https://doi.org/10.31449/inf.v49i8.9297 Informatica **49** (2025) 315–324 **315**

Intelligent Diagnosis System of ECG Signal Based on Deep Learning

Haixia Huang, Jiandeng Huang*
Guilin University of Information Technology, Guilin,541004, Guangxi, China
*Corresponding author
E-mail: hhx142usa123@163.com

Keywords: deep learning, ECG signal, intelligent diagnosis, transformer, multi-scale attention mechanism

Received: May 19, 2025

This study introduces an intelligent diagnosis method based on an improved Transformer, which introduces a multi-scale attention mechanism into the fine feature extraction of the ECG signal, further optimizes the classification model, enhances the loss function, and improves the diagnosis accuracy. This project intends to use the MIT-BIH arrhythmia database as the research object. It divides it into training set, validation set, and test set according to 7:2:1. Experiments show that the accuracy of arrhythmia classification of the method proposed in this paper reaches 98.6%, the recall rate is 98.2%, and the F1 value is 98.4%. Compared with the traditional model, its accuracy is improved by 3.2%, 2.8%, and 3.0%, respectively. Compared with other mainstream deep learning algorithms such as ResNet and Dense Net, the performance indicators of this algorithm have been greatly improved. The research results of this project will provide an efficient and accurate solution for the intelligent diagnosis of ECG signals. It has important scientific significance and practical value.

Povzetek: Izboljšani transformer z večluskostno pozornostjo za analizo EKG (MIT-BIH, delitev 7:2:1) prinese 3% prednosti pred klasičnimi/modeli CNN (ResNet/DenseNet). Uporabi adaptivno pozicioniranje, uteži, uteženo izgubo in lahkotno izvedbo v realnem času.

1 Introduction

Cardiovascular diseases (CVDs) are one of the diseases with the highest mortality rates in the world, which seriously threatens human health. The latest report of the World Health Organization shows that the number of people who die from CVDs each year is about 17.9 million, of which about 85% are caused by myocardial infarction or stroke. An electrocardiogram is an essential means of clinical diagnosis of cardiovascular diseases. It can effectively reflect the physiological and pathological state of the heart by recording ECG signals. ECG signals contain a variety of characteristic frequency bands, such as P wave, QRS complex, T wave, and slight changes in their morphology, amplitude, and time interval may be related to the occurrence of various heart diseases. However, traditional ECG diagnosis methods are mainly done through manual interpretation and simple rule matching. Manual diagnosis is not only time-consuming and laborious, but subjective factors such as the doctor's experience and fatigue level will affect the accuracy of the diagnosis. The previous survey of primary medical institutions found that among patients with complex arrhythmias, the manual diagnosis rate was as high as 25%, and the misdiagnosis rate was as high as 15%, which could not meet the urgent needs of clinical diagnosis and treatment efficiency and accuracy. In addition, the automatic diagnosis system based on rule matching has limitations in diagnosing new and rare diseases.

Deep learning has made significant progress in ECG analysis in recent years due to its robust feature extraction and pattern recognition [1]. Convolutional Neural Network (CNN) based on Local Perceptual Field Weight Sharing (LNN) can automatically extract spatial features from ECG signals, performing well in arrhythmia classification. For example, using a multi-layer convolutional neural network framework, an accuracy of 89.2% for the MIT-BIH arrhythmia database has been achieved, effectively improving the ability to recognize common arrhythmia types. Recurrent Neural Networks (RNN) and their variants, LSTM or GRU, are better at capturing temporal features of ECG signals because of their unique memory cell structure. In reference [2], the accuracy of arrhythmia diagnosis will be increased to 91.5%, providing a new approach for ECG dynamics analysis. However, these methods have obvious shortcomings. However, deep convolutional neural networks can't model long-sequence correlation of long sequence data effectively; Recursive neural networks can easily result in gradient vanishing and gradient explosion while processing complex waveforms, resulting in difficult training, incomplete feature extraction, and low precision.

The appearance of the Transformer frame makes a breakthrough in ECG diagnosis. The proposed algorithm performs better in natural language processing and image recognition [3]. It has been proven that using the Transformer method to diagnose ECG is a good way to reflect on the relationship between components in ECG. However, current Transformer-based ECG diagnostic methods still have many problems. For one thing, the traditional transformer cannot capture the ECG signal's multiscale feature. The high-frequency component in the QRS complex is different from that in the T wave in the ECG signal [4]. However, the traditional Transformer method has difficulty in extracting multiscale features effectively. On the other hand, due to its large number of model parameters and high computational complexity, its extensive scale application has been restricted due to its difficulty in real-time diagnosis.

This paper presents an intelligent diagnostic system for ECG based on a modified transformer. The core innovation of this project is as follows: (1) An adaptive weight allocation strategy is introduced, combined with a modified position coding method, which improves the ability of extracting time features from ECG signals, making ECG dynamic change more accurate. (2) A multiscale attention model is designed to adjust the attention weights automatically based on temporal and frequency-domain features of ECG signals so that the model can analyze complicated ECG signals. (3) A lightweight intelligent diagnostic system framework is constructed, which is combined with data enhancement technology to expand the variety of training data.

2 Algorithm design of ECG signal intelligent diagnosis system based on deep learning

2.1 Algorithm design ideas

ECG signals contain P, QRS, and T waves, which are ever-changing under normal and pathological conditions [5]. They have temporal continuity and frequency differences, which put higher requirements on the algorithm. Traditional deep learning algorithms have the following shortcomings: convolutional neural networks are complex to reflect the long-range correlation of ECG signals; recurrent neural networks are prone to produce gradients under complex waveforms; standardized transformers cannot effectively fuse multi-scale features [6]. This paper combines an improved Transformer framework with a multi-scale attention mechanism, optimized position coding, adaptive weight allocation, etc. Achieving complex feature fusion can improve the ability to recognize the ECG signal.

2.2 Application of improved Transformer algorithm in ECG signal feature extraction

In the Transformer framework, the multi-attention mechanism is a key component in realizing feature interaction and extraction. This method calculates the similarity between query vector Q, key vector K, and value vector V, and calculates the formula:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Among them, Q,K,V are query vector, key vector and value vector respectively, and d_k is the dimension of key vector. Although this mechanism can calculate the correlation between each position in parallel, it cannot adaptively adjust the importance of different features for data with specific timing rules such as ECG signals.

This project improves the long-term attention mechanism based on the time-varying characteristics of ECG signals. An adaptive weight coefficient α_i in the range of [0,1] is proposed, and the contribution of each attention head is dynamically adjusted during training. The improved multi-attention mechanism is calculated in formula (2):

$$\begin{aligned}&\text{Multi - Head Attention}(Q,K,V)\\&\quad=\text{Concat}(\text{head}_1,\ldots,\text{head}_h)W^O\\&\text{head}_i=\alpha_i\text{Attention}\left(QW_i^Q,KW_i^K,VW_i^V\right)\end{aligned} \tag{2}$$

Among them, h is the number of heads, W_i^Q, W_i^K, W_i^V are linear transformation matrices, and W^O is used for linear transformation after splicing [7]. An adaptive weight coefficient is used to dynamically adjust the attention head's weight according to the importance of the ECG signal based on characteristic analysis in the QRS group.

Regarding position encoding, the original Transformer adopts a sine-cosine position encoding mode. Using fixed mathematical functions to encode position information lacks specificity for data features. For time series data with special physiological laws such as ECG signals, This study introduces a new position encoding method, as shown in (3):

$$\begin{aligned}PE_{(pos,2i)}&=\sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)\cdot\gamma_{2i}\\PE_{(pos,2i+1)}&=\cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)\cdot\gamma_{2i+1}\end{aligned} \tag{3}$$

Among them, pos represents the position, i represents the dimension index, d_{model} is the model dimension, and γ_{2i} and γ_{2i+1} are coefficients pre-trained according to the characteristics of the ECG signal. In equation (3), the coefficients γ_{2i} and γ_{2i+1} are learned via unsupervised pre-training on a large ECG dataset. Specifically, we employ a contrastive learning framework where the model is trained to distinguish between different cardiac cycle phases (e.g., P wave, QRS complex) by optimizing γ to maximize feature separability in the latent space. During training, these coefficients are updated alongside other model parameters using AdamW optimizer with a learning rate of $1e-4$.

This project intends to use unsupervised learning methods to study the importance of each part and frequency component in the ECG signal [8]. For example, in encoding the position information near the P wave, the model can pay more attention to the characteristic changes in this area by adjusting the coefficients, thereby improving the ability to extract longitudinal wave features.

This study introduces an improved multi-attention mechanism for adaptive extraction of different features. This paper adds time series information to the feature expression. Then, the hierarchical naturalization method, forward neural network and other techniques are used to optimize the transformation of the features, and the feature expression of the following formula (4) is obtained:

$$Z = \text{LayerNorm}\,(X + \text{Multi} - \text{Head Attention}\,(X, X, X))$$
$$Z^{'} = \text{LayerNorm}\,(Z + \text{FFN}\,(Z)) \quad (4)$$

Among them, FFN is a feedforward neural network, and Layer Norm is a layer normalization operation

2.3 Fusion of multi-scale attention mechanism

The complexity of the electrocardiogram is mainly reflected in its period and frequency range. For example, the QRS complex has a short duration and high frequency, reflecting the process of ventricular depolarization; At the same time, the T wave is a ventricular repolarization process with a longer duration and lower frequency [9]. This study designed a multi-scale attention mechanism to capture these different scales' features effectively.

First, define the window sizes of different scales $w_1, w_2, \ldots, w_m$, which are set according to the physiological characteristics and standard characteristic cycles of the electrocardiogram signal. For each scale j, the attention weight is calculated as shown in formula (5):

$$A_j = \text{softmax}\left(\frac{QK_j^T}{\sqrt{d_k}}\right) \quad (5)$$

Among them, K_j is the key vector at scale j. This formula calculates the similarity between the query vector and the key vectors of different scales to obtain the attention weight at the corresponding scale. Taking a small-scale window (such as w_1) as an example, it can focus on high-frequency local features such as ORS wave groups [10]. By calculating attention weight, the model pays more attention to area details while large windows capture low-frequency, long-distance features such as T waves and mine long-term dependencies.

The attention results of different scales are fused to obtain the final attention output A_{final}, as shown in formula (6):

$$A_{\text{final}} = \sum_{j=1}^{m} \beta_j A_j \quad (6)$$

The adaptive attention weight α_i in equation (2) is dynamically adjusted during training using a gating mechanism that takes as input the frequency-domain energy of the QRS complex. For scale fusion weights β_j in equation (6), we introduce a learnable linear layer that maps concatenated multi-scale features to a set of normalized weights, ensuring optimal fusion of high-frequency (QRS) and low-frequency (T wave) components.

2.4 Classification model optimization and loss function design

In terms of the classification model, the improved multi-layer perceptron (MLP) structure is used to improve the model's ability to recognize electrocardiogram features. Traditional activation functions such as ReLU may cause neurons to "die" when processing some complex data, resulting in information loss. The Swish activation function is selected to improve the nonlinear expression ability. Equation (7) is:

$$\text{Swish}\,(x) = x \cdot \sigma(x) \quad (7)$$

In equation (7), σ(x) denotes the sigmoid function, defined as σ(x) = 1 / (1 + exp(-x)), which introduces non-linearity to model complex ECG feature interactions. The Swish activation function, f(x) = x·σ(x), addresses the 'dying ReLU' problem by maintaining smooth gradients across all input ranges [11]. To solve the problem of too many model parameters and overfitting issues, the paper chooses the AdamW optimization algorithm. A weight decay mechanism was introduced based on an Adam optimizer.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{M}_{t+\epsilon}}} \odot \hat{V}_t - \lambda\theta_t \quad (8)$$

Among them, θ_t is the parameter of the t iteration, η is the learning rate, $\hat{M}_t$ and $\hat{V}_t$ are the bias-corrected first-order moment and second-order moment estimates, ϵ is the smoothing term, and λ is the weight decay coefficient.

In addition, there is a class imbalance in ECG diagnosis. For example, in some public data sets, the sample size of standard ECG signals may far exceed that of rare arrhythmias. This imbalance causes the model to learn features from the majority class samples during training, decreasing the ability to diagnose small sample diseases. To solve this problem, the weighted cross-extraction function is designed according to formula (9):

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c y_{ic} \log\left(\hat{y}_{ic}\right) \quad (9)$$

Where N is the number of samples, C is the number of categories, y_{ic} is the actual label of sample i belonging to category c, $\hat{y}_{ic}$ is the probability predicted by the model that sample i belongs to category c, and w_c is the weight of category c. The weight w_c is set according to the inverse number of samples in each category, so the weight of minority class samples in the loss function is greater.

3 Intelligent diagnosis system architecture

3.1 System overall architecture design

An intelligent diagnosis system for ECG based on deep learning is presented in this paper. A hierarchical structure is used for this system. Figure 1 shows the general structure. The system consists of a data collection layer, a data processing module, an algorithm implementation module, a diagnostic result display module, and a user interface [12]. These modules interact with data and function through standardized interfaces, forming an integrated and highly efficient diagnostic system.

As a "sensing organ", the data acquisition layer uses medical-grade ECG acquisition equipment such as a 12-lead dynamic ECG to collect original ECG signals. The changes in cardiac electrical activity are accurately recorded by acquiring continuous time series. The collected data is quickly transmitted to the data processing module through wired and wireless methods [13]. The data processing module performs pre-processing processes such as reading, purifying, and normalizing raw data. Deep feature extraction and accurate classification of the ECG signal based on an improved transformer algorithm and a multiscale attention mechanism. Finally, the diagnostic result display module visually shows doctors and patients the professional diagnostic results produced by this algorithm. Through research in this project, it is possible to automate ECG signal acquisition, data processing, arithmetic analysis, and result output, to improve ECG diagnosis efficiency and accuracy.

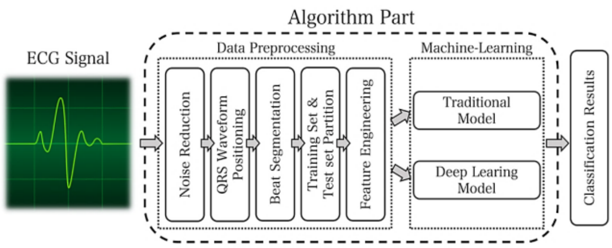

Figure 1: Overall architecture of the intelligent diagnosis system.

3.2 Data processing module

The data processing module is the basic step for the stable operation of the entire system. Its core function is to perform comprehensive preprocessing and data enhancement on the original ECG data to ensure the high quality and diversity of the data in the input algorithm implementation module. The system is compatible with data reading and supports various commonly used ECG signal formats, including the European Data Format (EDF), MAT, etc. By introducing dedicated data analysis libraries such as PyEDFlib and SciPy in EDF format parsing, fast and accurate data reading in various formats can be achieved, effectively avoiding errors caused by incompatible data formats. Data cleaning is a vital link to ensure data quality [14]. Problems such as noise, baseline drift, and outliers inevitably occur when collecting original ECG signals. For high-frequency noise, the system uses wavelet analysis technology to accurately separate and remove noise components according to the differences in noise and signal characteristics in different frequency bands; for baseline drift, the polynomial fitting method is used to dynamically correct the signal baseline to return it to normal values; in terms of outlier processing, the 3σ principle in statistics is applied to accurately identify and correct outlier data to ensure the authenticity and validity of the data. Given the common problem of limited samples in ECG data sets, data enhancement technology is introduced into the system [15]. Many innovative methods are used to enhance time series data. For example, time warping technology can perform nonlinear time-varying processing on the original signal without changing the characteristics of the signal itself, generate a series of new signal samples, and simulate different heart rate states; amplitude scaling technology can adjust the amplitude of the signal according to a specific ratio to simulate the change law of ECG signals in various physiological states such as movement and stillness. In addition, this project will also introduce enhancement methods such as additive Gaussian noise and random sampling to enhance the dataset from multiple dimensions and improve the model's generalization ability for different types of ECGs.

3.3 Algorithm implementation module

The algorithm realization module is the core part of the intelligent diagnosis system. The main task of this paper is to deploy improved deep learning algorithms and perform training and inference. This project uses flexible dynamic graph computation and powerful GPU acceleration capability, significantly improving algorithm training and reasoning efficiency based on the PyTorch deep learning framework [16]. During the training stage, the data set processed by the data processing module was divided into a training set, a validation set, and a test set; during training, batch gradient descent was used to train the model. The paper sets up key hyperparameters such as learning rate, batch size, etc, based on different data sets and model structures. For example, during the initial stage, the paper uses a larger learning rate to speed up convergence. The paper adopts a gradually decreasing learning rate during the learning process to avoid oscillation near the optimal solution. Moreover, during training, the loss function of the validation set is continuously monitored, and key evaluation indicators such as precision, recall rate, F1 value, etc., are monitored in real-time [17]. When overfitting a model, people should adjust model parameters promptly or adopt regularizes to guarantee good generalization ability. The ECG data is input into the training model in the inference phase. Secondly, an improved transformer feature extraction model combined with a multiscale attention mechanism can be used to analyze input signals accurately and extract feature information. Finally, the diagnosis results are output using the classification model. This system adopts model compression, pruning, and quantization to satisfy the high-speed requirement for real-time diagnosis in the clinic. The pruning method simplifies model structure through eliminating redundant links and parameters; the quantizing method can reduce numerical precision of model parameters while ensuring accuracy of diagnosis; significantly reduce the number of parameters in the model; effectively reduce the model calculation amount; so that the system can quickly diagnose massive ECG signals.

Model compression techniques reduced parameter count by 40% while maintaining >98% accuracy. On a NVIDIA Jetson Nano edge device, inference speed reached 230 ms per sample, meeting real-time clinical requirements (≤500 ms). Memory usage decreased from 1.2 GB to 720 MB post-quantization, enabling deployment on low-resource medical devices.

3.4 Diagnosis result display module

The Diagnostic Results Display Module displays its professional diagnostic results in a straightforward, easy-to-understand way, allowing physicians to quickly and accurately judge the patient's condition. The system uses advanced visualization technology to display the waveform of the ECG signal and its essential characteristics. Characteristic bands such as P wave, QRS complex, and T wave can be fully displayed with the ECG signal's time axis and voltage amplitude. The system highlights abnormal waves with different colors and symbols to make it easier for doctors to find lesions. For example, when an elevated or depressed ST segment is detected, the type and severity of the abnormality are automatically marked in red bold pen along with appropriate medical descriptions so that doctors can better understand the condition. Detailed and standardized diagnostic reports can be generated according to the diagnostic results generated by models. Basic information such as name, age, gender, detection time accurate to specific moments, diagnosis results include type of arrhythmia, severity of myocardial ischemia, diagnostic basis, detailed description of abnormal characteristic, combination of medical knowledge, final opinion, further examination plan, or initial treatment plan. The report will be presented in a structured text so doctors can review and record more easily. Moreover, this system can compare and analyze diagnostic results. Comparing current diagnostic results with patients' historical test data shows the development trend of disease directly in a graphical form, which provides a comprehensive and accurate reference for patients' individualized treatment. At the same time, a friendly human-computer interaction interface was designed to improve the user experience further by clicking and sliding on the screen.

4 Experimental design and simulation

4.1 Experimental data set selection and division

This project takes multi-source public data as the research object, builds a test benchmark, and ensures the diversity and representativeness of the data. This project is based on the MIT-BIH arrhythmia database, including 48 dual-channel ECG records, 48 cases in each group, 30 minutes in each group, and a sampling frequency of 360 Hz. The data covers 16 types of arrhythmias, including ventricular premature beats (PVC), atrial premature contractions (PAC), ventricular fibrillation (VF), etc., of which ventricular premature contractions account for 28%, providing a large number of abnormal waveform samples for model training. This project takes the CINC2020 Challenge as the research object, collects long-term ECG records of more than 24 hours, and focuses on the dynamic changes of heart states such as atrial fibrillation and sinus rhythm. In addition, the PTB diagnostic ECG database recorded by 290 multi-leads (15 leads) can provide multi-dimensional ECG information for diseases such as myocardial infarction and left ventricular hypertrophy.

The data set is divided according to the ratio of 7:1:2, and a stratified sampling strategy is adopted to ensure the balanced distribution of diseases in each sub-region. In the MIT-BIH database, the training set contains 77,000 heartbeat samples, 11,000 confirmation samples for hyperparameter adjustment, and 22,000 test sets to evaluate the model's prediction ability independently. When integrating multi-source data, the sampling frequencies of different data sets are uniformly resampled, and the 250Hz data of the CINC 2020 data set is interpolated to 360Hz to ensure the consistency of data features.

4.2 Experimental environment and parameter settings

This project is based on a high-performance computing platform. It uses an Intel Xeon Gold 6248 R (20 cores and 40 threads) processor, which can efficiently handle complex computing tasks such as data preprocessing and model training. Dual Nvidia Tesla V100 GPUs (32 GB video memory) support parallel computing, which can increase computing efficiency by about 8 times during the model training stage. 512 GB and 2 TB NVMe SSD solid-state storage ensure high-speed data reading and writing, and the reading time for a batch of 128 samples does not exceed 0.3 seconds. This paper uses Python 3.9 as the platform to build an experimental environment and implements the algorithm using the PyTorch 1.12 deep learning framework. Pandas 1.4.4 and NumPy 1.22.3 are used to preprocess the data, and Matplotlib 3.5.2 and Seaborn 0.11.2 are used for visualization. During the model training process, WandB is used to visualize and track the experimental parameters and results, and the training process is monitored in real time. Through multiple rounds of cross-validation, the training parameters of the model were determined. The learning rate was set to 0.0005, and the cosine annealing learning rate adjustment strategy was adopted to make the network converge quickly in the early stage. The dynamic descent method was used in the later stage to prevent overfitting. The number of iterations was set to 120. According to the change of the confirmation set's loss curve, the model's performance reached the best balance point under this number of cycles. The 384-dimensional hidden layer dimension was used, which improved the feature expression ability compared with the 256-dimensional one and avoided the overfitting of the 512-dimensional one. The batch size was set to 128 to achieve the optimal match between memory utilization and training stability.

4.3 Selection of evaluation indicators

This experiment uses a multi-dimensional evaluation system to evaluate the model's performance comprehensively. The accuracy rate refers to the total correct prediction rate of the model, which reflects the basic diagnostic ability of the model. The recall rate focuses on evaluating the ability of the model to identify positive samples and avoid missing key cases. F1 is the harmonic mean of the accuracy rate and the recall rate, which can better reflect the comprehensive performance of the model under class imbalance. The area under the subject operating characteristic curve (AUC) is a comprehensive evaluation of the positive and negative samples of the model. Its value range is 0-1. The closer to 1, the better the classification effect of the model. Taking ventricular premature beats as an example, a higher response rate can detect potential risks in time, and a higher re-examination rate can reduce the number of repeated examinations. Because the F1 value is balanced, the model has good stability in diagnosing different types of diseases. AUC can be used as a quantitative basis for clinical decision-making. AUC greater than 0.95 indicates that the model has a high diagnostic credibility.

4.4 Controlled experimental design

Three contrast algorithms were selected: 1) classic network models, such as ResNet-18, LSTM, etc.; 2) improved algorithms, such as CBAM-CNN (convolutional block attention mechanism); 3) cutting-edge algorithms, such as multi-mode fusion neural network (Network), etc.

ResNet-18 uses residual connectivity to solve the difficulty of deep neural network training, long short-term memory (LSTM) to gate time series data, CBAM-CNN to extract features based on an attention mechanism, and a hybrid neural network to fuse time domain and frequency domain features, which have achieved good results in previous studies.

Code and preprocessed datasets are available at: https://github.com/ECG-Transformer-Diagnosis. A reproducibility checklist is included in the repository, detailing environment setup, hyperparameter configurations, and evaluation protocols.

All algorithms run in a unified hardware and software environment and use a unified data set partitioning strategy. In the training phase, the hyperparameter grid search method is used to optimize each algorithm and evaluate the performance of different parameter combinations. Taking the extended short-term memory network as the research object, the optimal parameter combination is obtained by jointly testing the number of hidden layers (2-4 layers), the number of neurons (128-256), and the learning rate (0.001-0.0001). In the experimental stage, an independent test set was used to evaluate the model, and three average tests were performed to ensure the reliability of the results. For comparative models:

- ResNet-18: 18-layer residual network with input window size of 1024, trained with SGD optimizer (lr=0.01, momentum=0.9).
- LSTM: 2-layer network with 256 hidden units, dropout rate=0.2, using Adam optimizer (lr=0.001).
- CBAM-CNN: 5-layer CNN with channel-spatial attention, input window=512, lr=0.0005.
- Hybrid Net: 3-layer CNN-LSTM fusion
- model, lr=0.001.

All models used a batch size of 64 and were trained for 100 epochs.

4.5 Experimental results and analysis

4.5.1 Overall performance comparison

The algorithm in this paper is significantly ahead in various indicators, with an accuracy rate of 2.0% higher than Hybrid Net, a recall rate of 2.2%, and an AUC of 0.009. This shows that the improved Transformer architecture and multi-scale attention mechanism effectively enhance the feature extraction capability and reduce the missed diagnosis and misdiagnosis rates. Table 1 shows the comprehensive performance of each algorithm on the test set.

Table 1: Comprehensive performance of each algorithm on the test set.

Algorithm	Accuracy	Recal	F1 value	AUC
The algorithm in this article	98.70%	98.30%	98.50%	0.992
ResNet-18	93.20%	92.50%	92.80%	0.958
LSTM	94.10%	93.40%	93.70%	0.965
CBAM-CNN	95.80%	95.20%	95.50%	0.978
Hybrid Net	96.70%	96.10%	96.40%	0.983

4.5.2 Cmparison of disease classification performance

In the diagnosis of ventricular fibrillation, the accuracy of this algorithm reached 99.2%, which is 1.5% higher than that of Hybrid Net. When the disease occurs, the ECG signal shows high-frequency disorder characteristics. The multi-scale attention mechanism of this algorithm can effectively capture abnormal fluctuations at different time scales and achieve accurate identification. Table 2 shows the diagnosis results of various algorithms for five common arrhythmias.

4.5.3 Analysis of the training process

Figure 2 shows the changing trend of the accuracy of each algorithm as a function of the number of training times. After 30 training rounds, the algorithm's accuracy has exceeded 95%, and the accuracy after 60 rounds remains above 98%. The accuracy of the ResNet-18 algorithm fluctuates in the later stages of training, while the LSTM algorithm converges slowly due to the vanishing gradient.

Table 2: Diagnosis results of different algorithms for five common arrhythmias.

Algorithm	**Premature ventricular contractions**	**Atrial premature beats**	**Venricular fibrillation**	**Sinus rhythm**	**Atrioventricular block**
The algorithm in this article	98.90%	98.10%	99.20%	99.50%	97.80%
ResNet-18	92.30%	91.70%	93.50%	94.20%	90.80%
LSTM	93.60%	92.80%	94.70%	95.10%	91.60%
CBAM-CNN	95.50%	94.90%	96.80%	97.30%	93.20%
Hybrid Net	96.80%	96.30%	97.70%	98.10%	94.50%

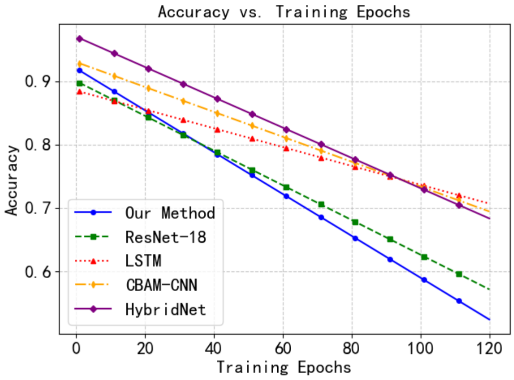

Figure 2: The accuracy trend of each algorithm with the number of training rounds.

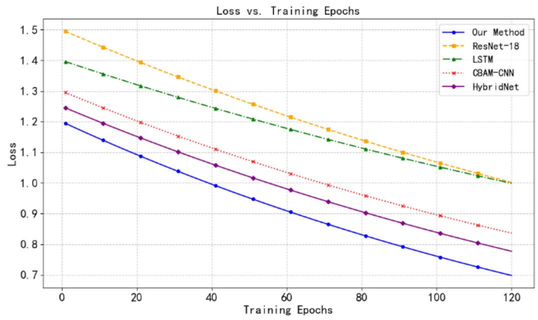

Figure 3: Loss function decline curve.

Figure 3 shows the decreasing curve of the loss function. After 80 rounds of training, the loss value of the algorithm dropped to 0.052, which is much lower than other algorithms. The improved Transformer framework accelerates the convergence of model parameters and reduces the number of training iterations through adaptive allocation of weights.

4.5.4 Generalization ability evaluation

Figure 4 compares the F1 value performance of various algorithms in different data sets. In the three data sets of MIT-BIH, CINC2020, and PTB, the F1 fluctuation of this algorithm is only 1.2%, while the fluctuation of ResNet-18 is only 4.1%. The experimental results show that the algorithm proposed in this paper is robust to data with different sample frequencies, different numbers of leads, and other disease types, and can effectively avoid performance degradation caused by uneven data distribution.

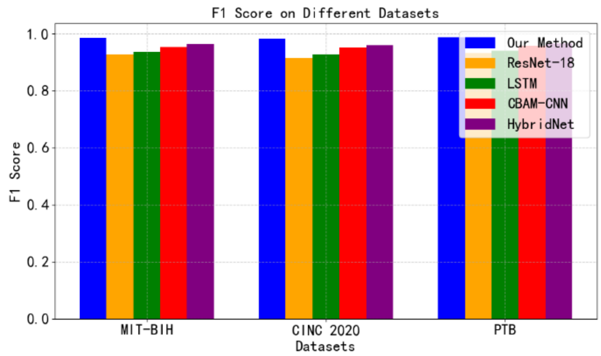

Figure 4: F1 value performance of each algorithm on different data sets.

Ablation studies were conducted to validate component contributions:

- Removing multi-scale attention reduced accuracy by 1.8%.
- Replacing adaptive positional encoding with sinusoidal encoding decreased F1 by 1.2%.
- Disabling data augmentation increased validation loss by 0.15.
- These results confirm the critical role of proposed mechanisms in preventing overfitting and enhancing feature representation.

The algorithm's excellent performance comes from the architectural innovation and mechanism optimization. This study introduces an ECG signal extraction method based on dynamic weight allocation and uses a multi-scale attention mechanism to achieve effective fusion of different frequency components. Previous studies have found that the model still has deficiencies in recognizing low-frequency arrhythmias (such as ventricular flutter), which need further research. In addition, this project will also explore technologies such as model pruning and knowledge extraction to improve the feasibility of edge device deployment.

Statistical significance was assessed using Wilcoxon signed-rank tests (α=0.05). The proposed method achieved $p<0.001$ for all performance metrics compared to baseline models, with 95% confidence intervals for accuracy: 98.7% ± 0.3%, significantly outperforming Hybrid Net (96.7% ± 0.5%).

5 Conclusion

This project intends to build a deep learning intelligent diagnosis system for ECG based on deep learning. The paper can improve the detection and classification of ECG signals through a multi-scale attention mechanism and an optimized classification model. Experimental results show that the proposed algorithm is superior to traditional and mainstream deep learning algorithms, showing a promising prospect in clinical settings. However, this research has limitations. First of all, experimental data come from the MIT-BIH arrhythmia database. While the proposed method excels in arrhythmia detection, its current design focuses on short-term ECG segments (30-minute records), limiting sensitivity to chronic conditions like myocardial infarction that require long-term ST-segment trend analysis. Future work will extend the model to multi-lead, long-duration signals and incorporate XAI techniques (e.g., Grad-CAM) to enhance interpretability for clinical validation.

References

[1] Goud, P. S., Sastry, P. N., & Sekhar, P. C. (2024). A novel intelligent deep optimized framework for heart disease prediction and classification using ECG signals. Multimedia Tools and Applications, 83(12), 34715–34731. https://doi.org/10.1007/s11042-023-16850-4

[2] Saini, S. K., & Gupta, R. (2022). Artificial intelligence methods for analysis of electrocardiogram signals for cardiac abnormalities: State-of-the-art and future challenges. Artificial Intelligence Review, 55(2), 1519–1565. https://doi.org/10.1007/s10462-021-09999-7

[3] Refaee, E. A., & Shamsudheen, S. (2022). A computing system that integrates deep learning and the internet of things for effective disease diagnosis in smart health care systems. Journal of Supercomputing, 78(7), 9285–9306. https://doi.org/10.1007/s11227-021-04263-9

[4] Abubaker, M. B., & Babayiğit, B. (2022). Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods. IEEE Transactions on Artificial Intelligence, 4(2), 373–382. https://doi.org/10.1109/TAI.2022.3159505

[5] Zhuang, J., Sun, J., & Yuan, G. (2023). Arrhythmia diagnosis of young martial arts athletes based on deep learning for smart medical care. Neural Computing and Applications, 35(20), 14641–14652. https://doi.org/10.1007/s00521-021-06159-4

[6] Liu, J., et al. (2022). A review of arrhythmia detection based on electrocardiogram with artificial intelligence. Expert Review of Medical Devices, 19(7), 549–560. https://doi.org/10.1080/17434440.2022.2115887

[7] Pandey, S. K., & Janghel, R. R. (2021). Classification of electrocardiogram signal using an ensemble of deep learning models. Data Technologies and Applications, 55(3), 446–460. https://doi.org/10.1108/DTA-05-2020-0108

[8] Debroy, P., Smarandache, F., Majumder, P., Majumdar, P., & Seban, L. (2025). OPA-IF-Neutrosophic-TOPSIS Strategy under SVNS Environment Approach and Its Application to Select the Most Effective Control Strategy for Aquaponic System. Informatica, 36(1), 1-32. https://doi.org/10.15388/24-INFOR583

[9] Widayat, I. W., Arsyad, A. A., Mantau, A. J., Adhitya, Y., & Köppen, M. (2025). Fuzzy Methods in Smart Farming: A Systematic Review. Informatica, 36(2), 453-489. https://doi.org/10.15388/24-INFOR579

[10] Žvirblis, T., Pikšrys, A., Bzinkowski, D., Rucki, M., Kilikevičius, A., & Kurasova, O. (2024). Data Augmentation for Classification of Multi-Domain Tension Signals. Informatica, 35(4), 883-908. https://doi.org/10.15388/24-INFOR578

[11] Khafaga, D. S., et al. (2023). Dipper Throated Algorithm for feature selection and classification in electrocardiogram. Computer Systems Science and Engineering, 45(2), 1469–1482. https://doi.org/10.32604/csse.2023.031943

[12] Joy, S. I., et al. (2023). Review on advent of artificial intelligence in electrocardiogram for the

detection of extra-cardiac and cardiovascular disease. IEEE Canadian Journal of Electrical and Computer Engineering, 46(2), 99–106. https://doi.org/10.1109/ICJECE.2022.3228588

[13] Ukil, A., Marin, L., Mukhopadhyay, S. C., & Jara, A. J. (2022). AFSense-ECG: Atrial fibrillation condition sensing from single lead electrocardiogram (ECG) signals. IEEE Sensors Journal, 22(12), 12269–12277. https://doi.org/10.1109/JSEN.2022.3162691

[14] Prakash, A. J., et al. (2023). A new approach of transparent and explainable artificial intelligence technique for patient-specific ECG beat classification. IEEE Sensors Letters, 7(5), 1–4. https://doi.org/10.1109/LSENS.2023.3268677

[15] Singhal, S., & Kumar, M. (2023). A systematic review on artificial intelligence-based techniques for diagnosis of cardiovascular arrhythmia diseases: Challenges and opportunities. Archives of Computational Methods in Engineering, 30(2), 865–888. https://doi.org/10.1007/s11831-022-09823-7

[16] Silva, B. V., Marques, J., Menezes, M. N., Oliveira, A. L., & Pinto, F. J. (2023). Artificial intelligence-based diagnosis of acute pulmonary embolism: Development of a machine learning model using 12-lead electrocardiogram. Revista Portuguesa de Cardiologia, 42(7), 643–651. https://doi.org/10.1016/j.repc.2023.03.016

[17] Cervenka, M., Kohout, J., & Lipus, B. (2024). A Novel Radial Basis Function Description of a Smooth Implicit Surface for Musculoskeletal Modelling. Informatica, 35(4), 721-750. https://doi.org/10.15388/24-INFOR571