HA-ResNet: Residual Neural Network With Hidden Attention for ECG Arrhythmia Detection Using Two-Dimensional Signal

Yuxia Guan, Ying An, Jingrui Xu, Ning Liu, and Jianxin Wang

Abstract—Arrhythmia is an abnormal heart rhythm, a common clinical problem in cardiology. Long-term or severe arrhythmia may lead to stroke and sudden cardiac death. The electrocardiogram (ECG) is the most commonly used tool to diagnose arrhythmia. However, the traditional diagnosis relies on experts for manual interpretation, which is time-consuming and laborious. In recent years, many automatic arrhythmia detection methods have emerged due to advancements in deep learning. These methods can reduce manual intervention and improve diagnostic efficiency. However, extracting useful features from raw ECG signals for arrhythmia detection is still challenging due to the low frequency of ECG signals and noise distribution. In this paper, we propose a novel hidden attention residual network (HA-ResNet) for automated arrhythmia classification. In this model, the one-dimensional ECG signals are first converted into two-dimensional images and fed into an embedding layer to obtain the relevant shallow features in ECG. Then, a hidden attention layer combining Squeeze-and-Excitation (SE) block and Bidirectional Convolutional LSTM (BConvLSTM) is used to further capture the deep Spatio-temporal features. We evaluate our HA-ResNet on two public datasets and achieve F1 scores of 96.0%, 96.7%, and 87.6% on 2s segments, 5s segments, and 10s segments, respectively, which significantly outperform the existing state-of-the-art approaches. The experimental results demonstrate the effectiveness and generalization of our method.

Index Terms—ECG classification, arrhythmia, dedep learning

1 Introduction

Arrhythmia is a common problem in cardiology, which can be caused by slow, fast, or irregular heartbeat in the atrium, ventricle, other parts, or multiple parts of the heart. In healthy people, the conduction tissue of the heart transmits electrical signals in the form of sequential activation waves [1]. This ordered electrical conduction may be interrupted by changing pulse formation (focal activity) or changing conduction (re-entry), resulting in arrhythmia [2]. Clinical manifestations can range from asymptomatic to symptoms such as palpitations, shortness of breath, fainting, and even sudden cardiac death. Therefore, early screening and accurate diagnosis of arrhythmia are of great clinical significance.

The electrocardiogram (ECG) is the most common tool for diagnosing arrhythmias. Using a sensor placed on the skin, the potential of the central muscle tissue of each cardiac cavity is continuously activated during each heartbeat. The morphological changes and temporal relationship of individual heartbeat provide diagnostic clues about the origin and type of arrhythmia [3]. However, due to the complexity and nonstationarity of ECG signals, ECG heartbeat classification is a difficult task for researchers [4]. Especially for some intermittent arrhythmias, the diagnosis requires doctors to analyze long-term continuous ECG signals and make subjective judgments based on relevant knowledge and experience, which is extremely time-consuming and laborious [5]. Therefore, there is an urgent need to design automatic ECG analysis tools for effectively improving medical quality and efficiency. [6].

Early arrhythmia detection mostly involves three main steps that are pre-processing of the ECG signal, feature extraction, and classification [7]. Martis et al. [8] used the higher order spectra (HOS) method to extract features, performed independent quantity analysis (ICA) to select high-dimensional significant features, and finally used the k-nearest neighbour classifier to classify atrial fibrillation (AFIB), atrial flutter (AFL), and normal (NSR) ECG signals. Elhaj et al. [9] proposed a technique for the arrhythmia classification of ECG signals based on a combination of linear and nonlinear feature extraction techniques. Acharya et al. [10] extracted 14 entropy features from ECG signals and fed them into a decision tree classifier to identify NSR, AFIB, AFL, and Ventricular Fibrillation (VFIB). Most of these methods often require researchers to manually select features based on experience. However, hand-crafted features may change due to noise, scaling, and translation, which makes them not generalize well on unseen data and thus limits the classification performance.

- *Yuxia Guan, Jingrui Xu, Ning Liu, and Jianxin Wang are with the School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China. E-mail: {guanyx1997, xujinrui, liun1996, jxwang}@csu.edu.cn.*
- *Ying An is with Big Data Institute, Central South University, Changsha, Hunan 410083, China. E-mail: anying@csu.edu.cn.*

Manuscript received 3 April 2022; revised 21 July 2022; accepted 7 August 2022. Date of publication 16 August 2022; date of current version 26 December 2023.
This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFF1201200, in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1909208, and in part by the Science and Technology Major Project of Changsha under Grant kh2202004.
(Corresponding author: Jianxin Wang.)
Digital Object Identifier no. 10.1109/TCBB.2022.3198998

In recent years, deep learning has been successfully applied in the fields of computer vision [11], image classification, and medical data analysis [12], [13], [14], [15], [16]. A large number of deep learning-based methods for automatic ECG detection have been proposed. For example, Acharya et al. [6] proposed an 11 layer CNN to automatically extract and classify the four categories of NSR, AFIB, AFL, and VFIB of 2s and 5s ECG segments. Fan et al. [17] proposed a multiscale convolution neural network to detect atrial fibrillation from single-lead ECG records and achieved 98.13% classification accuracy for 20s ECG segments. Petmezas et al. [18] used a fusion model of CNN and LSTM to classify four types of heart rhythms such as atrial fibrillation, atrial flutter, and finally achieved 97.87% sensitivity and 99.29% specificity. Hannun et al. [3] proposed a deep convolution neural network that includes 16 residual blocks to classify 12 rhythms using a single lead ECG signal, and obtained 83.7% F1 score, which is comparable to cardiologists.

Most of the above methods take one-dimensional ECG signals as input and utilize deep neural networks to automatically extract relevant latent features in an end-to-end manner without manual intervention. However, the information provided by one-dimensional ECG signals is relatively limited, which greatly affects the effectiveness of these methods. Recently, some studies have found that more abundant feature information can be captured in the two-dimensional images converted from the one-dimensional ECG signals, which can effectively improve the classification performance of the model. For example, Huang et al. [19] first converted the original ECG signals into time-frequency spectrums through the short-time Fourier transform (STFT), which was then fed into a 2D-CNN to classify five types of arrhythmias. The experimental results show that its classification accuracy is nearly 8.07% higher than the direct one-dimensional signals method. Alqudah et al. [20] compared four different ECG spectrum representations and four CNN architectures for classifying six different types of ECG arrhythmias. Finally, they proposed an ensemble method that integrated the outputs of multiple different CNN models using majority voting to improve the accuracy of arrhythmia classification. Naz et al. [21] proposed a 2D-CNN for ventricular arrhythmias detection using images transformed from ECG signals. Unfortunately, these models usually integrate features in a simple linear fashion and cannot adequately capture the complex spatial dependencies and temporal dynamics in ECG signals.

To address the above-mentioned issues, we propose a novel end-to-end ECG automatic classification model that employs hidden attention residual neural network (HA-ResNet) to learn relevant spatial and temporal features from ECG signals for efficient arrhythmia detection. In this model, the one-dimensional ECG signals are first converted into two-dimensional images and fed into an embedding layer to obtain the relevant shallow features in ECG. Then, a hidden attention layer combining SE block and Bi-directional Convolutional LSTM (BConvLSTM) [22] is used to further capture the deep Spatio-temporal features.

Fig. 1 shows the overall flow chart. And, the main contributions of this work can be summarized as follows:

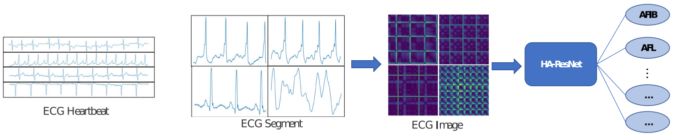


Fig. 1. Block diagram of the proposed system.

- We convert ECG signals into images through three different two-dimensional transformation methods to provide richer feature information.
- We propose a novel framework that includes hidden attention layers for automated arrhythmia classification. And the SE block and BConvLSTM are used to enhance the ability to learn the spatial dependencies and temporal dynamics from ECG signals.
- We evaluate our method for arrhythmia classification on two datasets and compare it with state-of-the-art methods. The results validate the efficiency and generalization of the proposed method.

The rest of this paper is organized as follows. We describe the methods in Section 2. Then, Section 3 illustrates the experiments and results. Finally, the discussions and conclusions are made in Section 4.

2 Methodology

The architecture of the HA-ResNet is shown in Fig. 2. In our approach, the one-dimensional ECG signals are first converted into images and fed to a 2-D convolution layer followed by a batch normalization (BN) layer and a max-pooling layer for shallow feature extraction. Then, we used four HA modules (HAM) with different parameters to capture the deep features in ECG signals. Finally, the deep features are sequentially input to an average pooling layer and a softmax layer to obtain the final classification result.

2.1 Data Preprocessing

We convert the raw one-dimensional data into images by three different two-dimensional transformation methods. The specific methods are described as follows. Moreover, an example of corresponding two-dimensional images obtained by converting 2s ECG segments of various arrhythmias through different conversion methods is also demonstrated in Fig. 3.

2.1.1 Image Formation by Recurrence Plot (RP)

The RP [23] can reveal the internal structure of time series and reflect the stationarity and internal phase of time series. Moreover, it can provide a priori knowledge of similarity, information content, and predictability. Therefore, it is especially suitable for short time-series data, such as Electroencephalography (EEG) and ECG. The image formed by RP is shown in Fig. 3a.

Let $q(t) \in R_d$ be a multi-variate time series. The recurrence plot is defined as:

$$RP = \theta(\varepsilon - ||q(i) - q(j)||), \tag{1}$$

where ε is a threshold, and θ is the Heaviside function.

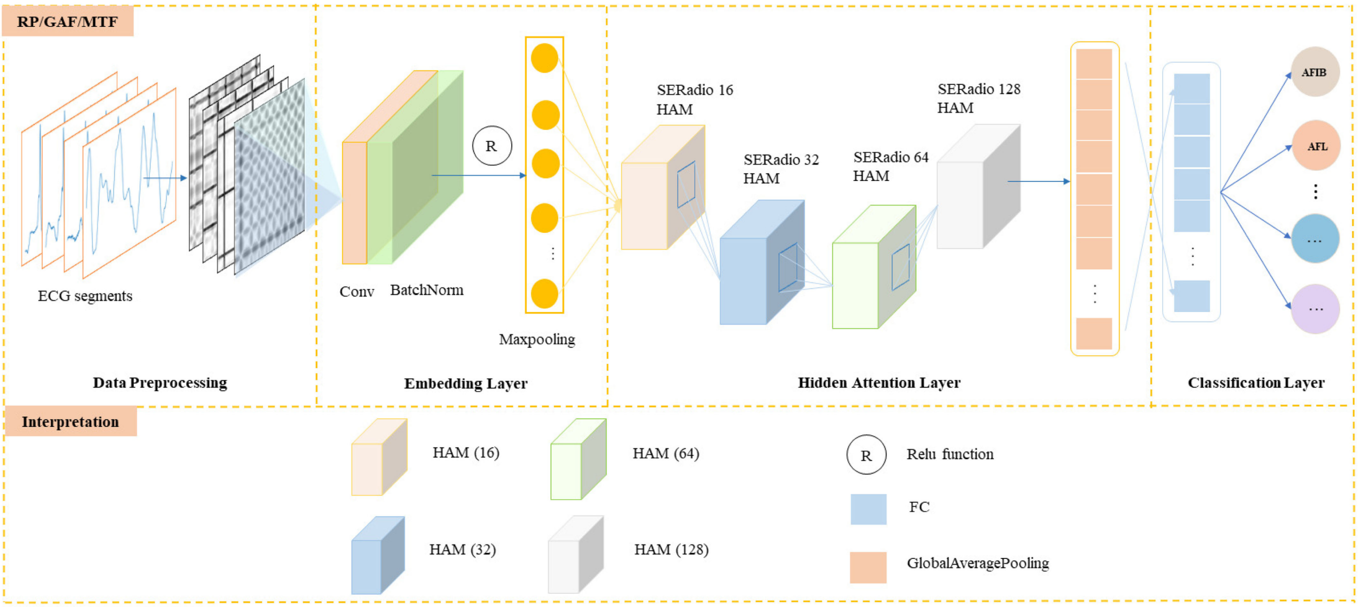


Fig. 2. A block representation of the proposed HA-ResNet model for arrhythmia detection.

2.1.2 Image Formation by Gramian Angular Field (GAF)

The GAF[24] is used to create matrixes of temporal correlations for time series. First, It rescales the time series in a range $[a, b]$ where $-1 \leq a < b \leq 1$. Then the arccos is token to compute the polar coordinates of the scaled time series. Finally, on the one hand, we can get the Gramian Angular Summation Field (GASF) by computing the cosine of the sum of the angles. On the other hand, we can get the Gramian Angular Difference Field (GADF) by computing the sine of the difference of the angles. The specific calculation process can be expressed as shown in Eq. (2). The image formed by GAF is shown in Fig. 3b.

$$\begin{aligned} \tilde{x}_i &= a + (b-a) \times \frac{x_i - \min(x)}{\max(x) - \min(x)}, \forall i \in \{1, \ldots, n\} \\ \phi_i &= \arccos\left(\tilde{x}_i\right), \forall i \in \{1, \ldots, n\} \\ GASF_{i,j} &= \cos\left(\phi_i + \phi_j\right), \forall i, j \in \{1, \ldots, n\} \\ GADF_{i,j} &= \cos\left(\phi_i - \phi_j\right), \forall i, j \in \{1, \ldots, n\}, \end{aligned} \tag{2}$$

where $x^{\sim}{}_i$ is a time series, a is the lower bound of scaling and b is the upper bound of scaling.

2.1.3 Image Formation by Markov Transition Field (MTF)

The MTF[24] discretizes a time series $X = x_1, x_2, \ldots, x_n$ into bins q_k. Then, a weighted adjacency matrix W is constructed by counting transitions among quantile bins in the manner of a first-order Markov chain. Finally, the Markov Transition Matrix is computed by W. The Markov transition field matrix is given by Eq. (3). The image formed by MTF is shown in Fig. 3c.

$$MTF = \begin{bmatrix} w_{lk|x_1 \in q_l, x_1 \in q_k} \cdots w_{lk|x_1 \in q_l, x_n \in q_k} \\ w_{lk|x_2 \in q_l, x_1 \in q_k} \cdots w_{lk|x_2 \in q_l, x_n \in q_k} \\ \vdots \\ w_{lk|x_n \in q_l, x_1 \in q_k} \cdots w_{lk|x_n \in q_l, x_n \in q_k} \end{bmatrix}, \tag{3}$$

where w_{lk} is the frequency with which a point in quantile q_k is followed by a point in quantile q_l.

2.2 Feature Extraction

2.2.1 Embedding Layer

The images transformed from ECG signals are fed into the embedding layer. There is a 2-D convolution layer, a batch normalization layer, and a max-pooling layer in the embedding layer. And we can obtain shallow features in the ECG data through it. Then, the shallow features will be fed into the hidden attention layer to further extract the deep features.

2.2.2 Hidden Attention Layer

As shown in Fig. 4, the hidden attention layer consists of 4 HA modules (HAM) and a global average pooling layer. Each HAM contains a residual block and a BConvLSTM block. Specifically, the input is split into lower-dimensional

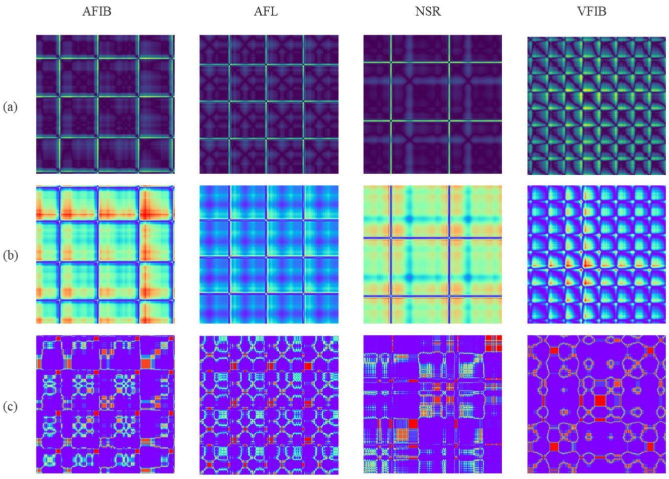


Fig. 3. An example of images for the two-second segments, (a) RP, (b) GAF, (c) MTF.

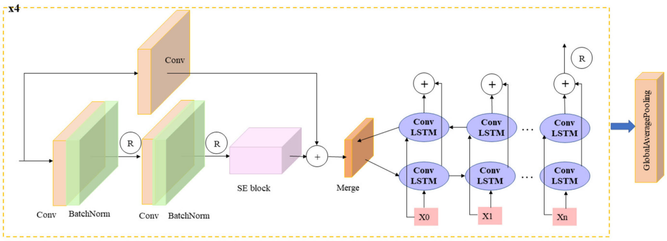


Fig. 4. Overview of the hidden attention layer.

embeddings by a 64 1 * 1 convolution layer with a stride of 1. After that, we apply batch normalization (BN) and a rectified linear activation (ReLU) to accelerate the convergence speed and prevent overfitting. Then, the output embedding vector will go through another convolutional layer with BN and ReLU and input into an SE block to further obtain the correlation in ECG features. Finally, the outputs of the SE block and a projected shortcut connection (done by a 64 1*1 convolution layer with a stride of 1) will be combined and fed into a BConvLSTM to obtain the representation of deep Spatio-temporal dependencies. The structure of the remaining HAM is the same as that of the first HAM, but some parameters are different. Table 1 shows the detailed information about the HAM. After going through four HAM, the global average pooling layer is added to regularize the structure of the whole network and strengthen the correspondence between feature mapping and categories. The key components in the hidden attention layer are described in detail below.

SE Block: The SE block [25] is shown in Fig. 5 which can give different weights to different channels of the feature maps. So we use it to increase the model's sensitivity to crucial features. First, we input the feature map into a global average pooling layer. Then we use two full connection layers to further extract features. The first fully connected layer is used for dimensionality reduction. The second fully connected layer is used to add more nonlinear processes and fit the complex correlation between channels. Finally, a complete multiplication operation is used to combine the original feature maps and their important relations to obtain enhanced feature representations.

BConvLSTM: The structure of the BConvLSTM is shown in Fig. 6. The BConvLSTM is more efficient than the ConvLSTM [26] for handling long time series since it can learn the dependencies in the sequence from both forward and backward directions to fully utilize the past and future information of the current state.

The standard ConvLSTM can be formulated as follows:

$$\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} * C_{t-1} + b_i)\\
f_t &= \sigma\big(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} * C_{t-1} + b_f\big)\\
C_t &= f_t \circ C_{t-1} + i_t \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c)\\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_{t-1} + b_c)\\
H_t &= o_t \circ \tanh(C_t),
\end{aligned} \tag{4}$$

where $*$ and $\circ$ denote the convolution and Hadamard functions, respectively. X_t is the input tensor (in our case X_e and X_i). H_t is the hidden state tensor, C_t is the memory cell tensor, and W_{x*} and W_{h*} are 2D Convolution kernels corresponding to the input and hidden state, respectively, and b_i, b_f, b_o, and b_c are the bias terms.

TABLE 1
Detailed Information of the HAM

Part	Layer	Reduction ratio	Kernel size	Number of kernel
HAM1	Conv	-	1x1	64
	BN+ReLU	-	-	-
	Conv	-	3x3	64
	BN+ReLU	-	-	-
	SE block	16	-	-
	Conv (shortcut connection)	-	1x1	64
	BconvLSTM	-	3x3	64
HAM2	Conv	-	1x1	128
	BN+ReLU	-	-	-
	Conv	-	3x3	128
	BN+ReLU	-	-	-
	SE block	32	-	-
	Conv (shortcut connection)	-	1x1	128
	BconvLSTM	-	3x3	128
HAM3	Conv	-	1x1	256
	BN+ReLU	-	-	-
	Conv	-	3x3	256
	BN+ReLU	-	-	-
	SE block	64	-	-
	Conv (shortcut connection)	-	1x1	256
	BconvLSTM	-	3x3	256
HAM4	Conv	-	1x1	512
	BN+ReLU	-	-	-
	Conv	-	3x3	512
	BN+ReLU	-	-	-
	SE block	128	-	-
	Conv (shortcut connection)	-	1x1	512
	BconvLSTM	-	3x3	512

For the BiConvLSTM, the forward ConvLSTM takes sequence of ECG segments in a regular order from $t = 1$ to $t = T_i$, as input and computes the forward hidden state $\overrightarrow{H_t}$. Similarly, the backward ConvLSTM reads ECG segments in a reverse order and calculates the backward hidden state $\overleftarrow{H_t}$. Then, the forward hidden state $\overrightarrow{H_t}$ and the backward hidden state $\overleftarrow{H_t}$ are combined by using a weighted sum to obtain the spatio-temporal feature

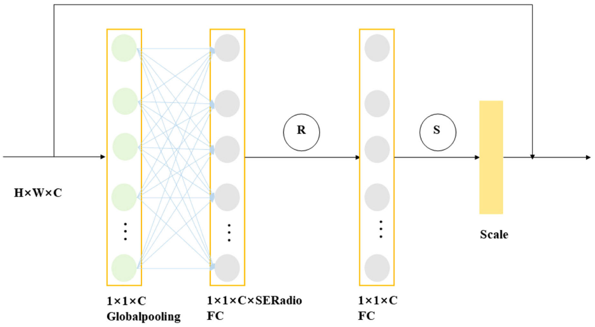


Fig. 5. The SE block used in this paper. R: Rectified linear unit, S: Sigmoid.

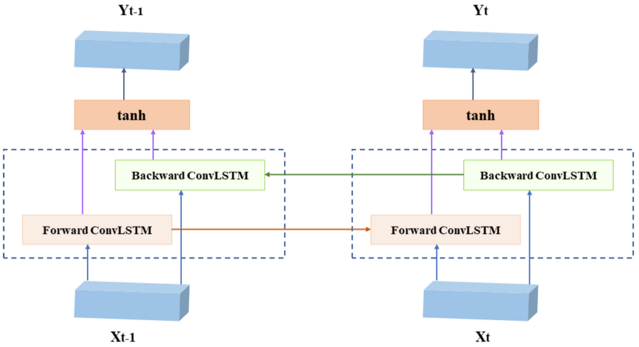


Fig. 6. The BConvLSTM used in this paper.

representation Y_t as follows:

$$\mathrm{Y}_t = \tanh(W_y^{\overrightarrow{H}} * \overrightarrow{H_t} + W_y^{\overleftarrow{H}} * \overleftarrow{H_t} + b), \tag{5}$$

where $W_y^{H\rightarrow}$ and $W_y^{H\leftarrow}$ are the weight matrixs, and b is the bias term. Moreover, tanh is the hyperbolic tangent which is utilized here to combine the output of both forward and backward states through a non-linear way.

Global Average Pooling Layer: Next, we use a standard average pooling layer for feature mapping instead of simply using the full connection layer. The average pooling layer gives each channel the actual category meaning which is more native to the convolution structure. What's more, there is no parameter to optimize so that can avoid overfitting.

2.3 Classification

Finally, we will feed the final ECG feature representation into a softmax layer to compute a probability distribution ŷ according to Eq. (6), and then the classification results can be obtained.

$$\hat{y} = softmax(z^i) = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_e}}, \tag{6}$$

where $\hat{y} \in \{0,1\}$, $z^{(i)}$ is the output value of the ith node, and C is the number of categories.

3 Experiment

In this section, we evaluate the performance of our proposed model by comparing it with some state-of-the-art methods.

3.1 Dataset Description

In this study, all our experiments are mainly conducted on the following two datasets. Dataset 1 is an intra-patient dataset that comes from the integration of three public databases. Dataset 2 is a large inter-patient dataset containing more than 10,000 subject records. Details of these datasets are given below.

3.1.1 Dataset 1

This dataset is obtained by extracting four ECG signals: AFIB, NSR, AFL, and VFIB from the following three public databases.

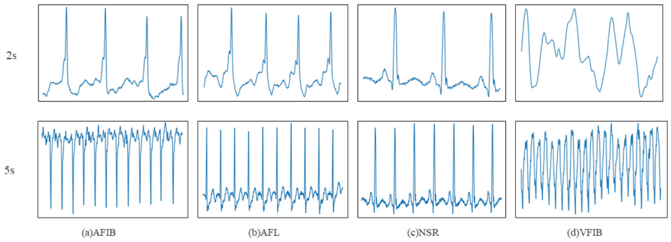


Fig. 7. ECG waveforms examples depicting (a) AFIB, (b) AFL, (c) NSR, and (d) VFIB of dataset 1.

- Creighton University ventricular tachyarrhythmia database (CUDB) : This database contains 35 eight-minute ECG signals of subjects who experienced episodes of sustained ventricular tachycardia, ventricular flutter, and ventricular fibrillation [27].
- MIT-BIH atrial fibrillation database (AFDB) : This database contains 25 long-term (10-h) ECG recordings of subjects with atrial fibrillation [28].
- MIT-BIH arrhythmia database (MITDB) : This database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects [29].

The frequency of ECG signals from CUDB and AFDB is 250Hz, while the frequency of ECG signals acquired from MITDB is 360Hz. First, we downsample the ECG signals from MITDB to 250Hz. Then, we use the Daubechies wavelet six [30] to denoise the signals and remove the baseline drift. Finally, we segment these denoised signals into two-second and five-second segments(shown in Fig. 7). Table 2 shows the details of Dataset1.

3.1.2 Dataset 2

This ECG dataset is collected by Chapman University and Shaoxing People's Hospital (Shaoxing Hospital Zhejiang University School of Medicine) [31]. It includes a large number of individual subjects more than 10,000 with 12-lead ECG signals sampled at a frequency of 500 Hz. In this database, there are 11 heart rhythms labelled by professional physicians: atrial flutter (AF), atrial fibrillation (AFIB), atrial tachycardia (AT), atrioventricular node reentrant tachycardia (AVNRT), atrioventricular reentrant tachycardia (AVRT), sinus irregularity (SI), sinus atrium to atrial wandering rhythm (SAAWR), sinus bradycardia (SB), sinus rhythm (SR), sinus tachycardia (SINT), and supraventricular tachycardia (SVT). The database comprises 10,646 subjects, and 12-lead ECG records are acquired over 10 seconds. Since some ECG recordings contain only zeros, and some channel values are missing, we use a total of 10,585 topics from this database. In addition, since the frequency range of normal ECG is from 0.5 Hz to 50 Hz, the Butterworth low pass filter [32] is used to remove the signal with a frequency above 50 Hz. The, the local polynomial regression smoother (LOESS) [33] is used to clear the effects of baseline wandering. Finally, the Non-Local Means (NLM) [34] technique is utilized to handle the remaining noise.

Due to the existence of severe class imbalance, we filter out four categories of arrhythmia signals (AT, AVNRT, AVRT, and SAAWR) whose sample sizes are too less. Thus,

TABLE 2
Distribution of Rhythm Categories on Dataset 1

Rhythms	Database	Total number of original 2-s segments	Total number of original 5-s segments
NSR	CUDB MITDB	902	361
AFIB	AFDB MITDB	18804	7407
AFL	AFDB MITDB	1840	736
VFIB	CUDB	163	65

TABLE 3
Distribution of Rhythm Categories on Dataset 2

Rhythms	Number of Samples	Age (Mean)	Number of Females	Number of Males
AF	438	71.14	182	256
AFIB	1780	73.35	739	1041
SI	397	34.88	175	222
SB	3888	58.33	1408	2480
SR	1825	54.37	1024	801
ST	1564	54.67	769	795
SVT	544	55.64	294	250
All	10436	57.48	4591	5845

a final dataset of 10,436 topics is obtained, and its related statistics are shown in Table 3.

3.2 Implementation Details

In our study, all the methods are implemented in Python 3.6.0 with Keras 2.2.2 and trained by Adam optimizer with a learning rate of 0.0001. We conduct the experiments on a server with Intel(R) Xeon(R) Gold 6230 CPU @ 2.10 GHz, 126 GB memory, and six GeForce RTX cards. We train each model by 10-fold cross-validation to enhance the generalization performance of models. Accuracy, Precision, Recall, and F1-score are used to evaluate the performance of our methods. What's more, we repeat every experiment 10 times and report the mean evaluation metrics for testing performance.

3.3 Result Analysis

3.3.1 Performance Comparison with Baselines

Experimental results on Dataset 1

We compare the performance of our proposed method against several representative baseline models on Dataset 1, the results are reported in Table 4. Desai et al. [35] adopt Recurrence Quantification Analysis features to classify ECG beats using ensemble classifiers. The ensemble classifiers include Decision Tree, Random Forest, and Rotation Forest. Acharya et al. [10] ranked the thirteen nonlinear features of ECG beats by ANOVA. And they applied K-Nearest Neighbor (KNN) and Decision Tree (DT) classifiers to realize the automatic classification of arrhythmias. Sree et al. [5] determined ECG segments to HOS and extracted 18 non-linear features from it. Then, they used a t-test to select significant features before classification. Acharya et al. [6] just used an eleven-layer CNN model to detect arrhythmia. Fujita et al. [36] used continuous wavelet transformation (CWT) for optimal extraction of necessary features before classification by CNN.

From the table, we can see that [5], [10], [35] are three machine learning based methods. Desai et al. [35] achieved an accuracy of 98.37% using a total of 3858 beats, and Acharya et al. [10] obtained 96.3% accuracy using 614526 beats. Sree et al. [5] achieved 93.94% accuracy using 21709 two-second ECG segments and 95.62% accuracy using 8683 five-second ECG segments. These methods must rely on experts to design and extract the characteristics of ECG signals, other potential information in the original signal is neglected. Although [10], [35] obtain relatively good results, their classification performance largely depends on the accuracy of QRS detection. In conclusion, the feature learning ability of the machine learning method is insufficient, and the dependence of feature selection on manual intervention also significantly limits its effectiveness and generalization. The studies [6], [36] try to use CNN to extract features automatically. However, the performance gains are poor. The main reason is that CNN lacks the learning ability of temporal dependency, which affects the model's overall performance. In contrast, we present a hidden attention layer that includes SE block and BConvLSTM to capture temporal dependencies in the ECG more efficiently. Thus, we achieve 99.2% accuracy using 21709 two-second ECG segments and 99.3% accuracy using 8683 five-second ECG segments which is obviously superior to other baselines. It's worth mentioning that most previous works directly extract features from the one-dimensional ECG signals, and we achieve better performance by using the two-dimensional images transformed by ECG signals. To some extent, it is proved that two-dimensional transformation can enhance the feature of ECG from another perspective.

Experimental results on Dataset 2

Dataset 2 is a large inter-patient dataset containing more than 10,000 subject records. To demonstrate the generalization of our method, we choose two other representative methods as baselines for performance comparison on Dataset 2. Yildirim et al. [37] used a single DNN model to classify one-dimensional ECG signals and obtained 80.04% of the F1 score. Baygin et al. [38] used a specifichomeomorphically irreducible tree (HIT) pattern and a maximum absolute pooling (MAP) decomposer to generate multilevel features. Then, a nonparametric Chi2 selector was used to select the most informative 1000 features; and SVM, to classify the chosen features. They finally obtained an F1 score of 84.01%.

Table 5 shows a detailed performance comparison. We finally achieve an F1 score of 87.6% which is 7.56% and 3.59% higher than the existing methods. This shows that our method also has good generalization ability on unseen data.

3.3.2 Effect of Various Data Preprocessing Methods

Our approach deploys a data preprocessing module to convert 1D ECG signals into 2D images, which are then fed into the model for feature extraction. To explore the impact of different data preprocessing methods on model

TABLE 4
Comparison of Some State-of-the-Art Study Methods on Dataset 1 *

Methods,Year	Database	Special Characteristics	ECG rhythms	Classifier	Performance
Desai et al. 2016[35]	MITDB AFDB CUDB	QRS detection performed 3858 ECG beats	AFIB AFL VFIB NSR	Decision Tree, Random Forest, and Rotation Forest	Accuracy=98.4%
Acharya et al. 2016[10]	MITDB AFDB CUDB	QRS detection performed 614526 ECG beats	AFIB AFL VFIB NSR	ANOVA ranking and decision tree	Accuracy=96.3%
Sree et al. 2021[5]	MITDB AFDB CUDB	No QRS detection performed 21709 2s ECG segment 8683 5s ECG segment	AFIB AFL VFIB NSR	HOS,Adasyn, Random Forest	2s:Accuracy=93.9% 5s:Accuracy=95.6%
Acharya et al. 2017[6]	MITDB AFDB CUDB	No QRS detection performed 21709 2s ECG segment 8683 5s ECG segment	AFIB AFL VFIB NSR	CNN	2s:Accuracy=92.5% 5s:Accuracy=94.9%
Fujita et al. 2019[36]	MITDB AFDB CUDB	No QRS detection performed 25459 2s ECG segment	AFIB AFL VFIB NSR	CNN	Accuracy=97.8%
Our Method	MITDB AFIB CUDB	No QRS detection performed 21709 2s ECG segment 8569 5s ECG segment	AFIB AFL VFIB NSR	HA-ResNet	**2s:Accuracy=99.2%** **5s:Accuracy=99.3%**

* The results in the table are extracted from the cited literatures.

performance, we take one-dimensional ECG segments of various lengths (2s, 5s, 10s) as the original input, and analyze the performance differences of our model when different two-dimensional transformation methods (RP, GAF, and MTF) are used in the preprocessing module.

The results are given in Table 6. We can see that HA-ResNet+RP and HA-ResNet+GAF outperform HA-ResNet +MTF significantly. Among them, HA-ResNet+RP achieves 96.0% F1 score which is the highest in the 2s segments. However, HA-ResNet+GAF achieves relatively higher performance on the 5s segments. This may be due to the fact that GAF employs an angle conversion to amplify the subtle numerical changes in the ECG signals, thereby better capturing fine-grained features in long ECG segments. However, under the inter-patient paradigm, ECG segments are collected from different individuals. While enhancing the key features related to classification tasks, GAF also amplifies irrelevant noises from individual differences in patients, thus greatly affecting the effectiveness of feature extraction. Therefore, as can be seen from the results on the 10s inter-patient dataset, the performance of HA-ResNet+GAF is significantly lower than that of HA-ResNet+RP. RP is obtained by calculating pairwise distances between the original data, which will not over-amplify the irrelevant noises. Thus, HA-ResNet+RP achieves the best overall performance under both inter-patient and intra-patient paradigms.

3.3.3 Benefits of Hidden Attention Module

To demonstrate the advantages of the HA-ResNet, we compare our model with its two variants. One is the ResNet, which is obtained by subtracting the SE block and BConvLSTM block from our model. Another is the SE-ResNet, which is obtained by removing the BConvLSTM block from our model.

Tables 7 and 8 show the performance of each model in 2s segments and 5s segments. As we can see, the F1 score of ResNet is 93.8% in 2s segments and 85.99% in 5s segments. Especially, the F1 score increases by 0.2% (94%-93.8%) in 2s segments and 2.02% (88.01%-85.99%) in 5s segments when

TABLE 5
Comparison of Some State-of-the-Art Study Methods on Dataset 2 *

Methods,Year	Dataset	Number of rhythms	Classifier	Precision(%)	Recall(%)	F1-Score(%)
Yildirim et al. 2020[37]	Dataset 2	7	DNN	80.3	80.2	80.0
Baygin et al. 2021[38]	Dataset 2	7	homeomorphically irreducible tree (HIT)	**90.2**	81.0	84.0
Our Method	Dataset 2	7	HA-ResNet+RP	87.6	**88.2**	**87.6**

∗ *The results in the table are extracted from the cited literatures.*

TABLE 6
The Results of Different Data Preprocessing Methods on Different Datasets

Dataset	Methods	Precisions (%)	Recall (%)	F1-Score (%)
Dataset 1 (2s)	HA-ResNet +MTF	86.4	82.2	84.2
	HA-ResNet +GAF	95.9	92.3	94.1
	HA-ResNet+RP	**96.2**	**95.4**	**96.0**
Dataset 1 (5s)	HA-ResNet +MTF	82.0	76.6	79.1
	HA-ResNet +GAF	**97.3**	**96.1**	**96.7**
	HA-ResNet+RP	95.4	92.3	93.7
Dataset 2 (10s)	HA-ResNet +MTF	78.3	74.8	74.7
	HA-ResNet +GAF	83.2	83.5	81.7
	HA-ResNet+RP	**87.6**	**88.2**	**87.6**

TABLE 7
Performance Comparison of HA-ResNet and its Modifications on 2S Segments

Methods	Class	Precision (%)	Recall(%)	F1-Score (%)
ResNet	AFIB	98.4	99.0	98.7
	AFL	90.0	85.3	87.6
	NSR	99.4	96.7	98.0
	VFIB	90.9	90.9	90.9
	Average	94.7	93.0	93.8
SE-ResNet	AFIB	98.4	99.0	98.7
	AFL	91.1	85.9	88.4
	NSR	98.3	97.8	98.1
	VFIB	90.9	90.9	90.9
	Average	94.7	93.4	94.0
HA-ResNet	AFIB	98.9	99.2	99.1
	AFL	92.5	90.5	91.5
	NSR	99.4	97.8	98.6
	VFIB	93.9	93.9	93.9
	Average	**96.2**	**95.4**	**96.0**

we add the SE block. It demonstrates that the SE block is effective for enhancing important features. Furthermore, compared with SE-ResNet, HA-ResNet obtains about 2% and 8.64% improvement in terms of F1 score on the 2s and 5s segments, respectively. It demonstrates that the proposed method is more competitive in mining the potential temporal dependencies of ECG signals.

Table 9 shows the performance of each model in the 10s segments. Since each ECG in this dataset comes from different patients, it increases the difficulty of classification. It can be seen that HA-ResNet can well extract the feature information related to categories, and obtain the performance that surpasses the other two models. In conclusion, the SE block and the BConvLSTM block can effectively enhance the model's feature capture ability and improve the generalization under the inter-patient paradigm.

TABLE 8
Performance Comparison of HA-ResNet and its Modifications on 5S Segments

Methods	Class	Precision (%)	Recall(%)	F1-Score (%)
ResNet	AFIB	98.1	95.6	96.8
	AFL	61.5	97.9	85.5
	NSR	96.3	72.4	78.8
	VFIB	100.0	88.7	92.9
	Average	86.5	88.1	86.0
SE-ResNet	AFIB	98.0	96.7	97.4
	AFL	68.9	87.9	77.3
	NSR	93.2	82.9	87.7
	VFIB	100.0	81.3	89.7
	Average	90.0	87.2	88.0
HA-ResNet	AFIB	99.3	99.1	99.2
	AFL	92.6	93.8	93.2
	NSR	97.3	98.6	97.9
	VFIB	100.0	92.9	96.3
	Average	**97.3**	**96.1**	**96.7**

TABLE 9
Performance Comparison of HA-ResNet and its Modifications on 10S Segments

Methods	Class	Precision(%)	Recall(%)	F1-socre(%)
ResNet	AF	19.0	18.2	18.6
	AFIB	72.1	77.0	74.5
	SB	95.5	97.4	96.7
	SI	64.3	22.5	33.3
	SR	83.1	88.5	85.7
	ST	88.5	87.9	88.2
	SVT	74.1	72.7	73.4
	average	82.9	83.5	82.8
SE-ResNet	AF	44.4	18.2	25.8
	AFIB	72.4	80.9	76.4
	SB	95.5	97.4	96.4
	SI	50.0	22.5	31.0
	SR	84.0	89.1	86.5
	ST	88.0	93.0	90.4
	SVT	77.8	76.4	77.1
	average	83.6	85.2	83.9
HA-ResNet	AF	45.2	31.1	36.8
	AFIB	76.2	86.5	81.1
	SB	96.7	98.5	97.6
	SI	75.0	45.0	56.3
	SR	91.5	88.5	90.0
	ST	90.1	92.4	91.2
	SVT	83.9	85.5	84.7
	average	**87.6**	**88.2**	**87.6**

4 Discussion and Conclusion

In this study, we developed a new hidden attention residual neural network to detect different types of arrhythmias. Our method gives full play to the advantages of ECG two-dimensional signal and uses the hidden attention residual model to extract the features that may be ignored by other models. We evaluate the proposed model on two datasets and achieve excellent results. The results prove that our

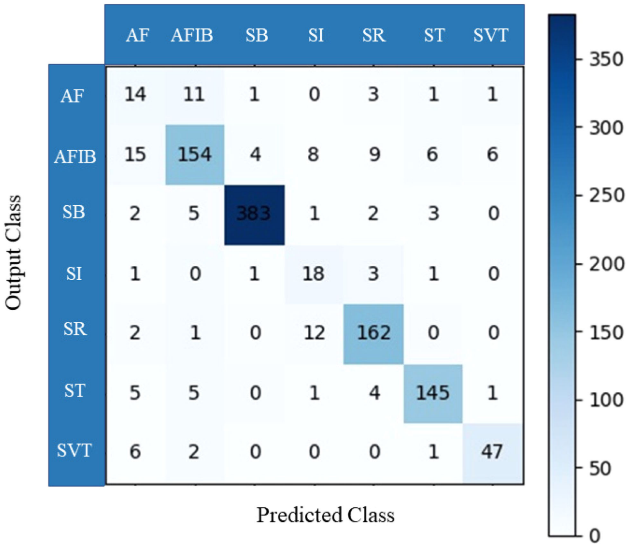


Fig. 8. The confusion matrix on dataset 2.

model can well summarize the invisible ECG signals and has good generalization performance.

However, in Dataset 2, due to the influence of unseen differences between patients, the performance of our method is not as good as that of Dataset 1. Fig. 8 shows the confusion matrix on Dataset 2. We can see that 18 samples of AF are misclassified as AFIB. 12 samples of SI are misclassified as SR. This is due to the fact that these arrhythmias have a greater similarity in ECG manifestations, which increases the difficulty of identification. On the other hand, the sample size of AF and SI is too small, so the model cannot learn enough effective features.

In the future, we will try to effectively fuse different types of features such as deep features and handcrafted features to further enhance the classification ability of the model, thereby improving the model's generalization ability to class-imbalanced datasets and inter-patient paradigm.

Acknowledgments

Yuxia Guan and Ying An are Co-First authors. This work was carried out in part using computing resources at the High Performance Computing Center of Central South University.

References

[1] F. Murat et al., "Exploring deep features and ECG attributes to detect cardiac rhythm classes," *Knowl.-Based Syst.*, vol. 232, 2021, Art. no. 107473.

[2] C. Antzelevitch and A. Burashnikov, "Overview of basic mechanisms of cardiac arrhythmia," *Cardiac Electrophysiol. Clin.*, vol. 3, no. 1, pp. 23–45, 2011.

[3] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat. Med.*, vol. 25, no. 1, pp. 65–69, 2019.

[4] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, "ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network," *Artif. Intell. Med.*, vol. 106, 2020, Art. no. 101856.

[5] V. Sree et al., "A novel machine learning framework for automated detection of arrhythmias in ecg segments," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 11, pp. 10 145–10 162, 2021.

[6] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Informat. Sci.*, vol. 405, pp. 81–90, 2017.

[7] E. K. Wang, X. Zhang, and L. Pan, "Automatic classification of cad ECG signals with SDAE and bidirectional long short-term network," *IEEE Access*, vol. 7, pp. 182 873–182 880, 2019.

[8] R. J. Martis, U. R. Acharya, H. Prasad, C. K. Chua, C. M. Lim, and J. S. Suri, "Application of higher order statistics for atrial arrhythmia classification," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 888–900, 2013.

[9] F. A. Elhaj, N. Salim, A. R. Harris, T. T. Swee, and T. Ahmed, "Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals," *Comput. Methods Programs Biomed.*, vol. 127, pp. 52–63, 2016.

[10] U. R. Acharya et al., "Automated characterization of arrhythmias using nonlinear features from tachycardia ECG beats," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2016, pp. 000 533–000 538.

[11] Z. Ahmad, K. Illanko, N. Khan, and D. Androutsos, "Human action recognition using convolutional neural network and depth sensor data," in *Proc. Int. Conf. Informat. Technol. Comput. Commun.*, 2019, pp. 1–5.

[12] S. Raghunath et al., "Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network," *Nat. Med.*, vol. 26, no. 6, pp. 886–891, 2020.

[13] J. Cheng et al., "Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 2, pp. 673–684, Feb. 2022.

[14] Y. Feng et al., "DCMN: Double core memory network for patient outcome prediction with multimodal data," in *Proc. IEEE Int. Conf. Data Mining*, 2019, pp. 200–209.

[15] J. Cheng et al., "COVID-19 mortality prediction in the intensive care unit with deep learning based on longitudinal chest X-rays and clinical data," *Eur. Radiol.*, vol. 32, no. 7, pp. 4446–4456, 2022.

[16] J. Cheng, J. Liu, H. Kuang, and J. Wang, "A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 1520–1532, Jun. 2022.

[17] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[18] G. Petmezas et al., "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomed. Signal Process. Control*, vol. 63, 2021, Art. no. 102194.

[19] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92 871–92 880, 2019.

[20] A. M. Alqudah, S. Qazan, L. Al-Ebbini, H. Alquran, and I. A. Qasmieh, "ECG heartbeat arrhythmias classification: A comparison study between different types of spectrum representation and convolutional neural networks architectures," *J. Ambient Intell. Humanized Comput.*, 2021, doi: 10.1007/s12652-021-03247-0.

[21] M. Naz, J. H. Shah, M. A. Khan, M. Sharif, M. Raza, and R. Damaševičius, "From ECG signals to images: A transformation based approach for deep learning," *PeerJ Comput. Sci.*, vol. 7, 2021, Art. no. e386.

[22] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convLSTM U-net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 406–415.

[23] J.-P. Eckmann et al., "Recurrence plots of dynamical systems," *World Sci. Ser. Nonlinear Sci. Ser. A*, vol. 16, pp. 441–446, 1995.

[24] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, 2015.

[25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[26] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Informat. Process. Syst.*, 2015, pp. 802–810.

[27] F. Nolle, F. Badura, J. Catlett, R. Bowser, and M. Sketch, "CREI-GARD, a new concept in computerized arrhythmia monitoring systems," *Comput. Cardiol.*, vol. 13, pp. 515–518, 1986.

[28] G. Moody, "A new method for detecting atrial fibrillation using RR intervals," *Comput. Cardiol.*, vol. 10, pp. 227–230, 1983.

[29] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[30] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digit. Signal Process.*, vol. 16, no. 3, pp. 275–287, 2006.

[31] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Sci. Data*, vol. 7, no. 1, pp. 1–8, 2020.

[32] S. Butterworth et al., "On the theory of filter amplifiers," *Wirel. Engineer*, vol. 7, no. 6, pp. 536–541, 1930.

[33] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting," *J. Amer. Statist. Assoc.*, vol. 83, no. 403, pp. 596–610, 1988.

[34] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[35] U. Desai, R. J. Martis, U. R. Acharya, C. G. Nayak, G. Seshikala, and R. Shetty K, "Diagnosis of multiclass tachycardia beats using recurrence quantification analysis and ensemble classifiers," *J. Mechanics Med. Biol.*, vol. 16, no. 01, 2016, Art. no. 1640005.

[36] H. Fujita and D. Cimr, "Decision support system for arrhythmia prediction using convolutional neural network structure without preprocessing," *Appl. Intell.*, vol. 49, no. 9, pp. 3383–3391, 2019.

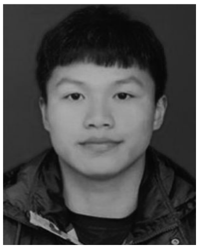
[37] O. Yildirim, M. Talo, E. J. Ciaccio, R. San Tan, and U. R. Acharya, "Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ECG records," *Comput. Methods Prog.s Biomed.*, vol. 197, 2020, Art. no. 105740.

[38] M. Baygin, T. Tuncer, S. Dogan, R.-S. Tan, and U. R. Acharya, "Automated arrhythmia detection with homeomorphically irreducible tree technique using more than 10,000 individual subject ecg records," *Informat. Sci.*, vol. 575, pp. 323–337, 2021.

Yuxia Guan received the BSc degree from the Wuhan University of Technology, China, in 2020. She is currently working toward the postgraduate degree in computer science with Central South University. Her currently research interests include medical data mining and deep learning.

Ying An received the BS degree in automatic control and the PhD degree in computer science from Central South University, China. He is an Associate Professor with Big Data Institute, Central South University. His research interests include Big Data analytics, machine learning, and its applications.

Jingrui Xu received the BSc degree from Henan University, China, in 2020. She is currently working toward the postgraduate degree in computer science with Central South University. Her currently research interests include bioinformatics and deep learning.



Ning Liu received the BS degree from the College of Computer Science and Information Technology, Guangxi Normal University, in 2018. He is currently working toward the MS degree with the School of Computer Science and Engineering, Central South University, Changsha, Hunan, China. His research interests include machine learning, deep learning, and medical image analysis.

Jianxin Wang (Senior Member, IEEE) received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is the vice dean and a professor with the School of Information Science and Engineering, Central South University, Changsha, Hunan, China. His current research interests include algorithm analysis and optimization, parameterized algorithms, bioinformatics, and computer network. He has published more than 150 papers in various international journals and refereed conferences.