Classification of ECG Signals Based on Hilbert-Huang Transform and 1D Convolution Neural Network

Tianyao Tang [1,a]

[1] Department of Applied Physics, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

[a] tty2004@nuaa.edu.cn

Abstract. Arrhythmia is a common phenomenon in cardiovascular diseases and its accurate diagnosis typically relies on a thorough analysis of electrocardiogram (ECG). Therefore, accurate identification and classification of ECG signals play a crucial role in the effective treatment of cardiac diseases. In this study, we propose a deep learning model based on Hilbert-Huang Transform (HHT) and one-dimensional convolutional neural network (1D-CNN), aiming to enhance the accuracy of ECG signal classification. Specifically, we first perform time-frequency analysis and feature extraction of the signal using the Hilbert-Huang Transform, after which we further extract features and classify the signal using 1D Convolutional Neural Network. The experimental results show that the model proposed in this paper performs well in classifying four types of ECG signals, with average classification accuracies of 97.73%, 99.16%, 99.50%, and 99.88%, respectively. This not only proves the effectiveness of our proposed method but also provides important technical support for the diagnosis and treatment of cardiovascular diseases, which has far-reaching clinical application value.

Keywords: Electrocardiogram Classification; Hilbert-Huang Transform; Convolutional Neural Network.

1. Introduction

Atrial fibrillation (AF), a highly prevalent and persistent arrhythmia, is one of the most common cardiovascular diseases. Early AF detection can enhance the effectiveness of clinical treatment and help prevent serious complications [1]. Electrocardiogram (ECG), first introduced by Muirhead in 1872, is a widely used non-invasive method that enables the analysis of ECG to diagnose clinical patients with AF and other types of arrhythmia [2]. Therefore, recognizing and classifying ECG signals is crucial for the treatment of cardiac diseases. It holds broad societal significance and high research value.

To enhance the accuracy of ECG signal recognition and classification, various automatic recognition methods are emerging [3]. These methods can be categorized into two main groups: those based on traditional machine learning algorithms and those based on neural networks and deep learning. Rabee and Barhumi's team [4] successfully classified 14 different heartbeat signals using support vector machines combined with discrete wavelet transform techniques, achieving a high accuracy rate. Celin [5] conducted a comparative study of various machine learning classifiers and found that the plain Bayes classifier performed well in terms of accuracy. [6] Employed an XGBoost classifier and a multi-stage processing technique, which included steps such as data acquisition, noise filtering, and feature extraction, for 45 feature descriptors. This approach significantly enhanced the classification accuracy of ECG signals. Mohebbanaaz et al. [7] conducted an early disease analysis using ECG and developed two classifiers: an optimized decision tree and an adaptive enhancement optimized decision tree. These classifiers effectively handle uncertain data and meet the need for high-precision classification.

In recent years, with the improvement of computational power, deep learning techniques have been widely used in ECG signal classification. Li Dan et al. [8] utilized a one-dimensional CNN to accurately classify five types of arrhythmias and achieved high accuracy on the MIT-BIH arrhythmia database. Mathur et al. [9] employed a CNN and a Feedforward Artificial Neural Network (FFANN) approach. Experiments demonstrated that the method performed well in terms of accuracy and efficiency. [10] utilized a deep Long Short-Term Memory (LSTM) network to address the ECG signal classification issue. The ECG time series were converted into spectral images, and key features such as instantaneous frequency and spectral entropy were extracted to train the LSTM network. This approach significantly enhanced the performance of the deep learning-based classifier. A new multi-task deep neural network was proposed by [11] to dynamically model the local and global information of ECG feature sequences. This model successfully performed ECG signal analysis on CPSC2018 and PTB-XL datasets, achieving average F1 scores of 0.827 and 0.833. Amin's team [12] utilized a two-dimensional CNN model to classify ECG signals. This model not only demonstrated high performance in terms of average classification accuracy but also exhibited significant advantages in evaluation metrics such as sensitivity and specificity.

Traditional machine learning algorithms have two main limitations: (1) They struggle to handle large-scale and diverse datasets, making it challenging to learn complex and high-dimensional features from vast amounts of data. This limitation hinders effective ground classification tasks. (2) For complex data such as ECG signals, traditional machine learning algorithms require extensive manual intervention for feature extraction and preprocessing. In contrast, deep learning models can automatically learn the features in the data, thus avoiding complex manual processing. In addition, deep learning is more efficient when dealing with large-scale datasets and is more likely to achieve better classification results.

Therefore, this paper proposes an ECG signal classification method that combines HHT and a one-dimensional CNN. The HHT is used to extract features in the frequency domain of the signal that is challenging to directly analyze with the model. Subsequently, the CNN leverages its robust learning capability for final signal classification. Our approach leverages the benefits of HHT in analyzing nonlinear and nonsmooth signals and the effectiveness of CNN in feature learning and pattern recognition. With this innovative combination, the accuracy and automation of ECG signal classification can be significantly improved.

2. Method

2.1 Hilbert-Huang Transform

The Hilbert-Huang Transform (HHT) is a powerful tool for signal feature extraction, providing distinct advantages over methods such as spectral mapping and wavelet analysis when handling nonlinear and nonsmooth signals [13]. The core algorithm of HHT involves Empirical Mode Decomposition (EMD) and the Hilbert Transform. Initially, complex signals are decomposed into multiple Intrinsic Mode Functions (IMFs) by EMD. Subsequently, features such as instantaneous frequency and instantaneous energy are extracted using the Hilbert transform. This process helps in understanding and classifying complex ECG phenomena to enhance classification accuracy. Next, I will introduce the principles of these two algorithms one by one.

2.1.1 Empirical Mode Decomposition (EMD)

EMD reveals the intrinsic properties of signals on different time scales by decomposing the dataset into a finite and typically small number of IMFs through adaptive time-frequency analysis. Among them, each IMF has a specific frequency and amplitude and possesses the following two explicit properties:

Across the entire dataset, the number of extreme points and the number of zero crossings are either equal or differ by at most one.

At any point, the average value of the envelope defined by the local maxima and the average value of the envelope defined by the local minima are zero.

The steps of EMD decomposition are as follows:

First, the local maxima and minima of the signal are identified. Applying the cubic spline fitting technique, the upper and lower envelopes of the signal are constructed, respectively. The average of these two envelopes is calculated to form a new curve m_1. The original signal $x(t)$ is then combined with the new curve m_1 to create a difference, i.e.:

$$h_1 = x(t) - m_1$$

Check h_1 whether the IMF conditions are satisfied. If not, the filtering process needs to continue. At this point, consider the h_1 as a new original signal, repeat the previous steps iteratively until h_k satisfies the IMF condition, and the first IMF is written as c_1. The first IMF is labeled as

$$c_1 = h_k$$

Subtract from the signal c_1 from the original signal to obtain the residual $r_1 = x(t) - c_1$. Repeat the process with the residual r_1 as the new original signal and repeat the above process, until the IMF cannot be extracted from the residuals, i.e.:

$$r_2 = r_1 - c_2, \cdots, r_n = r_{n-1} - c_n$$

Ultimately, the original signal $x(t)$ can be accurately represented as a series of IMFs with the upper final residuals r_n the sum of the final residuals on a series of IMFs:

$$x(t) = \sum_{i=1}^{n} c_i + r_n$$

Through the aforementioned steps, the EMD method can effectively decompose complex, non-smooth, and nonlinear signals into a series of IMFs reflecting the signal characteristics at various scales. This process enables a comprehensive analysis and understanding of the signals.

2.1.2 Hilbert transform

After EMD has successfully decomposed the signal into a series of IMFs, the Hilbert Transform can be performed on each IMF individually to further analyze the time-frequency characteristics of the signal, reveal the instantaneous frequency and amplitude of the signal, and gain insight into the nature of the signal. The process is as follows:

For a given IMF time series $c_i(t)$, performs a Hilbert transform to obtain the corresponding $H_i(t)$ that follows the mathematical expression below:

$$H_i(t) = \frac{1}{\pi} \int \frac{c_i(t')}{t - t'} dt'$$

This process generates a complex signal, called a resolved signal $z_i(t)$ which takes the original IMF $c_i(t)$ as the real part and its Hilbert transform result $H_i(t)$ as the imaginary part, in the specific form:

$$z_i(t) = c_i(t) + jH_i(t) = a(t)e^{j\theta(t)}$$

Here, the $a(t)$ represents the amplitude of the resolved signal and the $\theta(t)$ represents the phase angle. The amplitude and phase can be calculated by the following equations respectively:

$$a(t) = \sqrt{c_i^2(t) + H_i^2(t)}$$

$$\theta(t) = \arctan\left(\frac{H_i(t)}{c_i(t)}\right)$$

$a(t)$ provides instantaneous information about the signal strength, while the $\theta(t)$ the time derivative of the signal gives the instantaneous frequency $\omega(t)$:

$$\omega(t) = \frac{d\theta(t)}{dt}$$

After this, the resulting instantaneous frequency and amplitude can be combined to reconstruct the original signal and form a more comprehensive signal representation:

$$x(t) = Re \sum_{i=1}^{n} a_i(t) e^{j\int \omega_i(t) dt}$$

With the HHT algorithm, we can extract detailed information in the time-frequency domain from the original signal, which offers a powerful tool for analyzing and understanding ECG signals. With this approach, the dynamics of the signal on various time scales can be thoroughly explored, which is crucial for comprehending complex ECG signal systems and subsequent signal classification.

2.2 Convolution Neural Network

Convolutional neural network (CNN) is an efficient feature extraction tool [14]. Unlike 2D CNNs, which are mainly applied to high-dimensional tensors such as images, 1D CNNs are more adept at handling sequential data. They are very suitable for feature extraction of ECG signals in this paper, thanks to their core building blocks: convolutional layers and pooling layers. The orderly organization of these layers not only ensures that the model can extract features end-to-end but also provides a solid foundation for model training and optimization.

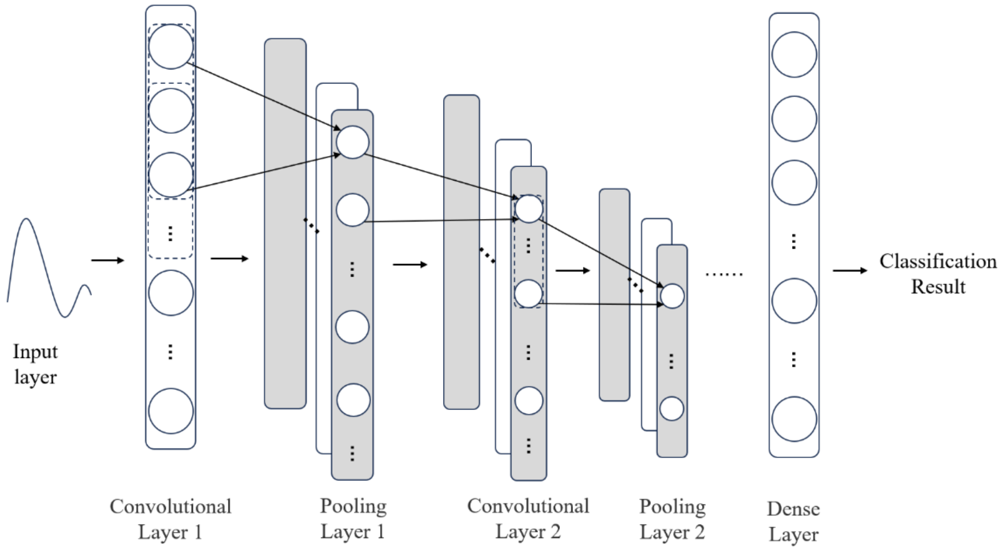


Figure 1 Structure of 1D-CNN

2.2.1 Convolutional layers

The convolutional layer in CNN achieves effective extraction of local features in the data by performing convolutional operations on local regions of the input data using multiple convolutional kernels. Specifically, the convolution operation can be described by the following equation:

$$X_j^k = f(\sum_{i \in M_j} w_{ij}^k * X_i^{k-1} + b_j^k)$$

$$f(z) = max(z, 0)$$

In the equation, X_j^k represents the output of the jth feature map in the kth layer. The function f signifies the application of a nonlinear activation function, with ReLU being the specific activation function used. M_j denotes the range of the local sensory field being considered. w_{ij}^k stands for the ith weight corresponding to the jth convolution kernel in the kth layer, and b_j^k represents the bias value associated with the feature map.

2.2.2 Pooling layer

The pooling layer's role is to conduct dimensionality reduction on the feature maps extracted from the convolutional layer. This process reduces model complexity and enhances processing performance. In the pooling operation, the two most common methods are mean pooling and maximum pooling. These methods achieve dimensionality reduction of features by calculating the mean or maximum value within the pooling window. In our experiments, we chose the maximum pooling method. The process can be expressed by the following equation:

$$y_{kij} = \max_{(p,q) \in R_{ij}} x_{kpq}$$

Where y_{kij} denotes the maximum output value associated with the kth feature map in the rectangular region R_{ij}, x_{kpq} denotes the element in R_{ij} located at (p, q), and $|R_{ij}|$ denotes the number of elements in R_{ij}.

3. Experiment

3.1 Experimental environment

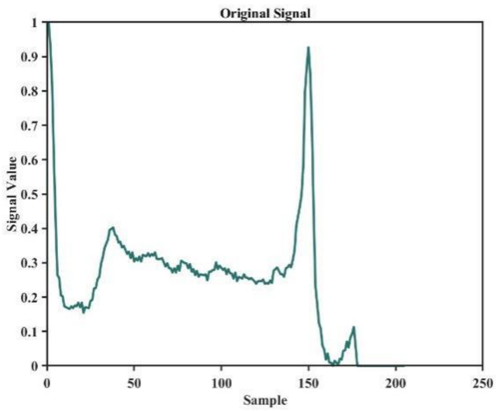
The experimental framework of this study was constructed on a computer platform equipped with an AMD Ryzen 7 5800H processor and an NVIDIA GeForce GTX 1650 graphics card. The data preprocessing and construction of the subsequent models were carried out using Python 3.11.2 language and the Keras deep learning framework. In addition, the experiments related to the HHT algorithm in this study were conducted in the MATLAB 2023b environment.

3.2 Dataset

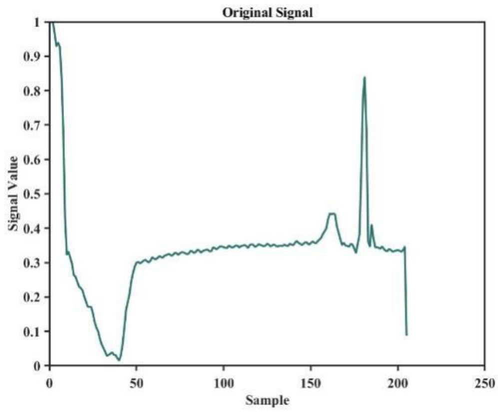
3.2.1 Introduction to the dataset

The dataset used for the study comes from the AliCloud Tianchi Learning Tournament[15]. The ECG data records provided cover a series of heartbeat signal sequences. The dataset classifies the signals into four categories: 0, 1, 2, and 3. The aim of this study is to predict the categories of heartbeat signals based on the various signal sequences.

After organizing the original dataset, two sets of data were selected for visualization, and the results are presented below:



Original Signal A



Original Signal B

Figure 2 Original Signal

It was noted that the waveforms obtained were similar to the waveform characteristics of a typical ECG. Given that ECG monitoring typically employs a sampling frequency of up to 200 samples per second, and that each data sample contains 205 data points, this is highly consistent with the characteristics of ECG sampling. On this basis, it is reasonable to assume that each data sample likely represents a one-second fragment of the ECG signal extracted from a specific lead.

3.2.2 Data preprocessing

First, upon examining the dataset, it was observed that the signal length was 205. However, certain signal segments had zero values at their ends. For instance, the signal depicted in the figure above (left) only contains 177 valid sampling points. This phenomenon may be due to the fact that the sampling process was interrupted prematurely before it lasted a full second, or that the electrode tabs were removed prematurely before the end of the sampling, thus causing data interference from non-heartbeat signals. Therefore, we converted the trailing zeros of these signals to NaN. Subsequently, we established a threshold to assess and discard/truncate the signals, thereby eliminating these non-representative data segments. Here we choose 128 as the threshold value.

In addition, a significant imbalance was observed in the distribution of samples in the original dataset. This imbalance may cause the model to overfit to categories with larger sample sizes during the training process, thereby reducing the predictive performance for certain categories. To solve this problem, we scale the data features randomly by generating an array of random numbers that follow a specific normal distribution as scaling factors and multiplying them with the original data features. This method can generate new samples with slight variations in the feature space while maintaining the original data distribution. This process increases the diversity of data, providing a more balanced and comprehensive database for model training. The following figure shows the sample distribution of the dataset before and after equalization:

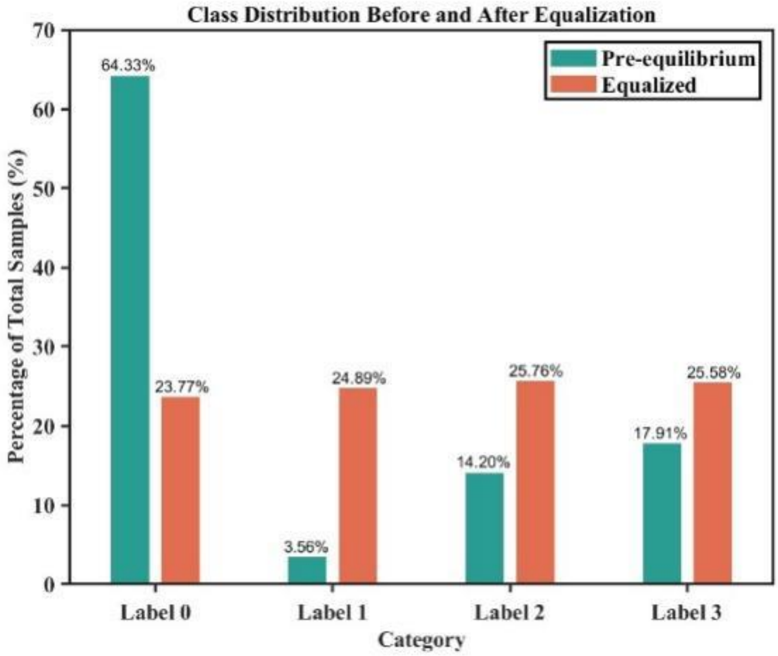


Figure 3 Class Distribution Before and After Equalization

After expanding the dataset, to reduce training costs, we randomly selected 10,000 signal samples from the expanded dataset as the base dataset for this experiment. Each signal sample was numbered, and all subsequent experiments and improvements were conducted based on this dataset.

3.3 Evaluation metrics

In this study, Accuracy, Precision, Recall and F1_Score are used as the core evaluation metrics to comprehensively assess the performance of the ECG signal classification model. These metrics comprehensively reflect the performance of the model in handling complex multi-classification tasks from various dimensions. They encompass both the overall accuracy of the model and the evaluation of the model's fine-grained discriminative ability. The definitions of each metric are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is a fundamental measure of the overall classification accuracy of the model. It indicates the proportion of correctly classified samples by the model, offering an intuitive basis for evaluating the model's performance on the entire dataset.

$$Precision = \frac{TP}{TP + FP}$$

Precision reflects the proportion of samples predicted to be in the positive category that are truly positive. High accuracy in ECG signal classification implies confidence in predicting abnormal rhythms and reduces the likelihood of misdiagnosis.

$$Recall = \frac{TP}{TP + FN}$$

Recall measures the proportion of all true positive class samples that are correctly identified by the model. High recall ensures that critical heart rate abnormalities are not missed, which is valuable for patient care.

$$F1\_Score = \frac{2 \times (Recall \times Precision)}{Recall + Precision}$$

The F1 score is the harmonized average of precision and recall, serving as a more comprehensive performance evaluation metric that balances the importance of precision and recall to offer a thorough assessment of model performance.

In the above equation, true positive cases (TP) refer to the number of samples correctly identified as abnormal rhythms by the model; true negative cases (TN) refer to the number of samples correctly identified as normal rhythms; false positive cases (FP) represent those normal samples incorrectly determined to be abnormal rhythms; and false negative cases (FN) are those samples that are actually abnormal rhythms but are incorrectly determined to be normal by the model.

3.4 Comparative experiments

The experiments were initially conducted to classify the ECG signals by using the signal values directly as features. XGBoost, Random Forest, SVM, and 1DCNN were selected as the classification models. The core parameters of each model were chosen as indicated in the following table:

Model	Parameter Name	Parameter Value
XGBoost	n_estimators	100
	learning_rate	0.1
	max_depth	6
RandomForest	n_estimators	100
	max_features	'sqrt'
	criterion	'entropy
SVM	C	1.0
	kernel	'rbf'
	gamma	'scale'
1DCNN	filters	256
	kernel_size	3
	epochs	150
	batch_size	256
	activation	'relu'

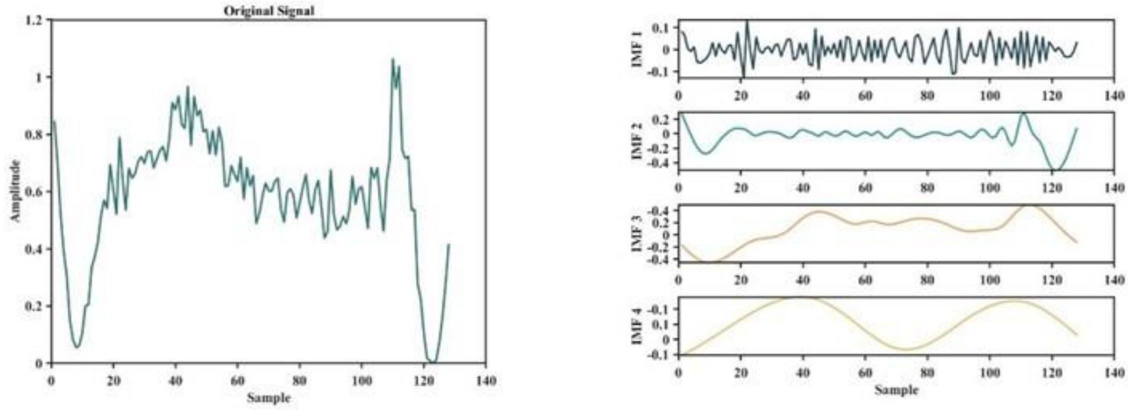
For each group of models in the table, we utilized a five-fold cross-validation method to acquire the experimental results as follows:

Model	Accuracy	Precision	Recall	F1_Socre
XGBoost	95.15%	95.10%	95.10%	95.08%
Random Forest	95.41%	95.37%	95.37%	95.35%
SVM	93.75%	93.70%	93.67%	93.66%
1DCNN	96.73%	96.71%	96.67%	96.67%

It can be seen that 1DCNN shows a relative advantage in ECG signal classification tasks. Therefore, we use 1DCNN as our fundamental model in the following modules.

3.5 HHT Experiment

After obtaining the optimal 1DCNN model in the previous step, we utilize the HHT method to extract additional frequency domain features from the signal. First, the signal is decomposed using the EMD method, resulting in a series of IMFs can be obtained for each signal. For instance, considering signal 2023 from the experimental dataset, the figure below illustrates the original signal and the IMFs obtained through its EMD decomposition:



Original Signal IMFs

Figure 4 EMD descomposition

After EMD decomposition of the signals, the obtained IMFs are then processed using the Hilbert transform to derive the respective sequences of instantaneous frequency and instantaneous energy for each signal.

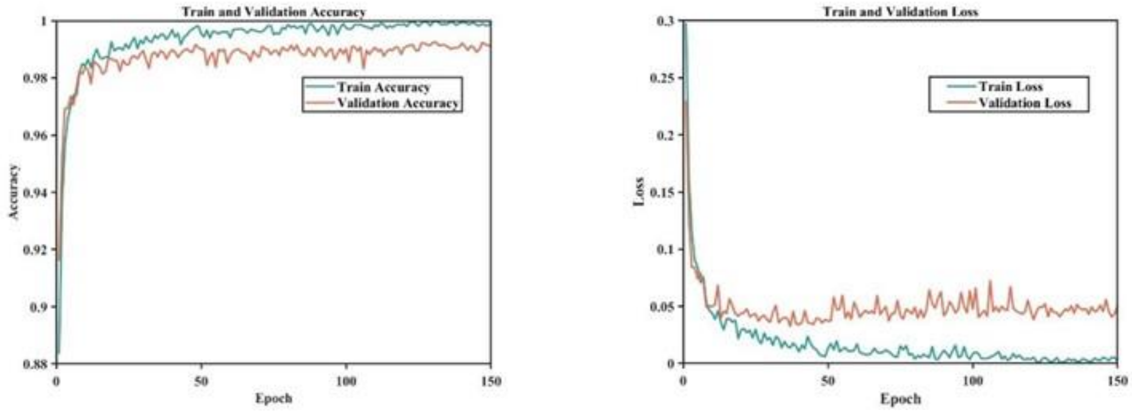
Different instantaneous frequencies and instantaneous energies are selected as features, respectively, and combined for experiments. It is noted that features involving frequency (mean_nu) result in a significant reduction in accuracy and other indicators. Therefore, only instantaneous energies are used as features thereafter. The energy of the first IMF (first_E), the energy of the last IMF (last_E), and the mean value of the energy of all IMFs (mean_E) are used as the features, and the results are obtained as follows.

Model	Accuracy	Precision	Recall	F1_Socre
mean_nu	85.08%	84.85%	84.87%	84.80%
mean_E	99.09%	99.09%	99.08%	99.08%
first_E	99.02%	99.01%	99.01%	99.01%
last_E	99.00%	98.99%	98.99%	98.99%

From the table above, it can be seen that the classification is best characterized by the mean energy of all IMFs.

3.6 Analysis of results

Analyzing the "mean_E" feature set, the training set and test set loss function values change with the number of training iterations, as illustrated in the figure below:



Accuracy Curve Loss Curve

Figure 5 Accuracy and Loss Curve

As the training progresses, the accuracy on the validation set increases and converges to around 99% after approximately 50 iterations. At this stage, the network is capable of effectively learning the features of different types of ECG signals. The classification confusion matrix for the test set is displayed below:

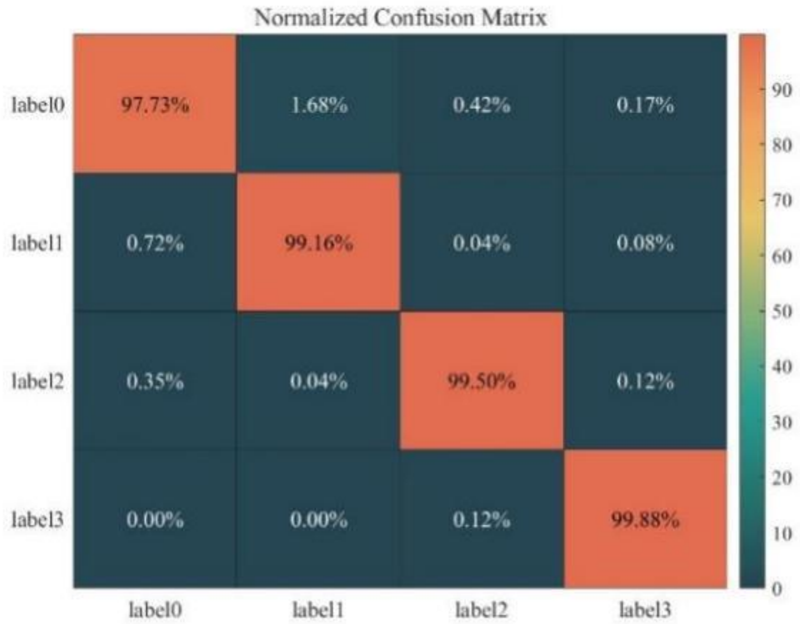


Figure 6 Normalized Confusion Matrix

From the figure, it can be clearly concluded that the classification accuracy of ECG signals in categories 1, 2, and 3 reaches 99.16%, 99.50%, and 99.88%, respectively. These values are close to 100%, indicating that the model can categorize these three categories of ECG signals almost without error. The accuracy of category 0 is slightly lower, but still reaches 97.73%.

In summary, it can be concluded that the HHT+1DCNN-based model proposed in this paper can effectively learn the features of various classes of ECG signals and performs well across all performance metrics for ECG signal classification.

4. Conclusion

In this study, we propose a deep learning model that combines the Hilbert-Huang transform (HHT) and a 1D convolutional neural network (1D-CNN) using ECG data provided by AliCloud Tianchi Learning Tournament as a dataset. This approach effectively enhances the accuracy of ECG signal classification. In the experimental process, we first compare the classification performance of four different classification models: XGBoost, Random Forest, SVM, and 1D-CNN, when using signal values as features directly. The results show that the 1D-CNN model outperforms the other models in all evaluation metrics (accuracy, precision, recall, and F1 value). Therefore, we have selected the 1D-CNN as our benchmark model. Additionally, we process the ECG signals using HHT to extract more comprehensive frequency domain features. The experimental results show that all four evaluation metrics exceed 99% when the mean energy (mean_E) of all IMFs is used as the feature combination. The average accuracies of classifying the four types of ECG signals are 97.73%, 99.16%, 99.50%, and 99.88%, respectively. This demonstrates the progress and effectiveness of our proposed method. However, there are still some shortcomings in this study. For instance, the model's generalization ability is not strong enough, and there is a lack of diversity in the combination of feature selection. Future studies can aim to explore more diverse datasets with varied data distributions and utilize a wider range of feature combinations for experiments to enhance the accuracy of ECG signal classification. This improvement can offer more robust technical support for the diagnosis and treatment of cardiovascular diseases.

References

[1] Lau D H, Linz D, Schotten U, et al. Pathophysiology of paroxysmal and pers istent atrial fibrillation: rotors, foci and fibrosis[J]. Heart, Lung and Circulation, 2017, 26(9): 887-893.

[2] **ong Z, Stiles M K, Zhao J. Robust ECG signal classification for detection of atrial fibrillation using a novel neural network[C]//2017 Computing in Cardiology (CinC). IEEE, 2017: 1-4.

[3] Clifford G D, Liu C, Moody B, et al. AF classification from a short single lead ECG recording: the PhysioNet/computing in cardiology challenge 2017[C]// 2017 Computing in Cardiology (CinC). IEEE, 2017: 1-4.

[4] Rabee A, Barhumi I. ECG signal classification using support vector machine based on wavelet multiresolution analysis[C]//2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA). IEEE, 2012: 1319-1323.

[5] Celin S, Vasanth K. ECG signal classification using various machine learning techniques[J]. Journal of medical systems, 2018, 42(12): 241.

[6] Pandeyz S K, Janghel R R, Gupta V. Cardiac arrhythmia detection and classification from ECG signals using XGBoost classifier[J]. Machine Learning Algorithms and Applications, 2021: 141-157.

[7] Mohebbanaaz, Kumari L V R, Sai Y P. Classification of ECG beats using optimized decision tree and adaptive boosted optimized decision tree[J]. Signal, Image and Video Processing, 2022, 16(3): 695-703.

[8] Li D, Zhang J, Zhang Q, et al. Classification of ECG signals based on 1D convolution neural network[C]//2017 IEEE 19th International Conference on e- Health Networking, Applications and Services (Healthcom). IEEE, 2017: 1-6.

[9] Mathur P, Sharma T, Veer K. Analysis of CNN and feed-forward ANN model for the evaluation of ECG signal[J]. Current Signal Transduction Therapy, 2023, 18(1): 1-8.

[10] Kłosowski G, Rymarczyk T, Wójcik D, et al. The use of time-frequency moments as inputs of lstm network for ecg signal classification[J]. Electronics, 2020, 9(9): 1452.

[11] Geng Q, Liu H, Gao T, et al. An ECG Classification Method Based on Multi-Task Learning and CoT Attention Mechanism[C]//Healthcare. mdpi, 2023, 11(7). 1000.

[12] Ullah A, Anwar S M, Bilal M, et al. Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation[J]. Remote Sensing, 2020, 12(10): 1685.

[13] Soorma N, Singh J, Tiwari M. Feature extraction of ECG signal using HHT algorithm[J]. arXiv preprint arXiv:1403.1660, 2014.

[14] Irungu J, Oladunni T, Denis M, et al. A CNN Transfer Learning-Electrocardiogram (ECG) Signal Approach to Predict COVID-19[C]//2023 15th International Conference on Computer and Automation Engineering (ICCAE). IEEE, 2023: 367-371.

[15] https://tianchi.aliyun.com/competition/entrance/531883/information