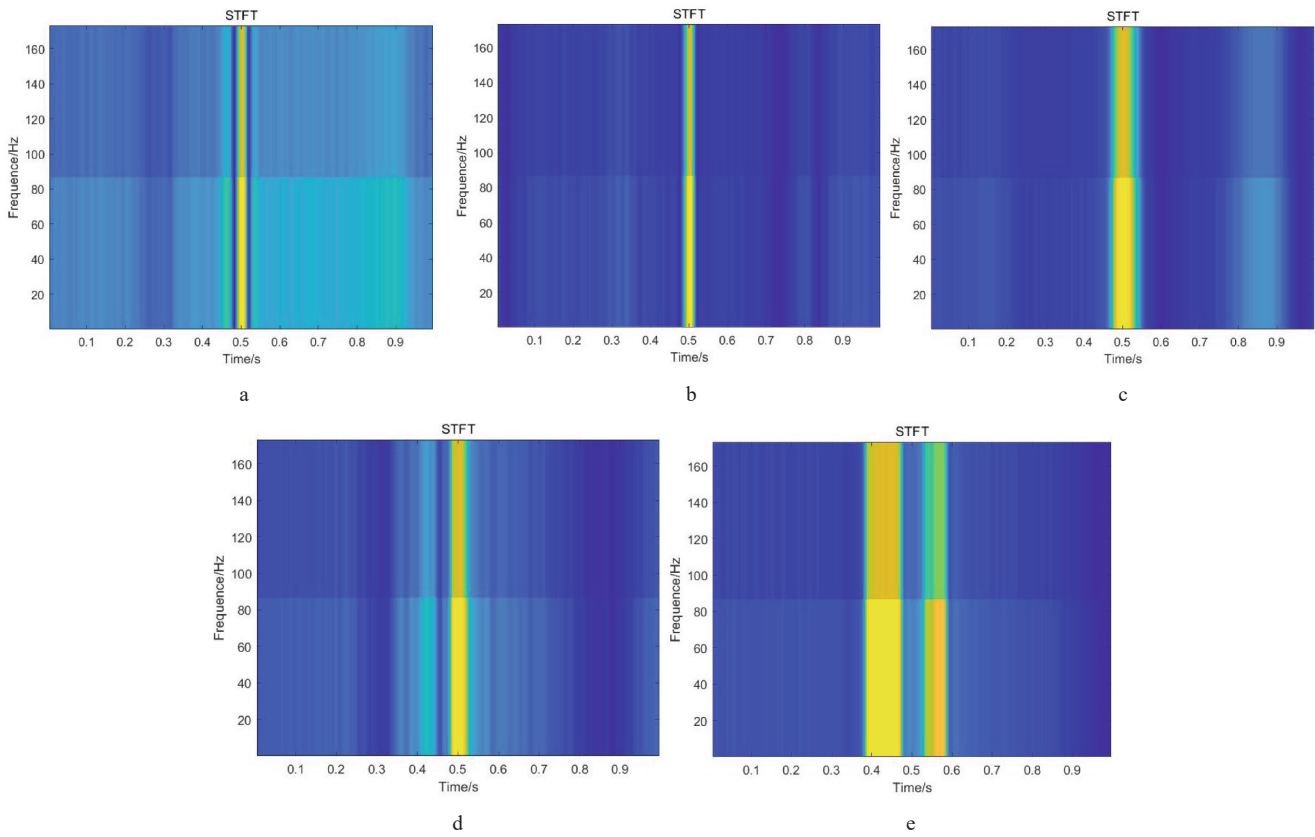


Fig. 3. 3-a to 3-e shows the time–frequency analysis of N, S, V, F, and Q beats of the ECG signal.

extraction of deep convolution neural network, these features information will be exposed to the network, providing a prerequisite for the classification of the network.

3.3. DA-DCNN structure

To classify and predict STFT images of signals in MIH-BIH dataset, we use deep learning convolution network to construct DA-DCNN shown

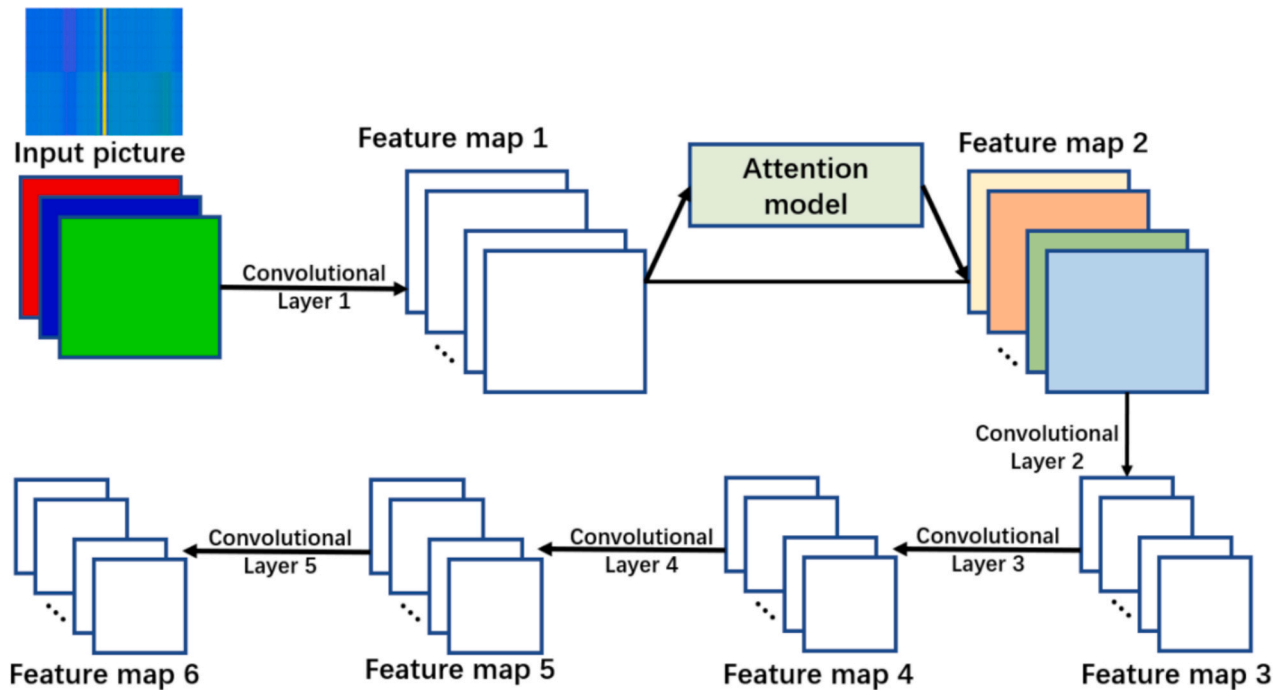


Fig. 4. Deep convolutional neural network framework based on dual attention mechanism. The pre-processed ECG signals are assigned feature weights through Attention model to focus on important feature channels, and feature fusion is performed by subsequent convolutional layers.

in Fig. 4. The network has five convolution layers, and a dual-channel attention mechanism is added after the first convolution layer.

In DCNN, convolution operation can effectively extract useful features from images or sequences, such as edges, textures, shapes, etc., which can be used to distinguish key information of different categories. After each convolution, data normalization is performed using Batch Normalization (BN) [24–27], which can speed up the convergence of model training, make the process of model training more stable, and avoid gradient explosion or gradient disappearance. Then the activation function ReLU introduces nonlinear factors into the network to enable it to deal with nonlinear problems [28–30] and improve the expression ability. Finally, MaxPooling is used to downsample the input data [31], which can reduce the amount of data and calculation, and extract the main features of input data.

The specific convolution process is shown in Fig. 5, and the input image size is $3 \times 224 \times 224$, where 3 represents the three channels of the image RGB (Red, Green, Blue, respectively), 224×224 represents the size of the matrix that makes up the image, that is, each channel has 224×224 pixels. When the picture passes through the first convolution layer, the convolution core is 48, which can be understood as the extraction of 48 different features or the convolution of the image has 48 channels. The size of each convolution kernel is 7×7 , and the step size of convolution is 3, that is, the next convolution after each convolution of the first layer moves 3 pixels to the right. Convolution plus padding is 1. After the convolution layer, BN linear correction unit and ReLU activation function are added to change the negative elements in the feature image into zero, thus increasing the nonlinear elements of the network. Finally, the characteristic diagram of $48 \times 74 \times 74$ is obtained. Then through the maximum pooling layer, the size of the pooling layer is a matrix of 2×2 , and the pooling step is 2. Therefore, the size of the output image is $48 \times 37 \times 37$ feature images. The image size after each convolution and each pooling can be calculated by formula (2):

$$X = \frac{M + 2 \cdot P - N}{S} + 1 \quad (2)$$

where M is the size of the input image, N is the size of the convolution or pooling core, P is the size of the padding, S is the step size of convolution or pooling, and X is the size of the convoluted image.

According to the calculation method of formula (2), the input ECG signal can pass through the network convolution channel shown in Fig. 5, and the output picture with the size of $128 \times 6 \times 6$ can be obtained, which is the Feature Map 6 in Fig. 4. The output of feature graph after Attention Model does not affect its own size, but only generates a new channel weight to improve the network's attention to important features.

Attention mechanism is a technology used in the fields of natural language processing and computer vision [17,32]. It can automatically select the important parts from the input data for weighted aggregation, improving the performance of the model. The basic idea is that for each position in the input sequence, the model can calculate the correlation

between the position and the target output, assign a weight to each position, and then use those weights to weight the input sequence to get the output. The application of attention mechanism in convolution neural network is mainly to enhance the network's attention to important information by weighted average. Specifically, the attention mechanism divides the input feature map of the ECG signal into multiple channels, and then calculates the weight of each channel. For different types of heartbeat signals, the key features and local details we need to pay more attention to are different. Therefore, the attention mechanism can give more weight to these important features and details, increasing their influence in the network and making the network pay more attention to these features in the subsequent deep learning, while reducing the amount and depth of network learning of some of the same features. This can improve the accuracy and generalization performance of the network, accelerate the convergence speed to the optimal value, and enhance the robustness of the network.

Fig. 6 shows the dual attention mechanism proposed in this paper. In the attention mechanism, the network selects the channel for the input feature by learning the channel weight. To put it simply, the network acquires channel weights by introducing an attention module that aggregates each channel of the input feature map into a global pool. These channel weights are used to weight each channel in the input feature map to obtain a new feature map with more emphasis on the important channels.

For the input tensor x , its size is usually $[C, H, W]$, where c represents the number of channels, H and W represent the height and width, respectively. Such As, $x \in \mathbb{R}^{C \times H \times W}$.

The attention module contains two sub-modules, one is SoftMax sub-module and the other is sigmoid sub-module, which use similar processing methods. The following shows the specific calculation form of the SoftMax sub-module:

First, the SoftMax submodule extracts features through a series of convolution layers to generate $\alpha \in \mathbb{R}^{1 \times H \times W}$:

$$\alpha = \text{Softmax}(f_\theta(x)) \alpha = \text{Softmax}(f_\theta(x)) \quad (3)$$

$$\begin{aligned} f_\theta(x) &= \text{Conv2d}(\text{BN}(\text{ReLU}(\text{Conv2d}(x)))) f_\theta(x) \\ &= \text{Conv2d}(\text{BN}(\text{ReLU}(\text{Conv2d}(x)))) \end{aligned} \quad (4)$$

where the f_θ is a sub-network composed of Convolution, Batch Normalization, and ReLU functions and controlled by trainable parameter θ .

Similarly, the Sigmoid submodule can also extract features through a series of convolutional layers $g_\psi g_\phi$ to generate $\beta \in \mathbb{R}^{1 \times H \times W}$ $\beta \in \mathbb{R}^{1 \times H \times W}$:

$$\beta = \text{Sigmoid}(g_\psi(x)) \quad (5)$$

$$g_\psi(x) = \text{Conv2d}(\text{BN}(\text{ReLU}(\text{Conv2d}(x)))) \quad (6)$$

The attention weight $\gamma \in \mathbb{R}^{C \times 1 \times 1}$ $\gamma \in \mathbb{R}^{C \times 1 \times 1}$ is smooth-controlled to integrate the attention generated by each spatial position, in the form of:

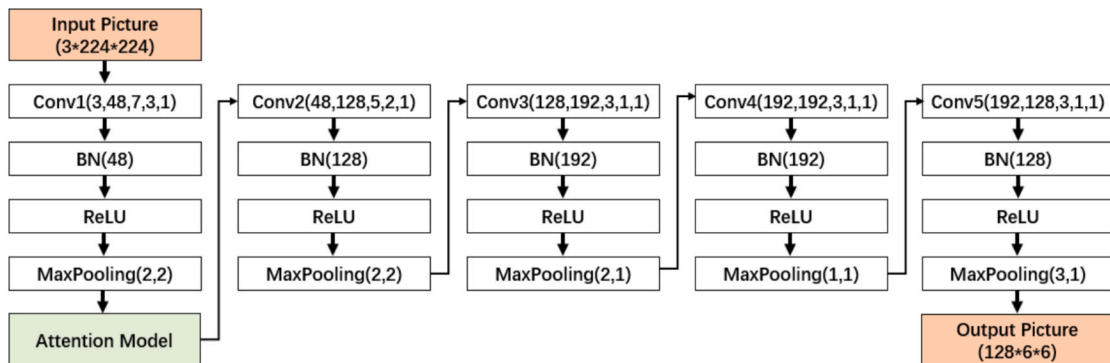


Fig. 5. Construction of deep convolutional neural networks.

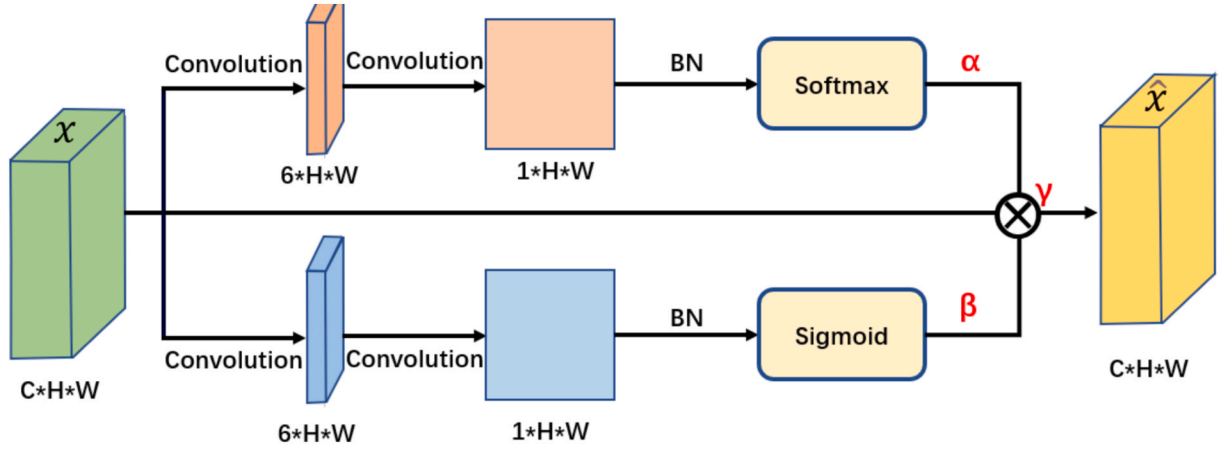


Fig. 6. Construction of dual attention mechanism.

$$\hat{x} = \gamma(x \odot \alpha \odot \beta) + x \quad (7)$$

Specifically, in DA_DCNN model, the attention module extracts signal features through two convolutional layers and transmits these features to a group of sub-modules based on SoftMax and Sigmoid functions. These sub-modules can automatically calculate the attention weights of the feature map in each spatial location to show how much input information is introduced in different locations and how much contribution to the final output of the task. These attention weights are used by the model to adjust the importance of each pixel in the feature map for classification by the network. This process will enable the network to capture the interaction more accurately between input data points and improve the performance of the model when dealing with more complex and noisy data. And the attention module uses the learnable parameter gamma for smooth control. When considering the interaction between the input data, the attention weights and the residual structure, it can effectively use the representations that the network has learned to further optimize the parameters of the model.

3.4. Fully connected layer

Through the previous DA-DCNN, we get a 128*6*6 Feature Map 6 containing the characteristic elements of ECG signal. The full connection layer [33,34] converts high-dimensional input features into more discriminative low-dimensional features by learning appropriate weight parameters. This feature abstraction capability enables the network to

extract more meaningful and differentiated features from the original data, which helps to improve the classification performance.

The structure of the fully connected layer is shown in Fig. 7. Firstly, the Feature Map 6 is expanded into 4608 feature values, and then connected into 2048 feature values through the first fully connected layer. Then, 1024 feature values are connected through the second layer of fully connected layer, and then 512 feature values are connected through the third layer of fully connected layer. Dropout technology has been added to the fully connected layers of the second and third layers. Dropout randomly disrupts signal flow paths in the network by dropping some neurons (that is, setting the output of those neurons to 0), preventing the model from overfitting any particular neuron or combination of neuron. It can improve the generalization ability of the network, reduce the dependence on some features, make the network learn more effective features, and improve the robustness of the network. Finally, through the fourth layer of the fully connected layers, the EGC signal is divided into five categories, namely F, N, S, V, Q.

4. Results and discussion

4.1. Training data

Training data is the basis of the training model and plays an important role in deep learning. In the process of network training, the STFT time-frequency image of ECG signal in MIT-BIH dataset is divided into training set, validation set and test set, with a ratio of 8:1:1. The

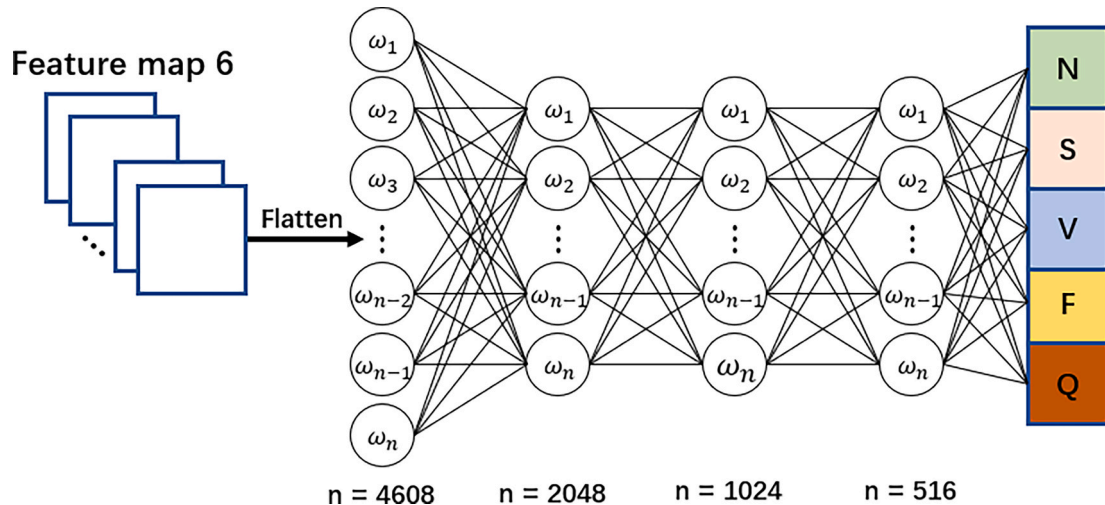


Fig. 7. Fully connected layer.

number of each ECG signal in the dataset is shown in Table 1.

The training set and validation set are sent into the network for training. The addition of the validation set can help the DA_DCNN network in parameter selection. For example, the selection of hyperparameters such as network structure and learning rate, it is necessary to use the validation set to evaluate the model performance under different hyperparameter combinations, so as to find the best hyperparameter Settings. The validation set can also help us to determine when to stop training to avoid overfitting based on the changes in training error and validation error during training. Because when the verification error begins to increase, it means that the model has been overfitted, and continued training may lead to a decline in generalization performance. Therefore, using the validation set to monitor the verification error can detect and stop the training in advance to prevent over-fitting. In addition, it should be noted that the validation, along with the pre-processing and attention model mentioned before can reduce the data imbalance in the MIT-BIH dataset. The training set is the basis for the training of our classification network. The data parameters of the final model are selected based on the training set, and the validation set will not participate in this process. Moreover, these model parameters will affect the classification performance of the classification network on the final test set, so the quality of the training set is crucial to the performance of the classification network.

4.2. Network parameters

The optimizer we use is Stochastic Gradient Descent (SGD). It is one of the most commonly used optimizers in deep learning because it is easy to implement and interpret, and performs well on large data sets. The optimization process of SGD is to update the model weight parameters by calculating the loss of each small batch of data points to gradually reduce the global loss function. In addition, because it does not consider the average gradient of all data samples, but updates the parameters on a batch basis, it can reduce computational costs and alleviate memory bottlenecks.

As a rule of thumb, the number of network training is 30. The more training times, the better the convergence of the network, but it will also increase the training time and computational cost. The 30 training sessions provide some assurance that the model can fully fit the training data set and try to strike a balance between the effect gain and the computational cost of additional training times. The initial learning rate is 10^{-3} , which is reduced by 0.5 after 10 epochs. The learning rate refers to the factor that controls the updating step of the model weight, which indicates that the weight of each node in the previous layer must be adjusted in each iteration. The smaller the learning rate, the slower the convergence speed, while the larger the learning rate, the model may oscillate, not converge or diverge in the training process.

4.3. Evaluation index

In order to measure the performance of the proposed classification network, we introduce the parameters in (8) – (12) to describe the accuracy of the classification network:

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (8)$$

Table 1
Network training parameter.

ECG type	Total	Training set	Validation set	Test set
N	98,600	78,880	9860	9860
S	2781	2225	278	278
V	7235	5792	723	723
F	802	642	80	80
Q	33	27	3	3

$$Pre = \frac{TP}{TP + FP} \quad Pre = \frac{TP}{TP + FP} \quad (9)$$

$$Re = \frac{TP}{TP + FN} \quad Re = \frac{TP}{TP + FN} \quad (10)$$

$$Spec = \frac{TN}{TN + FP} \quad Spec = \frac{TN}{TN + FP} \quad (11)$$

$$F1 = 2 \frac{Pre \cdot Re}{Pre + Re} \quad F1 = 2 \frac{Pre \cdot Re}{Pre + Re} \quad (12)$$

where TP is True Positive, FN is False Negative, FP is False Positive, TN is True Negative. For class S, TP is the number of S beats that are correctly predicted as S beats and FN is the number of S beats that are mistakenly identified as other beats. FP is the number of other beats which predicted as S beats and TN is the number of other beats which did not predict as S beats.

Among them, the accuracy (Acc) is the proportion of the number of samples with correct classification to the total number of samples. Precision (Pre) is the percentage of samples that really belong to that category among all the samples classified into that category. The recall rate (Re) is the percentage of samples that are successfully classified into that category among all the samples that belong to that category. The higher the index, the stronger the ability of the classification model in detecting positive examples. In some cases, the recall rate may be more important than other evaluation indicators. Specificity (Spec) is one of the indicators used to evaluate the quality of the classifier in the classification problem, which refers to the ability of the classifier to correctly predict negative samples. F1 value (F1) is the harmonic average of accuracy and recall, indicating the comprehensive performance of the classifier.

4.4. Experimental results

First, STFT is used to process the data set, and then the processing results are sent to the DA-DCNN for training, and an excellent ECG signal classifier is obtained. Finally, the test set is fed into the classifier for verification, and the confusion matrix of the classification results is obtained, as shown in Table 2.

The confusion matrix has five rows and five columns. The row represents the real label and the column represents the label predicted by the classifier. For example, the second column of the first row indicates the number of real label N beats classified into prediction label S beats. It can be seen from Table 2 that most of the classification prediction results of ECG signals are concentrated on the diagonal line of the confusion matrix, which are labels that are correctly classified. Therefore, this shows that the accuracy of the network classifier is high, and the proposed classification system can be used to classify ECG signals well. Most of the predicted wrong beats are concentrated in N beats, because N beats are mainly normal heartbeats, and most of the features of ECG signals that cannot be accurately identified are more likely to be thought as N beats. The Kappa coefficients can be calculated according to Table 2, which reaches 0.94. It is higher than the Kappa coefficients for [10,36,37] and [15] through the confusion matrix, which are 0.82, 0.86, 0.86 and 0.89, respectively. These results further prove the excellent

Table 2
Confusion matrix of MIT-BIH dataset validation results.

Beats	Classification result				
	N	S	V	F	Q
N	9829	16	9	6	0
S	33	244	1	0	0
V	22	1	696	4	0
F	9	0	8	63	0
Q	0	1	1	0	1

performance of DA-DCNN network for ECG signals recognition.

4.5. Result discussion

To further facilitate the analysis of the performance of the proposed ECG signal classifier, the confusion matrix in Table 2 can be calculated based on the calculation method of the network performance index in (8)–(12), and the performance indexes are shown in Table 3. As can be seen from Table 3, N-beat has the best performance. The Acc, Pre and Re of N beat are 99.13 %, 99.35 % and 99.69 %, respectively, all of which are more than 99 %. This proves that the classification effect of the proposed ECG signal classifier is very good. The specificity of N beats is also very high, reaching 94.01 %, which can be well identified from other types of beats. The comprehensive score F1 of N-beat heartbeat also reaches 0.995, indicating that the classifier can recognize normal heartbeat well. The performance of S, V and F categories are also good, and the indexes are basically above 90 %. Although the Re of the Q category is low, since the Q beat is an unrecognizable heartbeat, the network has no effective features to learn from.

To further measure the performance of our ECG signal classifier, according to the performance parameters in Table 3, we compare the classification results of the proposed method with those of some recent ECG signal classification methods. The classification and comparison results of S-beat and V-beat are shown in Table 4 and 5.

As you can see from Table 4, the ECG signal classifier proposed in this paper has very good performance in dealing with MIT-BIH dataset. Among them, the recognition accuracy of S-beat is 99.5 %, which is 1.1 % higher than the best of other methods. In terms of precision, S-beat has a precision of 93.1 %, which is much higher than all other classification methods. The precision in [15] is 90.1 %, it is also slightly lower than our method by 3 %. The recall rate of S-beat reaches 87.7 %, 11 % higher than the other best in the table. Recall rate represents the ability of the model to correctly detect all real examples. Therefore, recall rate is very important for tasks that need to minimize false negative rates, such as case testing and security checks. If the recall rate is low, it means that the model does not correctly detect some true positive cases, called false negatives. In case testing or security checks, this means that many patients or safety issues are missed, which is very dangerous. Therefore, the importance of recall rate cannot be ignored. For the specificity of the network, S-beat also reaches 99.8 %, slightly higher than the 99.3 % in [38] and 99.5 % in [15]. Compared to other classification methods, due to the great improvement of our work in both Pre and Re, we also succeed in increasing the F1 value of the S-beat to 0.90, which is much higher than other methods.

In Table 5, we can find that the conventional convolutional neural network has a good effect on the identification of V-beat, and can identify V-beat more accurately. However, the proposed DA-DCNN model can identify V-beat ECG signals more accurately. Among them, the recognition accuracy of V-beat was 99.6 %, which was 0.8 % higher than the best recognition rate of other methods. In terms of precision, the precision of V-beat is 97.3 %, compared with other methods, we have improved the precision by nearly 3 %. The recall rate of V-beat reached 96.3 %, 1.3 % higher than the recall rate in [10]. However, other performances in [10] are not high. For example, the precision is only 89.5 %, 7.8 % lower than ours. The V-beat in Table 5 all had high specificity, above 98 %. However, our specificity was the highest,

Table 3
Classification performance of the network.

Class	Acc (%)	Pre (%)	Re (%)	Spec (%)	F1
N	99.13	99.35	99.69	94.01	0.995
S	99.52	93.13	87.77	99.83	0.904
V	99.57	97.34	96.27	99.81	0.968
F	99.75	86.30	78.75	99.91	0.824
Q	99.98	1	33.33	1	0.5

Table 4

Comparison of S-beat classification performance.

Methods	Acc (%)	Pre (%)	Re (%)	Spec (%)	F1
T. Ince [35]	96.1	56.7	62.1	98.5	0.59
S. Kiranyaz [10]	96.4	62.1	64.6	98.6	0.63
X. Zhai [36]	97.5	74.0	76.8	98.7	0.75
Y. Xia [38]	97.8	82.9	67.0	99.3	0.74
Q. Xie [15]	98.4	90.1	75.6	99.5	0.82
Our work	99.5	93.1	87.8	99.8	0.90

Table 5

Comparison of V-beat classification performance.

Methods	Acc (%)	Pre (%)	Re (%)	Spec (%)	F1
T. Ince [35]	97.6	87.4	83.4	98.1	0.85
S. Kiranyaz [10]	98.6	89.5	95.0	98.1	0.92
X. Zhai [36]	98.6	92.4	93.8	99.2	0.93
Y. Xia [38]	98.6	93.8	91.3	99.4	0.93
Q. Xie [15]	98.8	94.5	92.8	99.4	0.94
Our work	99.6	97.3	96.3	99.8	0.97

reaching 99.8 %, slightly higher than 99.4 % in [15] and [38]. Compared with other classification methods, we further improve the work of Pre and Re of V-beat, and maintain the accuracy of Acc and Spec. We also improved the F1 score of V-beat to 0.97, which was higher than other classification methods. Moreover, by comparing Table 4 with Table 5, it can be found that compared with the classification of S beats, the classification effect of the network for V beats is better. And the proposed method provides the possibility for the clinical application of ECG signal classification network, and can replace cardiologists for the preliminary diagnosis and treatment of some conditions.

In order to synthesize the overall performance of the network and compare it with other different classification methods, we define the following overall accuracy parameters:

$$Accuracy = \frac{\sum_{i=1}^n A_{ii}}{\sum_{i=1}^n \sum_{j=1}^n A_{ij}} \quad Accuracy = \frac{\sum_{i=1}^n A_{ii}}{\sum_{i=1}^n \sum_{j=1}^n A_{ij}} \quad (13)$$

where A_{ij} represents the element in row i and column j of the confusion matrix.

According to (13), we compare the overall accuracy with different heartbeat classification methods in recent years, and the results are shown in Table 6. It can be seen from Table 6 that the classification performance of CNN can be improved by pre-processing and using different methods to learn or process the characteristics of ECG signals. Essentially, these methods are designed to enable CNN to learn the different features of ECG signals more efficiently. With the improvement

Table 6

The comprehensive performance of the method is compared with that of various networks in recent years.

Article	Pre-processing	Method	Network	Accuracy
U. R. Acharya [11]	Daubechi es wavelet 6 filters	—	CNN	94.30 %
X. Zhai et al. [36]	—	Dual-beat coupling matrix	CNN	96.05 %
Q. Xie et al. [15]	Chebyshevtype I bandpass filter	STFT	CNN	96.97 %
F. Zhou et al. [39]	—	LSTM	CNN	98.03 %
N. Alamatsaz [40]	Median Filter	LSTM	CNN	98.24 %
Z. Dokur et al. [13]	—	Walsh functions	CNN	98.70 %
A. Shoughi et al. [41]	SMOTE	BLSTM	CNN	98.71 %
Our work	STFT	Attention Model	CNN	98.98 %

and introduction of different methods, the overall classification accuracy of the network is gradually improved. It is worth noting that the use of Walsh function in [13] enriches the features of ECG signals, and the overall accuracy is 98.70 %. Reference [41] uses BLSTM method to extract local features and correlation of data and the overall accuracy is 98.71 %. However, compared with other work, our DA-DCNN model performs the best and the overall accuracy of ECG signal classification for MIT-BIH dataset reaches 98.98 %. This shows that the network based on attention mechanism has good performance in the heartbeat classification model, which can make the network pay more attention to important features and improve the learning efficiency of the network, so as to better classify heartbeat signals and improve the classification accuracy and robustness of the network.

5. Conclusion

Heartbeat signals are complex and specific, including heart rate, rhythm, time series and other information. Each person's heartbeat signals are unique and vary individual to individual. However, heartbeat signals have some similarities in waveforms and peaks that can be used for identification and monitoring. Therefore, heartbeat classification technology plays a very important role in modern medical field and health monitoring, and its research has a wide range of practicability and application prospects for improving heart health, and preventing heart disease.

In general, due to the influence of noise, ECG signals lack feature richness, and it is difficult to extract features effectively, which makes it unable to pay more attention to the important features of ECG signals. Therefore, this paper first adopts the time-frequency analysis method to enhance the features richness of the initial data. Then, the DA-CNN model is constructed by introducing attention mechanism into the network. The model can learn more comprehensive and in-depth detailed features, and pay more attention to important features, and thus better complete the classification task with high classification accuracy and robustness.

The test results of arrhythmia signals in MIT-BIH dataset show that DA-DCNN model can accurately classify N, S, V, F and Q ECG signals with high classification accuracy. Compared with other methods, the model can classify and predict S-beat and V-beat more accurately. The total correct rate of the network is also higher than other classification methods, reaching 98.98 %. The network uses 10 % of the data in the MIT-BIH dataset as a validation set for cross-validation, and this system can be repeated. These results prove that our classification method can achieve higher performance in the classification of single S-beat or V-beat. And on the problem of multi-classification, the network can also maintain a high accuracy. In the future work, we can also build a stronger network, and pay more attention to data imbalance and noise problems and other factors, to achieve a more effective ECG signal classification algorithm.

CRedit authorship contribution statement

Xiaoshan Zhang: Writing – original draft. **Xiaolong Chen:** Validation. **Yinwei Li:** Writing – review & editing. **Yonggang Luo:** Data curation. **Yiming Zhu:** Funding acquisition.

Funding

This research was supported by the Artificial Intelligence Promotes Scientific Research Paradigm Reform and Empowers Discipline Advancement Plan under Grant Z-2024-302-061, in part by the Natural Science Foundation of Shanghai under Grant 21ZR1444300, in part by the National Nature Science Foundation of China under Grants 12105177 and 61988102. (Corresponding author: Yinwei Li.).

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.

Data availability

Data will be made available on request.

References

- [1] World health statistics 2022: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.
- [2] Y. Yeh, W. Wang, C. Chiou, A novel fuzzy c-means method for classifying heartbeat cases from ECG signals, *Measurement* 43 (2010) 1542–1555.
- [3] T. Li, M. Zhou, ECG classification using wavelet packet entropy and random forests, *Entropy* 18 (2016) 285.
- [4] P. De Chazal, M. O'Dwyer, M. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.* 51 (2004) 1196–1206.
- [5] S.M. Dima, C. Panagiotou, E.B. Mazomenos, J.A. Rosengarten, K. Maharatna, J. V. Gialelis, On the detection of myocardial scar based on ECG/VCG analysis, *IEEE Trans. Biomed. Eng.* 60 (2013) 3399–3409.
- [6] P. De Chazal, R.B. Reilly, A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features, *IEEE Trans. Biomed. Eng.* 53 (2006) 2535–2543.
- [7] Y. Kutlu, D. Kuntalp, Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients, *Comput. Methods Programs Biomed.* 105 (2012) 257–267.
- [8] G. De Lannoy, D. Francois, J. Delbeke, M. Verleysen, Weighted conditional random fields for supervised interpatient heartbeat classification, *IEEE Trans. Biomed. Eng.* 59 (2011) 241–247.
- [9] H. Yang, C. Kan, G. Liu, Y. Chen, Spatiotemporal differentiation of myocardial infarctions, *IEEE Trans. Autom. Sci. Eng.* 10 (2013) 938–947.
- [10] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, *IEEE Trans. Biomed. Eng.* 63 (2016) 664–675.
- [11] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, A deep convolutional neural network model to classify heartbeats, *Comput. Biol. Med.* 89 (2017) 389–396.
- [12] A.Y. Hannun, P. Rajpurkar, M. Haghighpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.* 25 (2019) 65–69.
- [13] Z. Dokur, T. Olmez, Heartbeat classification by using a convolutional neural network trained with Walsh functions, *Neural Comput. Appl.* 32 (2020) 12515–12534.
- [14] D.U. Jeong, K.M. Lim, Convolutional neural network for classification of eight types of arrhythmias using 2D time-frequency feature map from standard 12-lead electrocardiogram, *Sci. Rep.* 11 (2021) 20396.
- [15] Q. Xie, S. Tu, G. Wang, Y. Lian, L. Xu, Feature enrichment based convolutional neural network for heartbeat classification from electrocardiogram, *IEEE Access* 7 (2019) 153751–153760.
- [16] J. Hu, L. Shen, G. Sun, Squeeze-and-Excitation Networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, 7132–7141.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, In *NIPS*. (2017) 6000–6010.
- [18] L. Chen, S. Tan, Z. Pan, J. Xing, Z. Yuan, X. Xing, P. Zhang, A new framework for automatic airports extraction from SAR images using multilevel dual attention mechanism, *Remote Sens.* 12 (2020) 560.
- [19] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Tang, Residual Attention Network for Image Classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, 3156–3164.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 2818–2826.
- [21] R. Sultan, M. AbuElkheir, Sequential Recommendation Using Deep Reinforcement Learning and MultiHead Attention. In *2022 56th Annual Conference on Information Sciences and Systems (CISS)*. 2022, 258–262.
- [22] P. Wen, W. Yuan, Q. Qin, S. Sang, Z. Zhang, Neural attention model for recommendation based on factorization machines, *Appl. Intell.* 51 (2021) 1829–1844.
- [23] R. Mark, G. Moody, MIT-BIH Arrhythmia Database Directory. 1988.
- [24] Y. Wu, K. He, Group Normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, 3–19.
- [25] S. Ioffe, S. Christian, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167* (2015).
- [26] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How does batch normalization help optimization? *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [27] S. Ioffe, Batch renormalization: Towards reducing minibatch dependence in batch-normalized models, *Adv. Neural Inf. Process. Syst.* 30 (2017).

- [28] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML-10). 2010, 807–814.
- [29] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural. Inf. Process Syst.* 25 (2012).
- [30] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, *Proc. Icml.* 30 (2013) 3.
- [31] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 770-778.
- [32] S. Ren, K. He, R. Girshick, J. Sun, Faster RCNN: towards RealTime object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 1137–1149.
- [33] Y. Bengio, Deep learning of representations: Looking forward. In International conference on statistical language and speech processing. 2013, 1-37.
- [34] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* (2015) 436–444.
- [35] T. Ince, S. Kiranyaz, M. Gabbouj, A generic and robust system for automated patient-specific classification of ECG signals, *IEEE Trans. Biomed. Eng.* 56 (2009) 1415–1426.
- [36] X. Zhai, C. Tin, Automated ECG classification using dual heartbeat coupling based on convolutional neural network, *IEEE Access* 6 (2018) 27465–27472.
- [37] H. Li, J. Han, H. Zhang, X. Zhang, Y. Si, Y. Zhang, Y. Liu, H. Yang, Clinical knowledge-based ECG abnormalities detection using dual-view CNN-Transformer and external attention mechanism, *Comput. Biol. Med.* 178 (2024) 108751.
- [38] Y. Xia, N. Wulan, K. Wang, H. Zhang, Detecting atrial fibrillationn by deep convolutional neural networks, *Comput. Biol. Med.* 93 (2018) 84–92.
- [39] F. Zhou, L. Jin, J. Dong, Premature ventricular contraction detection combining deep neural networks and rules inference, *Artif. Intell. Med.* 79 (2017) 42–51.
- [40] N. Alamatsaz, L. Tabatabaei, M. Yazdchi, H. Payan, N. Lamatsaz, F. Nasimi, A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection, *Biomed. Signal Process. Control* 90 (2024) 105884.
- [41] A. Shoughi, M.B. Dowlatsahi, 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). 2021, 1-6.