



OPEN ACCESS

EDITED BY

Michael Spartalis,
National and Kapodistrian University of Athens,
Greece

REVIEWED BY

Rishi Raj Sharma,
Defence Institute of Advanced Technology,
India
Gokhan Altan,
Iskenderun Technical University, Türkiye

*CORRESPONDENCE

Yang Li
✉ iyang15846601685@163.com
Bing Wang
✉ 10490842@qq.com

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 14 March 2024

ACCEPTED 20 May 2024

PUBLISHED 07 June 2024

CITATION

Chen S, Wang H, Zhang H, Peng C, Li Y and Wang B (2024) A novel method of swin transformer with time-frequency characteristics for ECG-based arrhythmia detection.
Front. Cardiovasc. Med. 11:1401143.
doi: 10.3389/fcvm.2024.1401143

COPYRIGHT

© 2024 Chen, Wang, Zhang, Peng, Li and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A novel method of swin transformer with time-frequency characteristics for ECG-based arrhythmia detection

Siyuan Chen^{1†}, Hao Wang^{2†}, Huijie Zhang^{3†}, Cailiang Peng³, Yang Li^{1,3*} and Bing Wang^{1*}

¹Heilongjiang University of Chinese Medicine, Harbin, China, ²College of Computer Science and Technology, Harbin Engineering University, Harbin, China, ³First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin, China

Introduction: Arrhythmia is an important indication of underlying cardiovascular diseases (CVD) and is prevalent worldwide. Accurate diagnosis of arrhythmia is crucial for timely and effective treatment. Electrocardiogram (ECG) plays a key role in the diagnosis of arrhythmia. With the continuous development of deep learning and machine learning processes in the clinical field, ECG processing algorithms have significantly advanced the field with timely and accurate diagnosis of arrhythmia.

Methods: In this study, we combined the wavelet time-frequency maps with the novel Swin Transformer deep learning model for the automatic detection of cardiac arrhythmias. In specific practice, we used the MIT-BIH arrhythmia dataset, and to improve the signal quality, we removed the high-frequency noise, artifacts, electromyographic noise and respiratory motion effects in the ECG signals by the wavelet thresholding method; we used the complex Morlet wavelet for the feature extraction, and plotted wavelet time-frequency maps to visualise the time-frequency information of the ECG; we introduced the Swin Transformer model for classification and achieve high classification accuracy of ECG signals through hierarchical construction and self attention mechanism, and combines windowed multi-head self-attention (W-MSA) and shifted window-based multi-head self-attention (SW-MSA) to comprehensively utilise the local and global information.

Results: To enhance the confidence of the experimental results, we evaluated the performance using intra-patient and inter-patient paradigm analyses, and the model classification accuracies reached 99.34% and 98.37%, respectively, which are better than the currently available detection methods.

Discussion: The results reveal that our proposed method is superior to currently available methods for detecting arrhythmia ECG. This provides a new idea for ECG based arrhythmia diagnosis.

KEYWORDS

electrocardiogram, deep learning, arrhythmia, wavelet time-frequency map, swin transformer

1 Introduction

Currently, cardiovascular diseases (CVD) exhibit the highest morbidity and mortality rates worldwide, posing a serious threat to human health (1). According to the World Health Organization (WHO), CVD-associated deaths account for approximately 32% of the total number of deaths each year (2). Arrhythmia, a phenomenon that causes the

heart to beat irregularly due to abnormal functioning of the heart's electrical system, is one of the major manifestations of underlying CVD (3). In many cases, arrhythmia patients often remain asymptomatic and later lead to diseases that cannot be easily diagnosed, resulting in serious consequences, for example, heart failure, stroke, and even sudden cardiac death (4, 5). Therefore, accurate and rapid detection of arrhythmia is particularly important for better treatment outcomes and long-term survival of the patient. Therefore, detecting arrhythmia at an early stage can minimize the chances of life-threatening situations in the future.

Currently, there are several diagnostic methods for arrhythmia, such as electrocardiogram (ECG), cardiovascular magnetic resonance imaging (MRI), and cardiac computed tomography (CT) (6–8). As an objective indicator of the occurrence, propagation, and recovery process of cardiac excitation, ECG plays a crucial role in the diagnosis of heart diseases. Moreover, due to the non-invasive procedure and low-cost advantages, ECG is most commonly used for detecting arrhythmia in clinical practice.

ECG data are usually affected by multiple factors, and data preprocessing can improve the data quality. Noise reduction is widely used as a common method for preprocessing. Sharma et al. (9) proposed noise reduction based on eigenvalue decomposition of Hankel matrix, which achieved better performance. Zhang et al. (10) proposed noise reduction of ECG signals by using *coif3* wavelet and trap filter, which improves the accuracy of extracted ECG parameters by suppressing the noise of P-wave and T-wave. Among many noise reduction methods, wavelet thresholding is suitable for various types of signal noise reduction, including biomedical signals, images, audio, etc., so it has a wide range of application prospects (11–13). Therefore, in this paper, wavelet thresholding method is used to achieve ECG noise reduction.

Feature extraction is an important component of machine-learning-based ECG diagnosis. Currently, feature extraction processes have been optimized based on the time and frequency domain aspects of the ECG (14, 15). Time domain feature extraction mainly captures the dynamic characteristics of the signal with respect to time but this feature has limitations in dealing with non-stationary signals. Frequency domain feature extraction can reveal the frequency of components and spectral features of the ECG, but transient information may be omitted when extracting the signal. In contrast, time-frequency domain feature extraction combines both time and frequency domain information, making the signal analysis more comprehensive by capturing the dynamic changes and frequency characteristics of the ECG. Currently, ECG feature extraction based on time-frequency features has become more and more popular, Qurraie et al. (16) extracted time-frequency features and statistical features of ECG signals along the RR intervals for arrhythmia classification, and Sharma et al. (17) proposed time-frequency matrix-based modified features for detecting coronary artery disease (CAD), which have all achieved good results. Wavelet time-frequency maps can visualise the time-frequency characteristics of the signal and have significant advantages in feature extraction (18). Therefore, this study primarily focuses on the time-frequency domain and adopts the visualized wavelet time-frequency diagram to represent ECG features. This

combined approach can effectively extract almost all ECG time-frequency domain features and provide a guarantee for the subsequent accurate classification work.

With the continuous development of computer-aided technologies in the medical field, the number of research works on the ECG detection method is gradually increasing. Diker et al. (19) used the Pan-Tompkins algorithm and the discrete wavelet transform (DWT) to extract the key points of ECG signals for ECG classification, and improved the wavelet kernel limit learning machine to determine the wavelet coefficients. Shirin et al. (20) used electrocardiogram data from three different databases, combined with temporal and spectral analyses and nonlinear dynamics, and were able to efficiently distinguish between ventricular fibrillation (VF) and non-VF arrhythmias and applied B bifurcated decision tree (BDT) and support vector machine (SVM) classifiers for arrhythmia classification. These methods have not only improved the accuracy of cardiovascular disease diagnosis, but also reduced the time required for diagnosis. In recent years, deep learning methods have achieved better results in several fields, including biosignal analysis (21), face recognition (22), computer vision (23), and character recognition (24). In ECG detection, the application of deep learning methods has also become substantially widespread (25–27). Currently, traditional networks such as convolutional neural networks (CNN) and recurrent neural networks (RNN) are most frequently used for arrhythmia diagnosis (28, 29). However, these techniques have several limitations in facing long-range dependencies and dealing with global information capture. The new network architecture of the emerging Transformer (30) shows certain advantages over previous models in the field of deep learning. The Transformer model has been developed considering the weightage of each position over the others through a self-attention mechanism, and positional coding to convey information about the sequence structure. In addition, jump connections between the outputs and inputs of each sublayer have been added to this model. Compared with the traditional model, the Transformer can effectively solve more complex problems involving long-range dependencies and global information capture. Simultaneously, the Transformer can perform parallel computation by calculating the number of dependencies between different positions in the input sequence, which improves the training speed of the model. Developed based on the Transformer model, the Swin Transformer offers specific improvements for computer vision to increase the efficiency of image-specific processing, thus making the Transformer architecture more compatible and efficient for a wide range of large-scale computer vision tasks (31). Therefore, we employed the Swin Transformer model to classify the ECG features.

In this study, we first pre-processed the ECG data from the MIT-BIH arrhythmia database to improve the data quality, then extracted the time-frequency features of the ECG by wavelet time-frequency mapping. Swin Transformer model was used to classify various types of arrhythmia, which in turn enabled the effective detection of cardiac arrhythmia. We found that the Swin Transformer model was effective in prompting early

warning and auto-diagnosis of arrhythmia, which provided a new avenue for arrhythmia detection.

2 Materials and methods

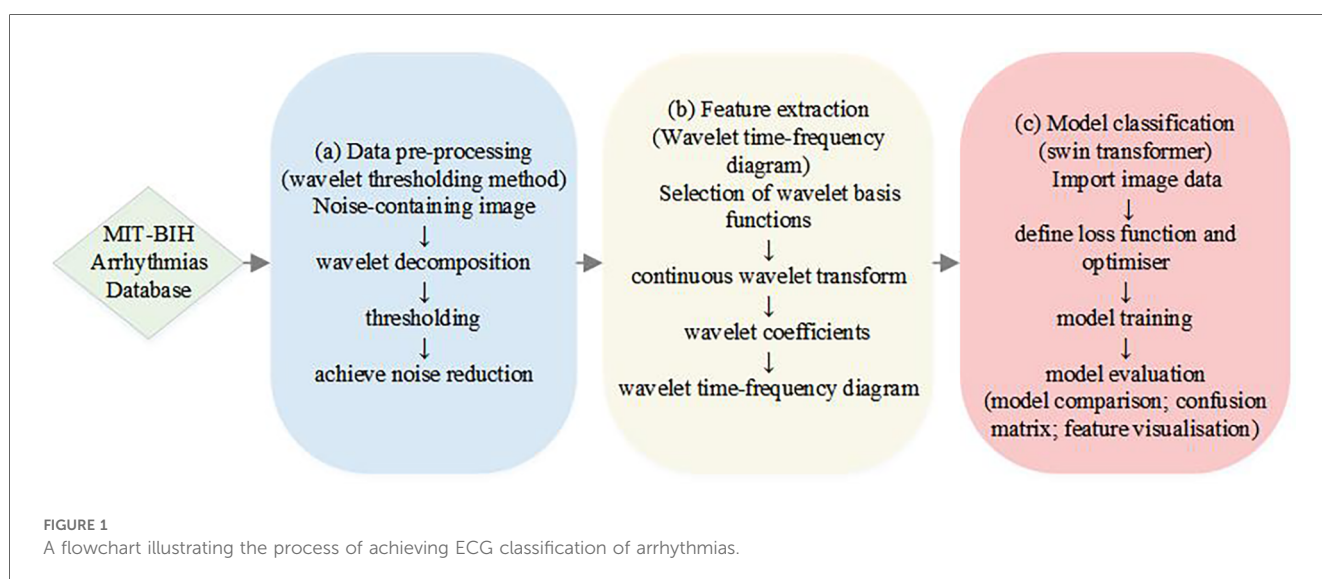
Arrhythmia classification has always been an important issue in the field of medicine, which is significant for diagnosis and treatment. With our proposed method, we are able to identify different types of arrhythmias more accurately and provide doctors with a more reliable auxiliary diagnostic tool, which further improves the efficiency and accuracy of patient diagnosis and treatment. In this study, we propose an ECG signal processing and classification method based on the wavelet threshold method and the Swin Transformer model, which has the potential for accurate ECG data analysis and classification in clinical applications. The overall flowchart of the experiments in this study is shown in [Figure 1](#), which consists of three main parts. The first part is the preprocessing of the data, which introduces the new application of wavelet thresholding method in ECG data denoising to remove the effects of high frequency noise, artefacts, electromyographic noise and respiratory motion produced on ECG. We perform wavelet transform on the original ECG signal, which makes the signal decomposed into wavelet components with different scales and frequencies, and choose the soft thresholding method for thresholding, which improves the quality of ECG. The second part is to complete the feature extraction of ECG data, we adopt the complex Morlet wavelet as the wavelet basis function, and then the continuous wavelet transform (CWT) is used to capture the feature information of the signal in time and frequency simultaneously, and the wavelet time-frequency diagram is drawn to show the ECG features, so as to achieve a more intuitive presentation of the time-frequency information of ECG, and to improve the accuracy and reliability of the analysis of ECG signals. The third part is to introduce the Swin Transformer model for ECG classification, which uses a hierarchical construction method similar to CNN,

reduces the computational complexity by calculating the self-attention within the window, and achieves window-to-window information transfer by moving window, which has a good accuracy in clinical diagnosis. We input the wavelet time-frequency maps into the Patch Partition module for chunking, sampling through the Patch Merging layer, and then construct feature maps of different sizes through four stages, and in order to combine local and global information, we use windowed multi-head self-attention (W-MSA) and shifted window-based multi-head self-attention (SW-MSA) in pairs. For each sample, the features are standardized, the mean and variance of the features are calculated, and the feature map is pooled along the spatial dimension. The globally pooled feature vector is used as input, and the output of the last fully connected layer is used as the final output of the model. Cross entropy loss is used as the loss function, and the Adam optimizer is selected for model optimization to achieve signal classification. This model has good accuracy in clinical diagnosis. Finally, model evaluation is achieved through model comparison confusion matrix and feature visualization.

2.1 Dataset

The MIT-BIH Arrhythmia database, created in collaboration between the Massachusetts Institute of Technology (MIT) and Beth Israel Hospital (BIH), is one of the most widely used resources in the field of arrhythmia detection and classification (32). The dataset covers many types of arrhythmias, including supraventricular and ventricular premature beats, atrial fibrillation, and atrial flutter.

In this study, we used the MIT-BIH Arrhythmia database for the classification of arrhythmias. This dataset included 48 ECG recordings from 47 volunteers, each of which lasted 30 min and contained more than 116,000 heartbeats, consisting of two signal channels V and II, and recorded at a sampling rate of 360 Hz and 11-bit resolution with a range of 10 mV (33–35).



In this experiment, we randomly divided the MIT-BIH Arrhythmia database into ten subsets, using nine folds of data as the training set and the remaining one fold as the validation set. The process was repeated ten times, choosing a different validation fold each time. During each validation process, the model is trained using the currently selected training set and the model performance is evaluated on the corresponding validation set. The accuracy, precision, sensitivity, specificity, F1 score and AUC values were recorded for each validation. Afterwards, the above metrics of each of the ten validations are averaged to obtain the final model evaluation results.

2.2 Pre-processing

Usually, the ECG recording includes artefactual contaminations, which incorrectly extract the local waveform of the ECG. Therefore, performing pre-processing is an essential step toward a precise and automatic ECG classification. Pre-processing of ECG data includes de-baseline drifting, filtering, noise reduction, and heartbeat detection, aiming to improve the quality of ECG signals (36). Among them, noise reduction is the most important step in experimental pre-processing. In this work, we used the wavelet thresholding method (37) to denoise the raw data, in order to remove the effects of high-frequency noise, artifacts, electromyographic noise, and respiratory movements on ECG. First, the db6 wavelet was used to decompose the signal into levels 1–3, and subsequently, decomposed signals were adjusted to the baseline using the rigrsure soft threshold selection method. Compared with the hard thresholding method, the soft thresholding method improves the defect of discontinuity at the threshold point, so smoother data are obtained in the soft thresholding process. The threshold processing formula of the soft thresholding method is expressed as Equation (1):

$$\hat{w} = \begin{cases} [\text{sgn}(w)](|w| - T), & |w| \geq T \\ 0, & |w| < T \end{cases} \quad (1)$$

where w is the decomposed wavelet coefficient, \hat{w} is the thresholded wavelet coefficient, and T denotes the threshold function.

For healthcare professionals to accurately identify and understand the ECG features, the Association for the Advancement of Medical Instruments (AAMI) has classified the MIT-BIH arrhythmia database into five AAMI heartbeat categories: N for normal beats, S for supraventricular abnormalities, V for ventricular abnormalities, F for fusion beats, and Q for unclassified beats (38, 39).

2.3 ECG feature extraction

The wavelet time-frequency diagram is one of the important methods to extract the time-frequency features of ECG, which converts one-dimensional signals into a two-dimensional time-frequency image through the CWT. Thus, the wavelet time-

frequency diagram could be helpful to further understand the essential properties of ECG, and enhance the analysis and diagnosis of diseases (40, 41).

Here, we used the complex Morlet wavelet for the wavelet analysis, which is expressed as Equation (2):

$$\psi(t) = (\pi F_b)^{0.5} e^{2i\pi F_c t} e^{-t^2/F_b} \quad (2)$$

where F_b denotes the bandwidth factor.

The wavelet coefficient of different scales and frequency bands of wavelet components that have been decomposed in the pre-processing step are obtained by wavelet transformation that gives different signal resolutions at different time-frequency characteristics, and the formula for the CWT is expressed as Equation (3):

$$U(\alpha, \beta) = \int_{-\infty}^{+\infty} x(t) \overline{\psi(t)} dt = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{+\infty} x(t) \overline{\psi\left(\frac{t-\beta}{\alpha}\right)} dt \quad (3)$$

where $U(\alpha, \beta)$ denotes the coefficient of the wavelet function; α and β denotes the scaling and translation factors; $x(t)$ denotes the original signal; $\psi(t)$ indicates the wavelet basis function; and $\overline{\psi(t)}$ refers to the conjugate complex of $\psi(t)$. These wavelets are generated by the mother wavelet ψ scaling and translation.

Following are the steps for calculating the time-frequency diagram of the wavelet components:

The actual frequency F_a corresponding to the scaling factor is denoted as Equation (4):

$$F_a = F_c \cdot \frac{f_s}{\alpha} \quad (4)$$

where F_c denotes the wavelet center frequency factor in Hz, and f_s denotes the sampling frequency.

The scale series t is expressed in the following form so that the changed frequency series is presented as an isotropic series A :

$$A = \{c/\text{totalscal}, \dots, c/(\text{totalscal} - 1), c/4, c/2, c\} \quad (5)$$

$$c = 2F_c \cdot \text{totalscal} \quad (6)$$

where the length of the scale sequence totalscal is set to 256, and c is a constant.

Substituting the expression of Equation (6) into Equation (5), we can obtain the desired scale sequence. Through the scale sequence and wavelet basis, the wavelet coefficient matrix can be derived from Equation (3), and the wavelet time-frequency diagram of the original vibration signal can be constructed by combining the time series and the actual frequency series.

2.4 ECG classification model

The Swin Transformer network model is a deep learning model proposed by Microsoft Research Asia in 2021. This model uses a

hierarchical structure to extract features by calculating attention within each window (31). In this study, we applied the Swin Transformer network model to segment an image with an input of $h \times w \times 3$ RGB into non-overlapping equal-sized $N \times (4 \times 4 \times 3)$ image blocks, where h and w were the height and width of the input image, respectively, and N was the effective length of the input sequence for the 31model.

When the linear embedding layer projects a tensor with feature dimension $h/4 \times w/4 \times 48$ to any dimension C , the feature dimension at this point becomes $h/4 \times w/4 \times C$. Hence, we first performed layer normalization (LN) by passing the image block sequence into two consecutive Swin Transformer blocks, and then conducted W-MSA analysis. The W-MSA divides the input sequence into different windows based on MSA to reduce the complexity of the calculation. Attention is calculated by extracting the correlation between local patches in the window. The basic formula of MSA can be expressed as Equation (7):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

Where Q , K and V denote the Query, Key and Value matrices, respectively; softmax denotes the weight assigned to all the Keys of each Query; d_k denotes the dimension of each element in the K matrix and applies weights to the Value matrix.

Assuming that the size of each local block is $m \cdot n$ and the dimension of the transformed matrix is $Z \cdot Z$, the computation of MSA can be done by Equation (8):

$$\Omega(\text{MSA}) = 4mnZ^2 + 2(mn)^2Z \quad (8)$$

Assuming that a graph is divided into H^*H local blocks, the computation of W-MSA can be estimated as Equation (9):

$$\Omega(\text{WMSA}) = 4mnZ^2 + 2H^2mn^2Z \quad (9)$$

Next, LN and multilayer perception (MLP) are performed to train the model with features at a deeper level, and this is the end of the first module. Subsequently, SW-MSA based on moving windows is carried out to realize mutual communication between windows, where W-MSA and SW-MSA are used in pairs, and their module calculations can be performed using Equation (10):

$$\begin{aligned} \hat{F}^l &= \text{WMSA}(\text{LN}(F^{l-1})) + F^{l-1} \\ F^l &= \text{MLP}(\text{LN}(\hat{F}^l)) + \hat{F}^l \\ \hat{F}^{l+1} &= \text{SW-MSA}(\text{LN}(F^l)) + F^l \\ F^{l+1} &= \text{MLP}(\text{LN}(\hat{F}^{l+1})) + \hat{F}^{l+1} \end{aligned} \quad (10)$$

where \hat{F}^l and F^l denote the output features of the (S)W-MSA module and the MLP module, respectively.

Finally, the output result is obtained by MLP, and the number of output sequences is the same as that of the input. In the second

stage, the adjacent 2×2 image blocks are stitched together by merging layers, and the output of the feature dimension after feature conversion is $h/8 \times w/8 \times 2C$. The process is repeated, and the eigendimensions of stages 3 and 4, respectively, give output as $h/16 \times w/16 \times 4C$ and $h/32 \times w/32 \times 8C$.

After completing the execution of the Swin Transformer block, the final ECG classification results are generated through normalization, global pooling, and full connection.

3 Results

3.1 Eigenvalue analysis

Here, we exploited wavelet time-frequency maps to transform one-dimensional ECG into two-dimensional images with time-frequency features, thus presenting ECG results in a better way, where the warm and cold colors of the wavelet time-frequency maps indicated the wavelet energy values of the signals in the time-frequency domain. The warm color represented the high-energy regions, while the cold color referred to the low-energy regions, and the horizontal and vertical axes of the images indicated the time and frequency, respectively. Figure 2 shows the raw signal maps of the five types of arrhythmias and the wavelet time-frequency maps presented in the time-frequency domain.

It can be observed that there are differences between the wavelet time-frequency maps corresponding to different types of cardiac beats. Different colors present an irregular block-like distribution, with the warm color area of the N-shaped heart beat being significantly larger than other types, and having the largest range and higher energy. The cold color area of Q-shaped heart beats is significantly more pronounced than other types. Compared to the S-shaped heart beat, the V-shaped heart beat exhibits a brighter overall color and a larger range. Additionally, the color range of F-type heart beats is the smallest compared to other types, indicating that the color distribution of wavelet time-frequency maps can effectively characterize significant differences between different types of arrhythmias. Therefore, wavelet time-frequency maps have the potential to serve as evaluation indicators for different types of ECG abnormalities. In this work, we extracted the time-frequency features of ECG data for different types of arrhythmias through wavelet time-frequency maps, laying the foundation for subsequent classification tasks.

3.2 Performance evaluation

We utilized 10-fold cross-validation to compare the performance of our proposed model with other commonly used models for arrhythmia detection. To enhance the credibility and robustness of our findings, we conducted intra-patient and inter-patient paradigm analyses, respectively. Intra-patient analyses evaluated the model's ability to track heart rate variability by partitioning the database into 10 groups, with one group serving as the test set and the remainder as the training

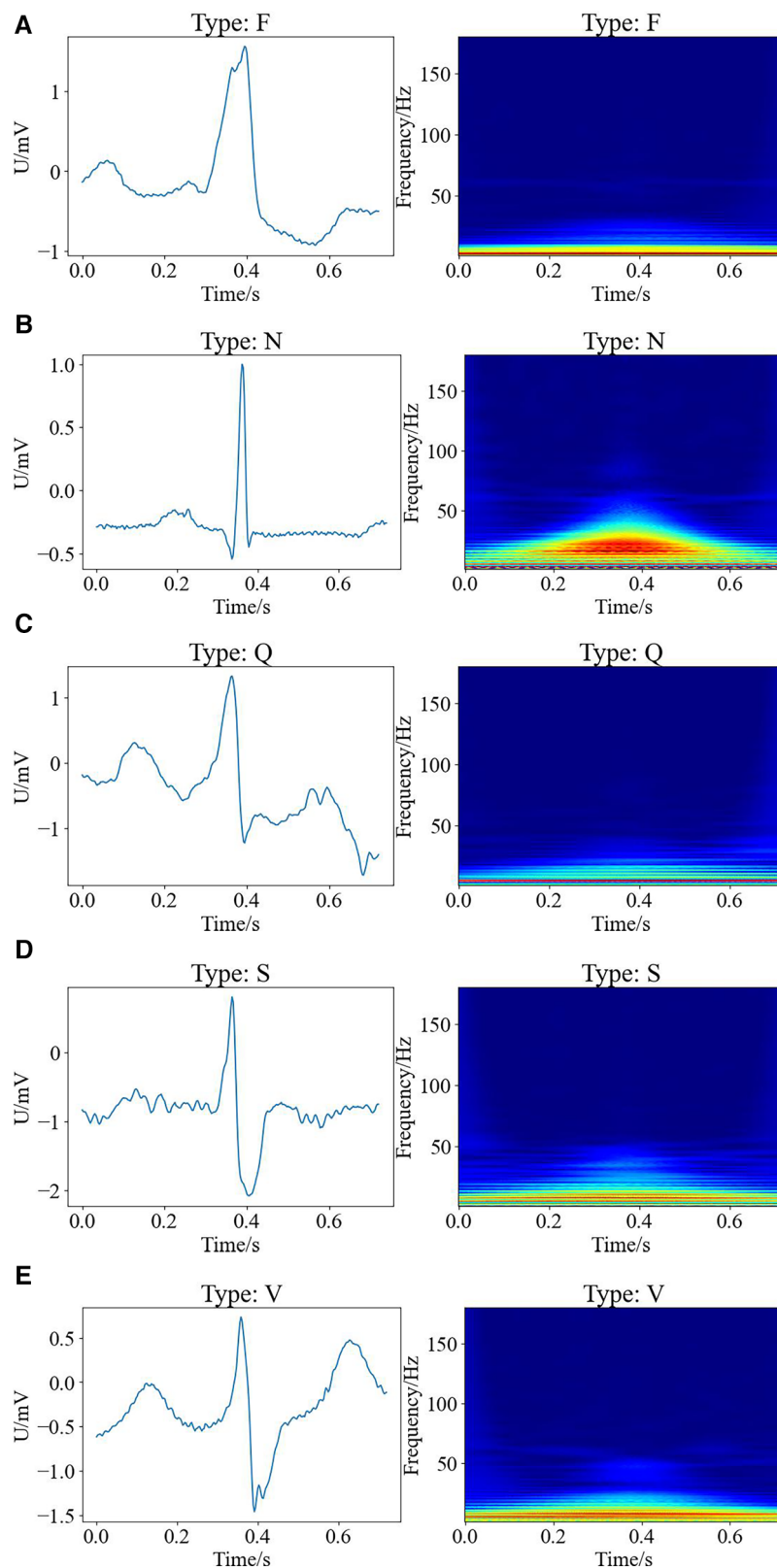


FIGURE 2

Raw signal maps and wavelet time-frequency maps of different types of arrhythmias, with raw signal maps on the left and corresponding wavelet time-frequency maps on the right. (A) F type; (B) N type; (C) Q type; (D) S type; (E) V type.

TABLE 1 Performance comparison between our model and previously reported models for arrhythmia detection in intra-patient paradigm. (unit: %).

Methodology	Accuracy (CI)	Precision (CI)	Sensitivity (CI)	Specificity (CI)	F1-Score (CI)	AUC (CI)
1D-CNN (42)	91.28 (89.49, 93.07)	93.58 (91.79, 95.37)	89.29 (87.50, 91.08)	92.83 (91.04, 94.62)	90.51 (88.72, 92.30)	93.27 (91.48, 95.06)
1D-CNN (28)	97.19 (94.93, 99.45)	95.62 (93.36, 97.88)	96.17 (93.91, 98.43)	91.73 (89.47, 93.99)	97.81 (95.55, 99.07)	96.51 (94.25, 98.77)
CNN-LSTM (43)	97.93 (96.86, 98.99)	95.37 (94.30, 96.44)	98.04 (96.97, 99.11)	97.62 (96.55, 98.69)	97.95 (96.88, 99.02)	97.38 (96.31, 98.45)
2D-CNN (42)	98.74 (97.10, 99.38)	97.69 (96.05, 99.33)	99.08 (98.44, 99.52)	94.82 (93.18, 96.46)	98.63 (97.99, 99.27)	98.14 (96.50, 99.78)
2D-CNN (44)	98.97 (98.83, 99.11)	94.24 (92.10, 96.38)	96.38 (94.24, 98.52)	99.53 (98.39, 99.67)	96.72 (94.58, 98.86)	98.95 (97.81, 99.08)
DBB, AdaBoos (45)	99.04 (98.25, 99.83)	98.37 (97.58, 99.16)	97.52 (96.73, 98.31)	99.31 (98.52, 99.47)	98.84 (98.05, 99.63)	97.62 (96.83, 98.41)
ECG DETR (46)	99.19 (98.56, 99.82)	97.91 (97.28, 98.54)	98.14 (97.51, 98.77)	99.27 (98.64, 99.90)	98.31 (97.68, 98.94)	99.08 (98.45, 99.71)
Proposed	99.34 (98.99, 99.69)	98.71 (98.36, 99.06)	99.49 (99.14, 99.84)	99.57 (99.22, 99.92)	98.96 (98.61, 99.31)	99.35 (99.00, 99.78)

CI: 95% confidence intervals (Lower–Upper bound).

The bold values represent the specific performance values of our proposed method.

set in each iteration. Table 1 illustrates an example of within-patient analysis, showcasing the specific performance of 8 different methods for arrhythmia detection at 95% confidence intervals. The evaluated metrics include accuracy, precision, sensitivity, specificity, F1 score, and AUC. Through the calculation of these confidence intervals in 10-fold cross-validation, we were able to assess the stability and variability of the models across different subsets of random data. This approach aids in identifying and addressing potential overfitting issues and ensures the model's robust generalization capabilities.

However, for clinical applications, the training and testing sets will not be sourced from the same patient. To address this challenge, we conducted an inter-patient paradigm analysis. We divided the 47 subjects into 10 groups. Each time, we randomly selected one group and used their ECG data with different labels as the test set, while the remaining ECG datasets were used as the training set. In other words, the same subject cannot be part of both the training and testing datasets simultaneously. We trained the model using the above-described data and repeated this process ten times, selecting different validation folds each time. Finally, we obtained the average performance indicators, as shown in Table 2.

Compared to the performance records of different methods in detecting arrhythmia, our model achieved classification accuracies of 99.34% and 98.37% in intra-patient and inter-patient case analyses, respectively. These values are 0.15% and 0.63% higher than the current best accuracy achieved by an arrhythmia classification model. Additionally, the classification accuracy of our model reached 98.71% and 96.64%, with sensitivity values of 99.49% and 97.20%, specificity values of 99.57% and 97.12%, F1

scores of 98.96% and 97.64%, and AUC values of 99.35% and 97.68%, respectively. The AUC value (47) provides an overall measure of the model's performance across all possible classification thresholds, so we have included this indicator in Tables 1, 2. The high AUC value of our proposed method demonstrates its strong ability to distinguish between different classes. In summary, our model exhibits superior performance compared to commonly used arrhythmia classification models.

Despite promising results of previous studies in classifying arrhythmias, long-distance dependency remains a major challenge in precise modeling for arrhythmia classification. However, the present study demonstrated improved performances of the model by adopting an advanced model architecture that comprehensively captured long-distance dependencies in ECG.

The confusion matrix is one of the most important methods of evaluating the performance of a classifier, and the confusion matrix for five types of arrhythmias for the test set is illustrated in Figure 3.

Figure 3 shows that the wavelet time-frequency map combined with the Swin Transformer model has a better recognition effect on different arrhythmia states, thus providing reliable support in arrhythmia detection. This result further confirmed that our model could achieve effective classification of arrhythmia and demonstrated the potential of practical applications of this model in the medical field.

3.3 Feature visualization

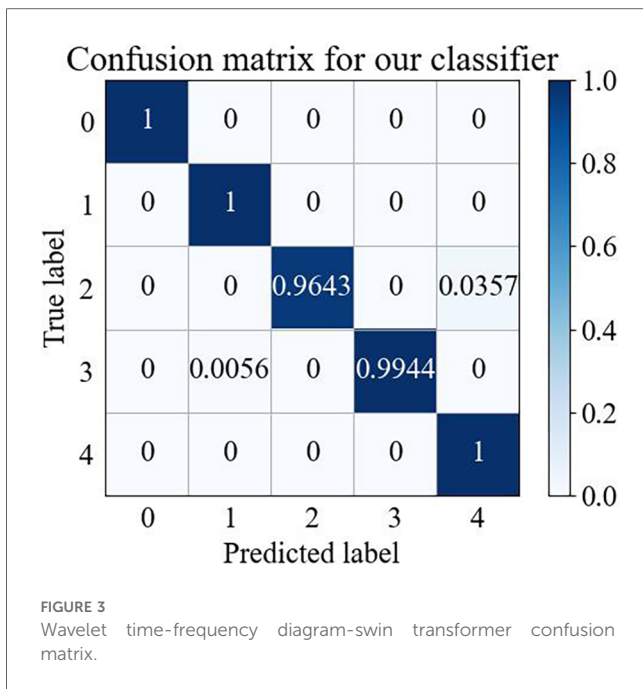
In this study, we used the t-distributed stochastic neighbor embedding (T-SNE) method (48) to visualize the extracted

TABLE 2 Performance comparison between our model and previously reported models for arrhythmia detection in inter-patient paradigm. (unit: %).

Methodology	Accuracy (CI)	Precision (CI)	Sensitivity (CI)	Specificity (CI)	F1-Score (CI)	AUC (CI)
1D-CNN (42)	89.92 (88.10, 91.74)	90.89 (89.07, 92.71)	87.68 (85.86, 89.52)	91.77 (89.95, 93.59)	87.61 (85.79, 89.43)	90.58 (88.76, 92.48)
1D-CNN (28)	94.48 (92.30, 96.66)	93.62 (91.44, 95.82)	94.35 (92.17, 96.53)	89.49 (87.31, 91.67)	95.44 (93.26, 97.62)	93.14 (90.96, 95.32)
CNN-LSTM (43)	96.54 (94.97, 98.11)	92.66 (91.09, 94.23)	95.12 (93.55, 96.69)	96.49 (94.92, 98.06)	96.42 (94.85, 97.99)	95.82 (94.25, 97.39)
2D-CNN (42)	97.28 (95.30, 99.26)	95.55 (93.57, 97.53)	97.93 (95.95, 99.91)	93.06 (91.08, 95.04)	97.01 (95.03, 98.99)	94.37 (92.39, 96.35)
2D-CNN (44)	97.74 (96.20, 99.28)	91.82 (89.28, 94.36)	94.44 (91.90, 96.98)	98.23 (95.69, 98.77)	93.95 (91.41, 96.49)	95.25 (92.71, 97.79)
DBB, AdaBoos (45)	96.73 (95.74, 97.72)	97.31 (96.32, 98.32)	96.14 (95.15, 97.13)	96.56 (95.57, 97.55)	95.98 (94.99, 96.97)	94.55 (93.56, 95.54)
ECG DETR (46)	97.25 (96.18, 98.32)	95.22 (94.15, 96.29)	96.29 (95.22, 97.36)	98.14 (97.07, 99.21)	97.18 (96.11, 98.25)	97.38 (96.31, 98.45)
Proposed	98.37 (97.75, 98.99)	96.64 (96.02, 97.26)	97.20 (96.58, 97.82)	97.12 (96.50, 97.74)	97.64 (97.02, 98.26)	97.68 (97.06, 98.30)

CI: 95% confidence intervals (Lower–Upper bound).

The bold values represent the specific performance values of our proposed method.



multidimensional ECG features in a two-dimensional space (Figure 4). T-SNE is a powerful downscaling and visualization tool for transforming dimensions. It achieves visualization mainly through iterative optimization of the position of high-dimensional data points in the low-dimensional space, to generate closely spaced similar data points in the low-dimensional space.

Figure 4A displays that the raw data of the MIT-BIH arrhythmia dataset has a high degree of imbalance, and the data are arranged haphazardly, with a large number of overlapping signals in different categories, thus presenting a confusing state. Whereas Figure 4B, as the processed ECG data, demonstrates

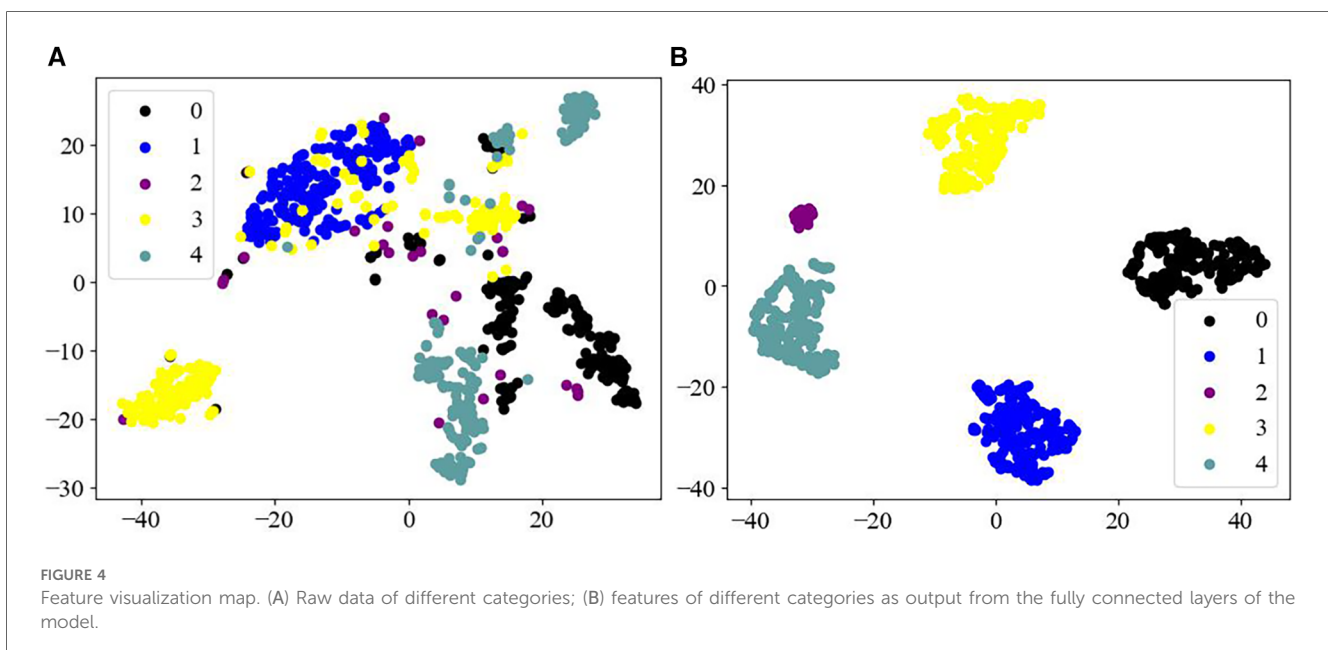
good separability, with five distinct categories of arrhythmias, achieving an effective output for the different categories of the dataset. Therefore, these observations further indicate that the Swin Transformer model may have a reliable classification ability for different forms of ECG datasets.

3.4 Time complexity

CWT is a highly suitable method for multi-scale time-frequency analysis of signals. CWT analyzes signal characteristics at various scales through wavelet transform, offering detailed scale and translation steps, thus enhancing signal processing detail. For each scale, CWT traverses the entire signal and performs convolution operations, with a time complexity of approximately $O(N^2)$, where N is the signal length. While this method demands significant computation, it yields a highly accurate representation of time and frequency, rendering it particularly suitable for the detailed demands of heart rate anomaly detection.

The Swin Transformer reduces complexity by constraining the computation of the self-attention mechanism to local windows, with a complexity of $O(w^2 \times d)$ within each window, where w is the size of the window and d is the feature dimension. Additionally, through a hierarchical window merging strategy, the Swin Transformer further decreases the global computational complexity to nearly $O[N \times \log(N)]$. This structural optimization makes the Swin Transformer efficient and highly expressive when processing large-scale data.

In summary, our method combines the detailed time-frequency analysis of CWT with the efficient data processing capabilities of the Swin Transformer, resulting in more accurate and detailed anomaly detection capabilities. We believe that, for heart rate anomaly detection applications requiring high-precision diagnosis, the associated computational cost is reasonable.



4 Discussion

4.1 Comparison with other feature extraction

ECG feature extraction refers to extracting the key information from the original ECG waveform. The high-dimensional time series data are then transformed into more resolved, simple, and representative low-dimensional features, which not only facilitate the early detection and diagnosis of diseases but also help observe the development of diseases and better understand the disease mechanism. The most common method of ECG feature extraction involves extracting either the time domain or frequency domain of ECG for research purposes.

Time domain analysis of ECG mainly includes measurement and analysis of RR intervals, P and QRS wave clusters, T wave duration, ST segments, and QT intervals. Alotaiby et al. (14) first pre-processed ECG data from the PTB database by detrending and inversion. Then, to construct feature vectors, the pre-processed ECG data were segmented and 11 statistical features were extracted from each segment. The median and mean values were used to describe the concentration trend of the ECG. While the standard deviation, range, and quartiles were utilized to measure the degree of dispersion of the ECG. By analyzing the kurtosis and skewness of the ECG, a deeper understanding of the amplitude distribution characteristics as well as the symmetry of the signals, can be obtained to facilitate further analysis of the ECG. Extracting ECG features from the time domain aspect has a low computational cost and is relatively simple to implement. However, the method may result in loss of information, rendering the extracted features non-representative (49).

ECG frequency domain analysis involves converting the ECG data from the time domain to the frequency domain by analyzing the ECG spectrum and frequency bands. Merri et al. (15) developed a model to characterize and quantify measurement errors introduced due to limited sampling frequency. The model takes into account the RR interval measurement errors caused by the ECG sampling frequency limitation and calculates the first- and second-order statistics of errors to evaluate the influence of the error on the heart rate variability power spectrum. We found that the error power spectrum had an additional high-pass filter-like term for the heart rate variability power spectrum, revealing the importance of an equilibrium between heart rate variability and the error power spectrum. The limitation of the sampling frequency might introduce errors to the extracted ECG features, suggesting that extracting only the frequency domain features may not accurately represent ECG.

In contrast, time-frequency domain analysis considers both the time and frequency characteristics of the signal within an integrated framework, which provides a more comprehensive understanding of the dynamic nature of the signal and reflects the better performance of the model in analyzing cardiac arrhythmias. Among them, wavelet time-frequency diagram belongs to one kind of time-frequency domain analysis with a

wide range of applications in biomedicine, signal processing, and image processing (50–53). Wavelet time-frequency diagram primarily relies on the characteristics of the signal, the selection of a suitable wavelet family as the basis function, and the calculation of the wavelet coefficients. Then the time-frequency diagram is constructed by calculating the amplitude or energy of the coefficients. Given that the wavelet time-frequency map can comprehensively reflect the time-frequency information of the ECG, this study focused on wavelet time-frequency maps to analyze the ECG features of arrhythmia.

4.2 Comparison with other classification models

Recent literature investigating the ECG signals using deep learning models is increasing. Kiranyaz et al. (54) proposed an adaptive one-dimensional convolutional neural network (1D-CNN) that could be trained for different patients to achieve an effective detection of ventricular and supraventricular ectopic beats. Based on the CNN, but 1D-CNN has a fixed requirement on the length of the input signal, which can easily lead to information loss. Jangra et al. (55) proposed an updated CNN, called visual geometry group network (VGGNet), to increase the depth and width of the network. Techniques such as small-size convolutional kernel, pooling layer, batch normalization layer, and dropout layer were used to achieve the extraction of more complex features, thereby enhancing the generalization of the model. The VGGNet achieved better results for arrhythmia classification than the CNN alone. Wang et al. (56) used a long short-term memory (LSTM) model to capture temporal information in ECG data to identify abnormalities in arrhythmias, but the model may suffer from gradient vanishing or gradient explosion when dealing with long sequences. Zhang et al. (29) employed an RNN model to learn strong correlations between consecutive ECG signal points and achieved effective classification of ECG signals at different heart rates, but model may be limited in their ability to model long-term dependencies. Jiang et al. (50) proposed a deep neural network model called Multi-Model Multi-Scale Network (MMnet) for more comprehensive analysis of ECG data, but the complexity of the model as well as the computational cost is high. Gokhan et al. (57) utilized a multi-stage classification system, incorporating ECG waveforms and second-order difference plot (SODP) features, along with a deep belief network (DBN) classifier, to successfully distinguish five types of cardiac arrhythmias. In a separate study, they introduced additional techniques such as wavelet packet decomposition, high-order statistics, morphology, and discrete Fourier transform to enhance feature extraction in a multi-class DBN framework (58). The results demonstrated the method's effectiveness in distinguishing a wide range of heartbeats, albeit requiring substantial data and computational training. While these models have achieved promising results, they also have certain limitations, which we summarize in Table 3.

TABLE 3 Comparison of advantages and disadvantages of different models.

Method	Advantage	Disadvantage
ID-CNN (54)	Able to effectively extract local features.	May lead to information loss.
VGGNet (55)	Capable of extracting complex features and exhibiting good generalization ability.	High computational and training costs.
LSTM (56)	Captures long-term dependencies in time series.	May encounter gradient vanishing or exploding issues when dealing with long sequences.
RNN (29)	Captures dynamic feature relationships in ECG signals.	Limited modeling ability for long-term dependencies.
MMnet (50)	Provides comprehensive recording of ECG information.	High model complexity and computational costs.
DBN (57, 58)	Effectively extracts features.	Requires a large amount of data and computational resources.

In common deep learning models, the chain rule is typically used to perform gradient multiplication across multiple layers. If the gradient was less than 1, the chain multiplication led to a decrease in gradient; if the gradient was greater than 1, the chain multiplication led to an increase in gradient. Either decrease or increase in gradients can seriously affect the model's performance, especially in dealing with long-range dependencies. The transformer delayed the process of gradient loss or increment by implementing certain manipulations, such as residual linking and layer normalization, to facilitate an uninterrupted gradient transfer across the layers. Ding et al. (59) used the Transformer-based ECG reduced-dimensional stacked self-encoder model to effectively overcome the long-distance and the long-term dependence problems as well as accurately limit the parallelization during signal processing for detecting arrhythmia. However, Transformer requires high computational and memory costs when facing large-scale images. While Swin Transformer (60), as an emerging deep learning model, has certain advantages in image processing, adopting the windowed self-attention mechanism to partition the image into multiple windows for reducing the computational cost, improving the semantic understanding of the image through global modeling, enhancing the information exchange between the windows by alternately executing the windowed self-attention and translational windowed self-attention, and adapting to different scales of feature extraction by adjusting the size of the window. Swin Transformer combines the advantages of CNN and Transformer in extracting local features and location information, which takes into account both local and global feature extraction capabilities to achieve a precise feature classification. Therefore, in this study, Swin Transformer was chosen to detect arrhythmia, aiming to accurately record ECG information and achieve rapid diagnosis of CVD.

4.3 Limitations and future research lines

Although our study demonstrates advancement in developing an arrhythmia detection model through precise classification of

ECG data, ECG-related studies are very complex in the sense that the model classifier can also be susceptible to multiple interfering factors such as muscle movement and external noise. There are still certain errors in our model, even after pre-processing the ECG dataset. When applying the classification model, individual differences between different clinical patients should be carefully considered. Therefore, in future studies, we will plan to analyze larger patient ECG datasets, pay more attention to the robustness of the model, test its accuracy, and make it closer to real-world clinical applications.

5 Conclusion

In this study, we propose a novel method for detecting arrhythmias using wavelet time-frequency maps and the Swin Transformer. This technology is particularly adept at capturing subtle and transient changes in ECG signals, which are crucial for early arrhythmia diagnosis, especially in asymptomatic patients. Accurate diagnosis facilitates timely interventions by healthcare professionals, enables the development of personalized treatment plans, and ultimately improves patient prognosis and quality of life, highlighting the effectiveness of the Swin Transformer classification model in diagnosing arrhythmias. In our approach, we first employed the wavelet algorithm to denoise the data from the MIT-BIH Arrhythmia Database, thereby enhancing the quality of the original signal. Subsequently, we utilized wavelet time-frequency maps to extract time-frequency features from the ECG signals, effectively representing their distribution. Finally, we applied the Swin Transformer for automatic ECG classification, achieving accuracy rates of 99.34% in intra-patient case analysis and 98.37% in inter-patient case analysis.

Traditional ECG analysis is often limited by time and frequency resolution. However, our proposed new method overcomes this limitation, which is crucial for understanding the complexity of arrhythmias. By combining wavelet time-frequency maps with the Swin Transformer, doctors can analyze electrocardiogram data more accurately and quickly detect arrhythmias, thereby improving diagnostic speed and accuracy. The introduction of the Swin Transformer aims to utilize its advanced self-attention mechanism, which demonstrates excellent performance in processing ECG signals with high spatiotemporal dynamics. Our method achieves high accuracy by analyzing ECG data at multiple frequency and time scales, a crucial aspect in clinical applications as it can provide a more comprehensive diagnostic perspective than traditional methods.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval was not required for this study in accordance with the local legislation and institutional requirements.

Author contributions

SC: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. HW: Writing – original draft. HZ: Writing – original draft. CP: Methodology, Writing – review & editing. YL: Supervision, Writing – review & editing. BW: Supervision, Writing – review & editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article.

References

- Ma LY, Chen WW, Gao RL, Liu LS, Zhu ML, Wang YJ, et al. China cardiovascular diseases report 2018: an updated summary. *J Geriatr Cardiol.* (2020) 17(1):1–8. doi: 10.11909/j.issn.1671-5411.2020.01.001
- Tang X, Ma Z, Hu Q, Tang W. A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated linear-kernel support vector machines. *IEEE Trans Biomed Eng.* (2019) 67(4):978–86. doi: 10.1109/TBME.2019.2926104
- Ahmed AA, Ali W, Abdullah TA, Malebary SJ. Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model. *Mathematics.* (2023) 11(3):562. doi: 10.3390/math11030562
- Jabaudon D, Sztajzel J, Sievert K, Landis T, Sztajzel R. Usefulness of ambulatory 7-day ECG monitoring for the detection of atrial fibrillation and flutter after acute stroke and transient ischemic attack. *Stroke.* (2004) 35(7):1647–51. doi: 10.1161/01.STR.0000131269.69502.d9
- Freedman B, Potpara TS, Lip GY. Stroke prevention in atrial fibrillation. *Lancet.* (2016) 388(10046):806–17. doi: 10.1016/S0140-6736(16)31257-0
- Yıldırım Ö, Plawiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med.* (2018) 102:411–20. doi: 10.1016/j.combiomed.2018.09.009
- Arevalo HJ, Vadakkumpadan F, Guallar E, Jebb A, Malamas P, Wu KC, et al. Arrhythmia risk stratification of patients after myocardial infarction using personalized heart models. *Nat Commun.* (2016) 7(1):11437. doi: 10.1038/ncomms11437
- Cundari G, Alkadhri H, Eberhard M. The role of CT in arrhythmia management—treatment planning and post-procedural imaging surveillance. *Br J Radiol.* (2023) 96(1151):20230028. doi: 10.1259/bjr.20230028
- Sharma RR, Pachori RB. Baseline wander and power line interference removal from ECG signals using eigenvalue decomposition. *Biomed Signal Process Control.* (2018) 45:33–49. doi: 10.1016/j.bspc.2018.05.002
- Zhang WW, Li M, Zhao JY. Research on electrocardiogram signal noise reduction based on wavelet multi-resolution analysis. *International Conference on Wavelet Analysis and Pattern Recognition* (2010). p. 351–4. doi: 10.1109/ICWAPR.2010.5576381
- Zhang S, Ma Y, Zhang Q, Gao Y. Motor imagery electroencephalogram de-noising method based on EEMD and improved wavelet threshold. *Chinese Control and Decision Conference* (2018). p. 5589–94. doi: 10.1109/CCDC.2018.8408106
- Jianhua Z, Qiang Z, Jinrong Z, Lin S, Jilong W. A novel algorithm for threshold image denoising based on wavelet construction. *Cluster Comput.* (2019) 22:12443–50. doi: 10.1007/s10586-017-1655-0
- Wu H, Zhang Y, Chen H, Xiao W, Huang L, Xiao Y, et al. The improved wavelet denoising scheme based on robust principal component analysis for distributed fiber acoustic sensor. *IEEE Sens J.* (2023) 23:22944–51. doi: 10.1109/JSEN.2023.3305532
- Alotaiby TN, Alrshoud SR, Alshebeili SA, Aljafar LM. ECG-based subject identification using statistical features and random forest. *J Sens.* (2019) 2019:1–3. doi: 10.1155/2019/6751932
- Merri M, Farden DC, Mottley JG, Titlebaum EL. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Trans Biomed Eng.* (1990) 37(1):99–106. doi: 10.1109/10.43621
- Sultan Qurraie S, Ghorbani Afkhami R. ECG Arrhythmia classification using time frequency distribution techniques. *Biomed Eng Lett.* (2017) 7:325–32. doi: 10.1007/s13534-017-0043-2
- Sharma RR, Kumar M, Pachori RB. Joint time-frequency domain-based CAD disease sensing system using ECG signals. *IEEE Sens J.* (2019) 19(10):3912–20. doi: 10.1109/JSEN.2019.2894706
- Lai D, Fan X, Zhang Y, Chen W. Intelligent and efficient detection of life-threatening ventricular arrhythmias in short segments of surface ECG signals. *IEEE Sens J.* (2020) 21(13):14110–20. doi: 10.1109/JSEN.2020.3031597
- Diker A, Avci D, Avci E, Gedikpinar M. A new technique for ECG signal classification genetic algorithm wavelet kernel extreme learning machine. *Optik (Stuttg).* (2019) 180:46–55. doi: 10.1016/j.ijleo.2018.11.065
- Hajeb-Mohammadipour S, Ahmadi M, Shahghadami R, Chon KH. Automated method for discrimination of arrhythmias using time, frequency, and nonlinear features of electrocardiogram signals. *Sensors.* (2018) 18(7):2090. doi: 10.3390/s18072090
- Gao Z, Dang W, Wang X, Hong X, Hou L, Ma K, et al. Complex networks and deep learning for EEG signal analysis. *Cogn Neurodyn.* (2021) 15(3):369–88. doi: 10.1007/s11571-020-09626-1
- Coşkun M, Uçar A, Yıldırım Ö, Demir Y. Face recognition based on convolutional neural network. In *2017 International Conference on Modern Electrical and Energy Systems (MEES); Kremenchuk, Ukraine.* (2017). p. 376–9. doi: 10.1109/MEES.2017.8248937
- Wu Q, Liu Y, Li Q, Jin S, Li F. The application of deep learning in computer vision. In *2017 Chinese Automation Congress (CAC); Jinan, China.* (2017). p. 6522–7. doi: 10.1109/CAC.2017.8243952
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* (1998) 86(11):2278–324. doi: 10.1109/5.726791
- Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* (2019) 25(1):65–9. doi: 10.1038/s41591-018-0268-3
- Hammad M, Ilyasu AM, Subasi A, Ho ES, Abd El-Latif AA. A multitier deep learning model for arrhythmia detection. *IEEE Trans Instrum Meas.* (2020) 70:1–9. doi: 10.1109/TIM.2020.3033072

27. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation*. (2022) 145(2):122–33. doi: 10.1161/CIRCULATIONAHA.121.057480
28. Ullah A, Rehman SU, Tu S, Mehmood RM, Fawad, Ehatisham-Ul-Haq M. A hybrid deep CNN model for abnormal arrhythmia detection based on cardiac ECG signal. *Sensors*. (2021) 21(3):951. doi: 10.3390/s21030951
29. Zhang C, Wang G, Zhao J, Gao P, Lin J, Yang H. Patient-specific ECG classification based on recurrent neural networks and clustering technique. In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed); Innsbruck, Austria*. (2017). p. 63–7. doi: 10.2316/P.2017.852-029
30. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. (2017) 30.
31. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021). p. 9992–10002. doi: 10.1109/ICCV48922.2021.00986
32. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol Mag*. (2001) 20(3):45–50. doi: 10.1109/51.932724
33. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. (2000) 101(23):e215–20. doi: 10.1161/01.CIR.101.23.e215
34. Shoughi A, Dowlatshahi MB. A practical system based on CNN-BLSTM network for accurate classification of ECG heartbeats of MIT-BIH imbalanced dataset. *International Conference, Computer Society of Iran* (2021). doi: 10.1109/CSICC52343.2021.9420620
35. Kuila S, Dhanda N, Joardar S. Feature extraction and classification of mit-bih arrhythmia database. *International Conference on Communication, Devices and Computing* (2020). p. 417–27. doi: 10.1007/978-981-15-0829-5_41
36. Escalona OJ, Mendoza M, Villegas G, Navarro C. Real-time system for high-resolution ECG diagnosis based on 3D late potential fractal dimension estimation. In *2011 Computing in Cardiology; Hangzhou, China*. (2011). p. 789–92.
37. Liu H, You Y, Li S, He D, Sun J, Wang J, et al. Denoising of laser self-mixing interference by improved wavelet threshold for high performance of displacement reconstruction. *Photonics*. (2023) 10(8):43. doi: 10.3390/photonics10080943
38. Chikh MA, Behadada O. A PVC beats recognition using fuzzy classifier. *J Mech Med Biol*. (2010) 10(02):327–39. doi: 10.1142/S021951941000337X
39. Sayantan G, Kien PT, Kadambari KV. Classification of ECG beats using deep belief network and active learning. *Med Biol Eng Comput*. (2018) 56(10):1887–98. doi: 10.1007/s11517-018-1815-2
40. Chang C, Wang Q, Jiang J, Jiang Y, Wu T. Voltage fault diagnosis of a power battery based on wavelet time-frequency diagram. *Energy*. (2023) 278:127920. doi: 10.1016/j.energy.2023.127920
41. Wu Z, Lan T, Yang C, Nie Z. A novel method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks. *IEEE Access*. (2019) 7:170820–30. doi: 10.1109/ACCESS.2019.2956050
42. Huang J, Chen B, Yao B, He W. ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access*. (2019) 7:92871–80. doi: 10.1109/ACCESS.2019.2928017
43. Oh SL, Ng EY, San Tan R, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med*. (2018) 102:278–87. doi: 10.1016/j.compbiomed.2018.06.002
44. Ullah A, Anwar SM, Bilal M, Mehmood RM. Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation. *Remote Sens (Basel)*. (2020) 12(10):1685. doi: 10.3390/rs12101685
45. Rajesh KN, Dhuli R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomed Signal Process Control*. (2018) 41:242–54. doi: 10.1016/j.bspc.2017.12.004
46. Hu R, Chen J, Zhou L. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Comput Biol Med*. (2022) 144:105325. doi: 10.1016/j.compbiomed.2022.105325
47. Altan G. DeepOCT: an explainable deep learning architecture to analyze macular edema on OCT images. *Eng Sci Technol*. (2022) 34:101091. doi: 10.1016/j.jestch.2021.101091
48. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res*. (2014) 15(1):3221–45. doi: 10.5555/2627435.2697068
49. Singh AK, Krishnan S. ECG signal feature extraction trends in methods and applications. *Biomed Eng Online*. (2023) 22(1):22. doi: 10.1186/s12938-023-01075-1
50. Jiang S, Li D, Zhang Y. A deep neural network based on multi-model and multi-scale for arrhythmia classification. *Biomed Signal Process Control*. (2023) 85:105060. doi: 10.1016/j.bspc.2023.105060
51. Gao D, Zhu Y, Wang X, Yan K, Hong J. A fault diagnosis method of rolling bearing based on complex morlet CWT and CNN. *Prognostics and System Health Management Conference* (2018). p. 1101–5. doi: 10.1109/PHM-Chongqingdoi: .2018.00194
52. He K, Li Q, Yang Q. Characteristic analysis of welding crack acoustic emission signals using synchro-squeezed wavelet transform. *J Test Eval*. (2018) 46(6):2679–91. doi: 10.1520/JTE20170218
53. Miao R, Gao Y, Ge L, Jiang Z, Zhang J. Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network. *Comput Ind*. (2019) 112:103115. doi: 10.1016/j.compind.2019.07.005
54. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans Biomed Eng*. (2015) 63(3):664–75. doi: 10.1109/TBME.2015.2468589
55. Jangra M, Dhull SK, Singh KK. ECG Arrhythmia classification using modified visual geometry group network (mVGGNet). *J Intell Fuzzy Syst*. (2020) 38(3):3151–65. doi: 10.3233/JIFS-191135
56. Wang L, Zhou X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors*. (2019) 19(7):1502. doi: 10.3390/s19071502
57. Altan G, Kutlu Y, Allahverdi N. A multistage deep belief networks application on arrhythmia classification. *Int J Intell Syst Appl Eng*. (2016) 4:222–8. doi: 10.18201/ijisae.270367
58. Altan G, Allahverdi N, Kutlu Y. A multistage deep learning algorithm for detecting arrhythmia. In: *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)* (2018). p. 1–5. doi: 10.1109/CAIS.2018.8441942
59. Ding C, Wang S, Jin X, Wang Z, Wang J. A novel transformer-based ECG dimensionality reduction stacked auto-encoders for arrhythmia beat detection. *Med Phys*. (2023) 50(9):5897–912. doi: 10.1002/mp.16534
60. He X, Zhou Y, Zhao J, Zhang D, Yao R, Xue Y. Swin transformer embedding UNet for remote sensing image semantic segmentation. *IEEE Trans Geosci Remote Sens*. (2022) 60:1–5. doi: 10.1109/TGRS.2022.3144165