Mathematical Biosciences and Engineering

MBE, 19(10): 9877–9894.
DOI: 10.3934/mbe.2022460
Received: 29 March 2022
Revised: 17 June 2022
Accepted: 30 June 2022
Published: 11 July 2022

http://www.aimspress.com/journal/MBE

Research article

A two-step method for paroxysmal atrial fibrillation event detection based on machine learning

Ya'nan Wang[1], Sen Liu[1], Haijun Jia[1], Xintao Deng[2], Chunpu Li[2], Aiguo Wang[2,*] and Cuiwei Yang[1,3,*]

[1] Center for Biomedical Engineering, School of Information Science and Technology, Fudan University, Shanghai 200433, China
[2] Department of Cardiology, Xinghua City People's Hospital, Jiangsu 225700, China
[3] Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai 200093, China

* **Correspondence:** Email: xhwag@163.com, yangcw@fudan.edu.cn.

Abstract: Detection of atrial fibrillation (AF) events is significant for early clinical diagnosis and appropriate intervention. However, in existing detection algorithms for paroxysmal AF (AFp), the location of AF starting and ending points in AFp is not concerned. To achieve an accurate identification of AFp events in the long-term dynamic electrocardiograms (ECGs), this paper proposes a two-step method based on machine learning. In the first step, based on features extracted from the calculated R-to-R intervals (RR intervals, the cycle of heart beat), the rhythm type of the ECG signal is first classified into three classes (AFp rhythm, persistent AF (AFf) rhythm, and non-atrial fibrillation (non-AF, N) rhythm) using support vector machine (SVM). In the second step, the starting and ending points for AF episodes of AFp rhythms predicted in the first step are further located based on heartbeat classification. By training a deep convolutional neural network with phased training, the segmented beats of AFp rhythms are divided into AF beats and non-AF beats to determine the beginning and end of any AF episode. The proposed two-step method is trained and tested on the 4th China Physiological Signal Challenge 2021 databases. A final score U of 1.9310 is obtained on the unpublished test set maintained by the challenge organizers, which demonstrates the advantage of the two-step method in AFp event detection. The work is useful for assessing AF burden index for AFp patients.

Keywords: atrial fibrillation event detection; machine learning; phased training; two-step method;

RR intervals

1. Introduction

Atrial fibrillation (AF) is a common cardiovascular disease characterized by tachyarrhythmia, which can lead to increased rates of stroke, heart failure, and other cardiovascular diseases [1]. Therefore, early detection of AF episodes is crucial for patients to access to treatment as early as possible.

According to the duration of AF episodes, it can be divided into three grades: (1) paroxysmal AF (AFp), (2) persistent AF (AFf), and (3) permanent AF. AFp has the characteristics of short duration of attack, high recurrence rate, difficult to capture on electrocardiogram (ECG) [2]. With the prolongation of the patient's course of disease, it will gradually evolve into AFf or permanent AF, threatening the patient's life [3]. Therefore, the detection of AFp events is very significant for patients to receive effective treatment in time. ECG is a convenient, rapid and non-invasive detection method, which is widely used in clinical practice [4,5]. The AF episode is generally associated with irregular RR intervals when atrioventricular conduction is not impaired [6]. In addition, the absence of distinct repeating P waves and the presence of f waves help to identify the AF rhythm [6]. However, for AFp detection, conventional 12-lead ECG has great limitation, while dynamic ECG has high detectable rate and accuracy (Acc) [7].

At present, many automatic detection algorithms have been proposed for AFp detection based on dynamic ECG. Liu et al. [8] proposes a detection method for AFp based on covariance descriptor and nuclear sparse coding. Petrenas et al. [9] designs a detector consisting of RR interval irregularity characterization, P-wave absence, f-wave presence, and noise level, which are used for fuzzy logic classification to detect brief AF episodes of AFp. Ganapathy et al. [10] proposes an automatic detection algorithm for AFp based on co-occurrence matrices computed from RR intervals. Xin et al. [11] applies a multi-scale wavelet analysis to extract features of heart rate variability (HRV), and then completes AFp detection based on a SVM classifier.

In addition to accurate detection of AFp, precise location of starting and ending points for AF episodes is also crucial for an effective AFp management, which provides more diagnosis and treatment information for doctors to implement appropriate medical intervention [12]. As such, AF burden associated with risk of stroke can be defined and measured better during a long-term AF monitoring period instead of device-detected AF duration as measured from cardiac implantable electronic devices (CIEDs) [13]. Previous algorithms aim at the detection of AFp, but fail to locate the starting and ending points for AF episodes of AFp. Recently, the 4th China Physiological Signal Challenge 2021 (CPSC 2021) [14] has provided two large public databases and an undisclosed test database for participants to implement algorithms for AFp event detection. We take this opportunity to propose an efficient two-step method that meet challenge rules for the starting and ending point location of AF episodes in AFp rhythm. In the first step, we divide all records with unequal length into (1) non-atrial fibrillation (non-AF, N) rhythm, (2) AFp rhythm and (3) AFf rhythm based on machine learning. In the second step, we locate the starting and ending points for AF episodes of AFp rhythm identified in the first step based on a convolutional neural network (CNN).

Tow-step detection allows us to select the most appropriate methods to achieve the goals in each step. As such, traditional feature extraction combines deep learning to maximize the usage of

beneficial information in different steps. In the first step, we directly extract artificial features from unequal-length ECG records based on RR intervals reflecting ventricular activations, advisably evading the requirement for fixed-length records during feature extraction by sliding window. In the second step, a CNN is applied to judge whether the single target beat constitutes AF rhythms or not, which is essentially a heartbeat binary classification problem. Besides, due to large amount of heartbeat segments are available from the public CPSC2021 databases, we train the CNN in multiple phases and choose different learning rate according to training situation as well as the size of the training set, thereby avoiding nonconvergence of model loss when model trained on a large dataset.

To summarize, the main contributions of this paper are listed as follows:

- We propose a two-step method to detect AF episodes, in which the SVM classifier is for classification and the 1-dimensional (1-d) CNN is for location.
- It is a method based on both ventricular responses and atrial activities that is suitable for analyzing unequal-length ECG records.
- We propose the phased training fashion that adjusts learning rate and the size of the training set to train the deep model based on large data.

2. Material and method

2.1. ECG dataset

CPSC 2021 [14] provides two public databases named as dataset I and dataset II, and an unpublished test set where partial data is from the same source as two public databases. Both public databases contain a large number of variable-length ECG record fragments, which are extracted from lead I and lead II of long-term dynamic ECG signals. The dataset I includes 730 records, which is collected from 54 subjects. The dataset II is collected from 51 subjects, including 706 records. There are three rhythm types, namely, non-AF rhythm, AFp rhythm, and AFf rhythm. Table 1 shows the amount distribution of rhythms in public databases. The duration of records ranges from 0.14 to 411.11 minutes, and average duration is 20.33 minutes. In these records, the original sampling rate is 200 Hz, and both non-AF episodes and AF episodes are not shorter than continuous 5 beats.

Table 1. The distribution of rhythms in CPSC 2021 public databases.

Rhythm distribution	N	AFp	AFf	Subject	Sum
Dataset I	481	96	153	0–53	730
Dataset II	251	133	322	54–104	706
Sum	732	229	475	–	1436

We use records from lead II in our experiment. For rhythm classification in the first step, we divide records from dataset I and dataset II into a rhythm training set and a rhythm test set. The amount distribution of rhythms in subsets is shown in Table 2, notably, the distribution in number of the rhythm training set is relatively balanced. In addition, rhythms belonging to the same class come from different subjects in the training rhythm set and the test rhythm set.

For heartbeat classification in the second step, we refer to heartbeats constituting AF episodes as AF beats, and heartbeats constituting non-AF episodes as N beats. Finally, a beat training set, a beat validation set, and a beat test set are generated from the public CPSC 2021 databases. For heartbeat

distribution in Table 3, the amount ratio of N beats and AF beats is about 1:1 in each subset. A portion of the beat training set and the whole beat training set are used under phased training fashion, which will be introduced in Section 4. In order to obtain a model with outstanding performance and convictive test results, heartbeat segments in each subset come from different records.

Table 2. The amount distribution of rhythms in subsets.

Rhythm distribution	N	AFp	AFf	Subject	Sum
Training rhythms from Dataset I	0	77	0	3, 25, 31, 32, 39, 40, 48,	77
Training rhythms from Dataset II	251	133	322	54–96, 99, 100, 102–104	706
Test rhythms from Dataset I	481	0	153	0–53	634
Test rhythms from Dataset II	0	19	0	97, 98, 101	19
Sum	732	229	475	–	1,436
The rhythm training set	251	210	322	–	783
The rhythm test set	481	19	153	–	653

Table 3. The amount distribution of heartbeats in subsets.

Heartbeat distribution	Non-AF beats	AF beats	Sum
A portion of the beat training data	372,616	368,836	741,452
The whole beat training set	629,425	627,188	1,256,613
The beat validation test set	32,257	31,134	63,391
The beat test set	104,381	15,615	119,996
Sum	1,138,679	1,042,773	2,181,452

2.2. Two-step detection

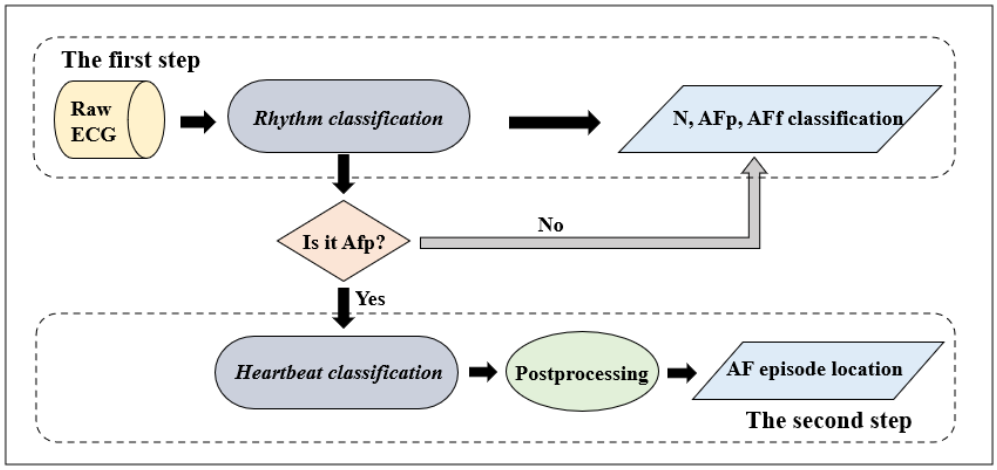
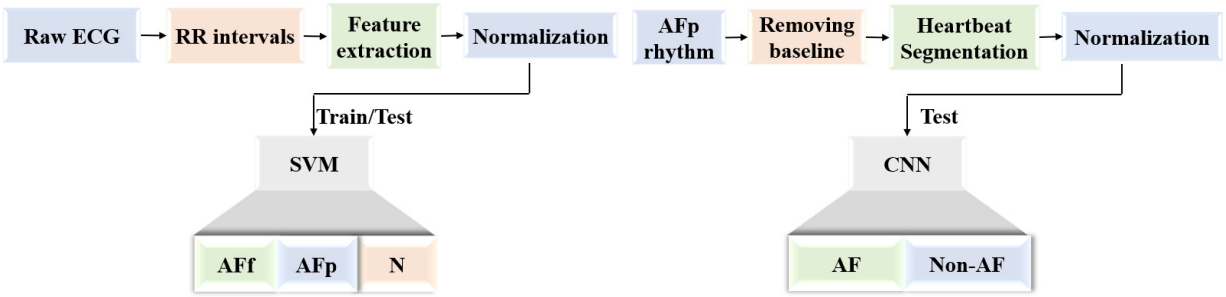


Figure 1. The flowchart of AFp event detection, including classification and location.

For AFp event detection, an effective method is first to determine whether the record is AFp rhythm, and then to locate the starting and ending points for AF episodes. Therefore, we divide the process of AFp event detection into two steps, namely, rhythm classification and heartbeat

classification. In each step, we choose the most appropriate algorithm and optimize accordingly. For rhythm classification in the first step, all records are classified into three rhythm types, including non-AF rhythm, AFf rhythm, and AFp rhythm. A traditional machine learning method is used to distinguish the rhythm type of each record. For heartbeat classification in the second step, a deep learning-based method classifies beats of AFp rhythms predicted in the first step into non-AF beats or AF beats. And then, a median filter postprocesses the results of heartbeat classification, which can effectively reduce false detection of AF beats. Finally, the starting and ending points of AF episodes are the location of changes between beat types. Figure 1 shows the flowchart of AFp event detection.



(a) (b)

Figure 2. The process of the two-step method. (a) The process of rhythm classification; (b) The process of heartbeat classification.

3. Rhythm classification

The rhythm classification is based on features calculated by RR intervals and classifier selection. Each record is taken as a single rhythm, either non-AF rhythm or AFp rhythm, or AFf rhythm. The process consists of three parts: (1) feature extraction, (2) feature normalization and (3) classification, as shown in Figure 2(a). For feature extraction, we extract representative features of each record. For feature normalization, each feature is processed by min-max normalization. For classification, a support vector machine (SVM) classifier is trained by grid search and 5-fold cross validation on the rhythm training set, and is tested on the rhythm test set. In order to show existing and representative methods for rhythm classification visually, we introduce these methods in detail besides our method and our optimized method in Table A1 (in Appendix).

3.1. Feature extraction

AF rhythm has obvious irregular RR intervals and analysis based on ventricular responses is more robust for the noise [15]. Thus, the rhythm type is judged according to characteristics of RR intervals. We extract features in time domain, frequency domain, and nonlinear domain, respectively. The features are extracted from variable-length records, but the number of features is fixed for each record, which meeting the requirement of the classifier with fixed input length. Hence, we do not need to crop records, nor do we need to be based on a fixed number of RR interval sequences.

- Time domain features: The features include average RR intervals (avgRR), percentage of differences greater than 50ms between RR intervals (pRR50), number of differences greater than

50ms between RR intervals (RR50), standard deviation of RR intervals (SDRR) and root mean square of successive differences of RR intervals (RMSSD), etc.

- Frequency domain features: The RR intervals are first resampled to 7 Hz. A frequency power spectrum is estimated using welch power spectral density estimate (hanning window; 4096 points; 50% overlap). The calculated frequency domain features include normalized low-frequency power (LFnorm), normalized high-frequency power (HFnorm), as well as ratio of low-frequency power and high-frequency power (rLH), etc.
- Non-linear features: The features are obtained based on the analysis of Lorenz plot. The features include standard deviation for the transverse direction (sd1), standard deviation for the longitudinal direction (sd2), cardiac sympathetic index (csi), cardiac vagal index (cvi), etc.

3.2. Feature normalization

The convergence of a classifier will slow down when the classifier is trained under different dimension of features, even to have a negative influence on the classifier's performance [16]. To solve this problem, we adopt min-max normalization for the rhythm subsets. It is worth noting that the min and max value of the rhythm training set are also used to normalize the rhythm test data.

3.3. Classification

The SVM classifier [17] that proposed by Boser has a strong generalization ability, which can convert low-dimensional data that are linearly indivisible to high-dimensional codes by kernel mapping, and mapping form depends on the form of kernel function. The process can be summarized as a convex optimization problem with optimal solution [18,19]. In our experiment, we choose a SVM classifier with a radial basis function (RBF) kernel to automatically identify records' rhythm classes. Optimal parameters are determined by 5-fold cross validation and grid search on the rhythm training data.

4. Heartbeat classification

Heartbeat classification is the basis of detection for various arrhythmias [18,20]. A single beat is short enough for 1-d CNN to automatically extract deep features based on its morphological characteristics. Considering that the aim is to locate the starting and ending points of AF episodes by classifying beats constituting AFp rhythms, we label beats either AF beats or non-AF beats, rather than labeling beats in term of the Association for the Advancement of Medical Instrumentation (AAMI) standard [21]. Therefore, when model tested on the beat test set, the variety of predicted heartbeat types between AF beats and non-AF beats can indicate the starting and ending points of AF episodes. The process of heartbeat classification is composed of five parts, respectively, removing baseline, heartbeat segmentation, heartbeat normalization, deep learning-based model building, as shown in Figure 2(b). Then, a median filter corrects misjudged heartbeats to a certain extent as a post-processor. The phase training fashion is also introduced in this section.

There are other outstanding algorithms for AF episode location by new ideas that have innovation and value whether in preprocessing or in model structure or in overall location process, which are provided by other participants in CPSC 2021 and have not been published.

4.1. Removing baseline

Considering that ECG signals in clinical contain various of noise, we only filter baseline drift and adopt no other denoising approaches in order to enhance the robustness to the noise of the proposed method. As the method adopted by Chazal [18] and Jiang [20], a 200 ms median filter first removes P waves and QRS complexes of an original signal, and then a 600ms median removes T waves, so that we can get the baseline drift signal. Finally, the ECG signal without baseline drift is obtained by subtracting the baseline drift signal from the original ECG record.

4.2. Heartbeat segmentation

The original sampling rate of ECG signals from CPSC2021 databases is used as the standard sampling rate, which is 200 Hz. The R-labels provided by CPSC 2021 databases is used as the fiducial R-points. Each segment is a single heartbeat with 200 sampling, which consists of 60 points in front of the R-point and 139 points after the R-point. The proportion of sampling points before and after the R-point is close to 3:7. Each heartbeat segment contains a complete P wave, a QRS complex, and a T wave if various waves exist.

4.3. Heartbeat normalization

The amplitude gains of ECG signals that are collected by different ECG monitoring devices have an obvious difference. Besides, the normalization of input data can contribute convergence and improve generalization of a model [22]. Thus, it is necessary to normalize data to a uniform scale. By Eq (1), the amplitude of heartbeat segments is mapped into the range of 0–1.

$$F_o = \frac{F - F_{\min}}{F_{\max} - F_{\min} + K} \qquad K = e - 06 \tag{1}$$

where F is a feature matrix; $Fmin$ and $Fmax$ are matrices, which are combined by minimum and maximum vectors from features, respectively; K is set to e-06 to preserve the divisor from being zero.

4.4. Model structure

A CNN can automatically extract deep features from ECG signals, and is an end-to-end fashion that combines feature extraction and classification [23]. We build a 1-d model in Figure 3, whose deepening fashion borrows that of VGG13 network [24]. The whole model consists of an automatic deep feature extractor and a classifier. The feature extractor is composed of convolutional layers, maximum pooling layers, and dropout layers. The preprocessed heartbeat segments are as the input of the feature extractor directly. Every two stacked convolutional layers have the same number filters, and are followed by a maximum pooling layer and a dropout layer to form a convolutional unit. Finally, five convolutional units are stacked in series to constitute the feature extractor. As the network deepens, the number of convolutional layers' filters in the five units increases or stays unchanged, and is 64, 128, 256, 512 and 512 in order. The kernel size and the stride of convolutional layers are all 3 and 1, respectively. Maximum pooling layers with step size 2 and kernel size 2 halve

the size of feature maps. For dropout layers, the probability of discarding neurons is set to 0.3 in order to reduce the risk of model overfitting. The rectified linear unit (RELU) activation function in convolutional layers is used for nonlinear mapping.

The classifier is composed of a flattening layer, three fully connected layer, and three dropout layers. Firstly, a flattening layer expands the feature matrix into a feature vector. Then, fully connected layers and dropout layers are stacked alternately. The probability of discarding neurons in the classifier's dropout layers is set to 0.5. The number of neurons of fully connected layers is 256, 128 and 1, respectively. The last fully connected layer is as the output layer, which only has one neuron and uses the sigmoid activation function for binary classification. The output of the classifier is the probability that the target beat is an AF beat. The feature extractor and classifier are connected for end-to-end training. Table 4 shows the hyperparameters of our model.

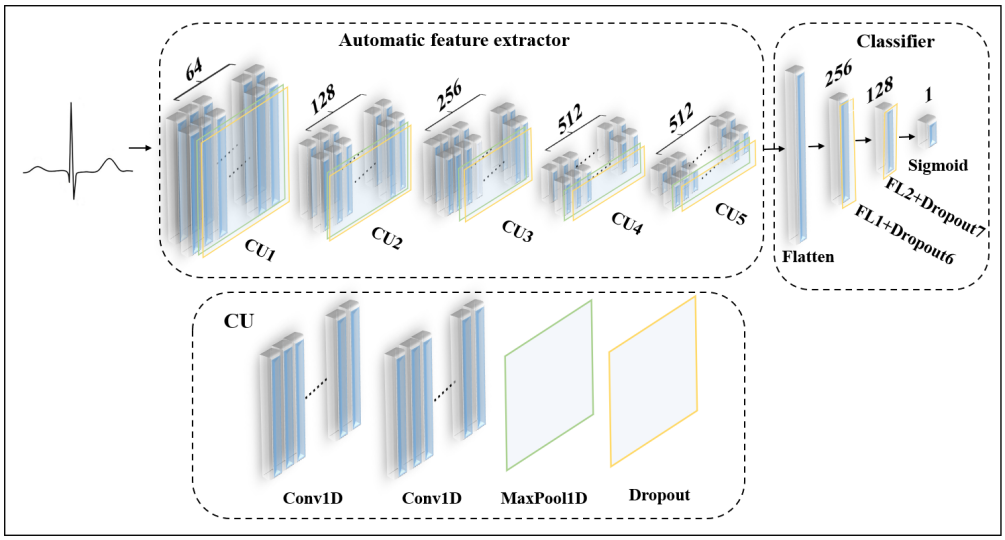


Figure 3. The model structure for heartbeat classification. Conv1D, MaxPool1D, CU and FL represent 1-D CNN layer, 1-D maximum pooling layer, convolution unit and fully connected layer, respectively.

Table 4. Parameters for the network.

Layer	Output	Kernel	Stride	Padding	Total parameter	FLOPs
Input	(200, 1)	–	–	–		
Convolutional unit 1	(100, 64)	3	1	same		
Convolutional unit 2	(50, 128)	3	1	same		
Convolutional unit 3	(25, 256)	3	1	same		
Convolutional unit 4	(12, 512)	3	1	same		
Convolutional unit 5	(6, 512)	3	1	same		
Flatten	(3072,)	–	–	–	3, 956, 289	0.148 G
FL 1	(256,)	–	–	–		
Dropout 6	(256,)	0.5	–	–		
FL 2	(128,)	–	–	–		
Dropout 7	(128,)	0.5	–	–		
Sigmoid	(1,)	–	–	–		

4.5. Phased training

The beat training set is so large that model training needs take a lot of time. In order to monitor training process and avoid failing to train, we propose a phased training fashion, and take every 300 epochs as a training phase. In addition, according to the situation of model convergence, we change the amount of training data or the learning rate in each phase. Figure 4 shows the operation. The model is trained under four phases, a total of 1200 epochs. We save the beat model with the highest F1 score on the beat validation training set in the fourth phase.

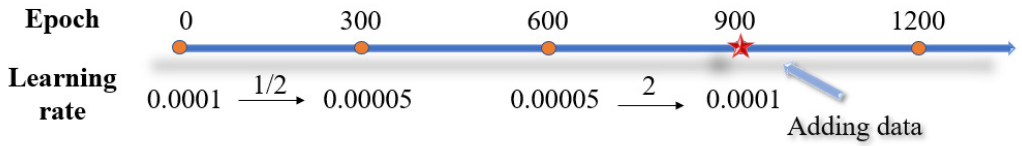


Figure 4. The variable learning rate in each phase.

4.5.1. Variable size of training set

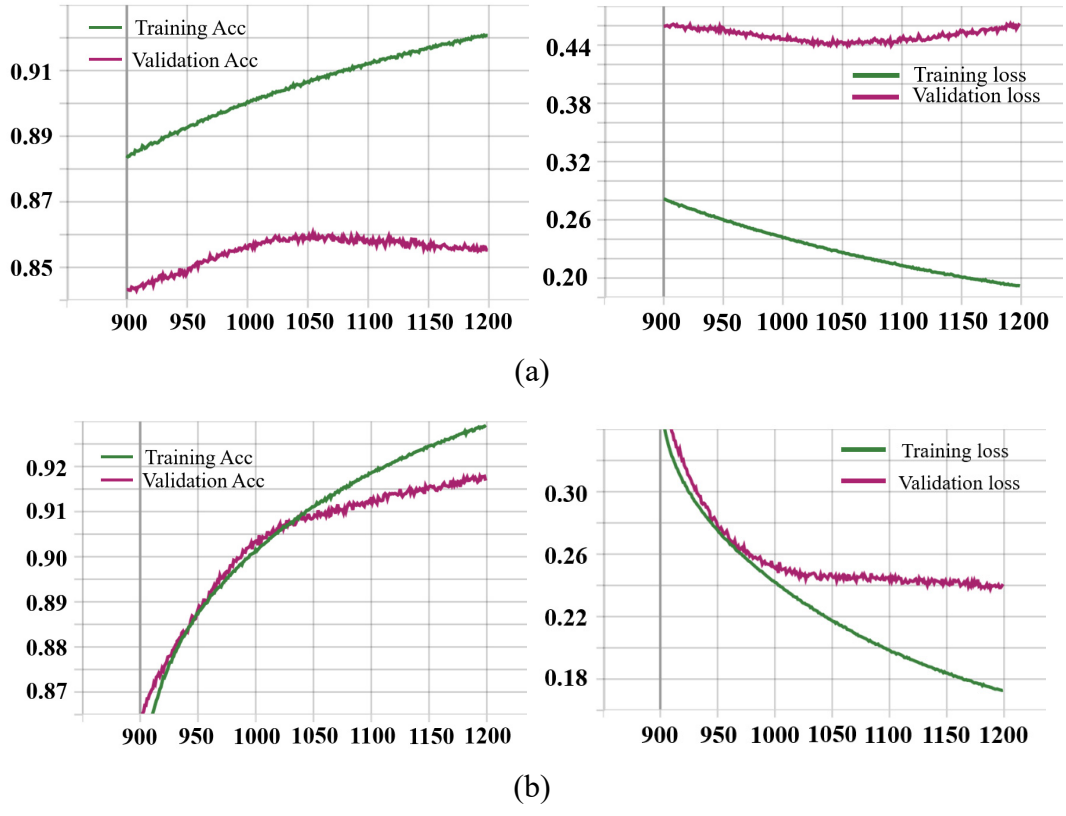


Figure 5. The training and validation Acc curves and loss curves during 900–1200 epochs. (a) on a portion of the beat training set; (b) after data expansion.

A large training data benefits the model performance, but may make training loss not easy to

converge, and increase the possibility of training failure [25]. Therefore, we first train a model on a portion of the beat training data until convergence, and then continue train it on the whole beat training set where abundant new data are added. The positive effect on the model performance can be observed clearly when the beat training set is extended. To the authors' knowledge, the trainable parameters are randomly initialized in the first training phase, and then, in the phase where training data are added, the optimized parameters are transferred into initial parameters in the current phase. The optimized initial values of trainable parameters can reduce the possibility of the model falling into a local optimal solution [26].

In our experience, a portion of the beat training data first make the model has the ability to distinguish different classes. As shown in Figure 5(a), after more than 1000 epochs, the loss of the beat validation set presents an upward trend, and the Acc of the beat validation set is in a significant downward trend. Next, the optimized parameters as the initial parameters are trained on the whole training set. Therefore, we expand the beat training set at the beginning of the fourth phase. That is, we fine-tuned the model obtained at the end of the third phase on the whole beat training set. Figure 5(b) shows the loss and Acc curves acquired after data expansion. A large number of new samples make the model learn new discriminative features quickly, and improve the model performance rapidly. At more than 1000 epochs, the validation loss is basically stable, and the validation Acc maintains above 90%. Finally, we obtain the best model at the 1043rd epoch, which has the highest F1 score on the beat validation set during the fourth phase.

4.5.2. Variable learning rate

The model is difficult to converge under a large learning rate, while a small learning rate tends to make the model fall into a local minimum [27]. Thus, it is necessary to adjust the learning rate based on the situation of model training. We use Adam optimizer to realize back propagation of training loss on the beat training set. Each batch contains 256 heartbeat segments. The model runs on GPU 3090 and Ubuntu system. We determine the learning rate at the beginning of each phase according the training loss and validation loss, so that the model can be well trained on the large beat training set.

In the first phase, the learning rate of Adam optimizer is 0.0001, and then we change into one second of the original at the beginning of the second phase and keep the learning rate constant until the beginning of the fourth phase. By observing the loss curve of the beat validation set, there is room for convergence. Therefore, we change the value of the learning rate into 0.0001 back at the beginning of the fourth phase to speed up the model's convergence.

4.6. Postprocessing

According to CPSC 2021, the number of continuous beats that constitute AF episodes or non-AF episodes is no less than 5. Therefore, we use a median filter with a window accommodating five beats to correct the predicted types by the model for heartbeat classification. Under postprocessing, the predicted result of the target beat is replaced by the median of the target and its four surrounding beats. The process corrects the outliers of the predicted results, and reduce the misdiagnosis of AF episodes. If the situation that the first beat or the last beat of an AFp rhythm is an AF beat is not considered, the starting points of AF episodes are the locations of R-points of AF beats before which

the beat is a non-AF beat, and ending points are the locations of R-points of AF beats after which the beat is a non-AF beat.

In the 1-d CNN training process, all rhythms are firstly segmented so that we can collect AF beats and non-AF beats, and then the 1-d CNN is trained to predict the types of heartbeats. During the test process for a single record, the rhythm type is predicted in the first step. The outputs of starting points and ending points of AF were [] (the [] is the output, representing the AF episode is none) in non-AF rhythm and [1, value equal to the length of the record] (1 represents the index of the first heartbeat) in AFf rhythm. In the second step, heartbeats are classified beat by beat for the record predicted as AFp rhythm in the first step, and the indexes of beats are also retained. After postprocessing by median filter, the location of AF episodes is the indexes of beats whose types change from AF beat to non-AF beat and vice versa.

5. Result

5.1. Assessment metrics

5.1.1. Indicators for classification

Four indicators, including Acc, sensibility, (Sen), positive predictive value (PPV) and F1 score, are used to evaluate the performance of our algorithms that are applied for rhythm classification and heartbeat classification. These indicators can be expressed by false positive (FP), true positive (TP), false negative (FN) and true negative (TN). The FP refers to the number of positive samples predicted to be negative, and the TP refers to the number of positive samples that are predicted correctly. The FN means the number of negative samples predicted to be positive, and the TN is the number of negative samples predicted to be correct. The four indicators are defined in Eqs (2)–(5), respectively.

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \tag{2}$$

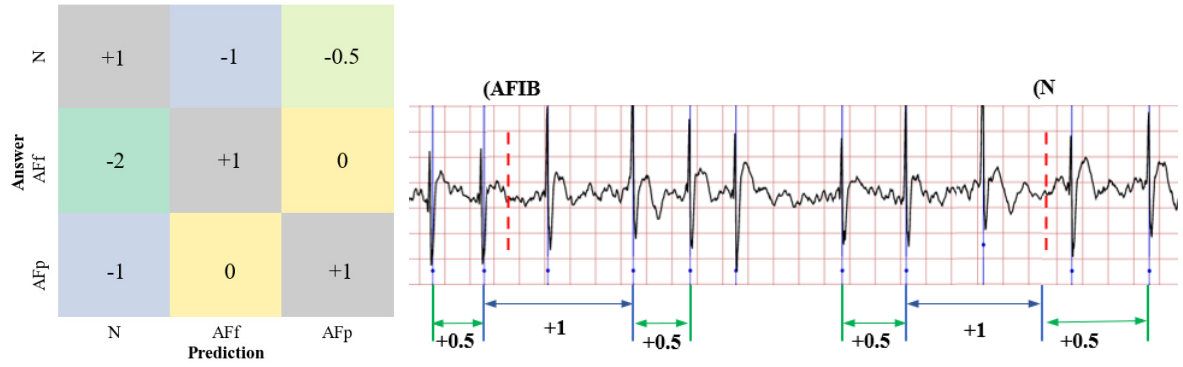
$$Sen = \frac{TP}{TP + FN} \tag{3}$$

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

$$F1 = \frac{(1+\beta)^2 \times Sen \times PPV}{\beta^2 \times Sen + PPV} \tag{5}$$

The overall Acc is the percentage of the samples that are accurately predicted. The Sen reflects the rate of missed diagnosis of a certain class, and the PPV presents the rate of misdiagnosis of a certain class. The F1 score provides a systematic evaluation of the algorithms' performance on an unbalanced test set. The β in Eq (5) is usually set to 1, therefore, the F1 score is the harmonic mean of Sen and PPV.

5.1.2. Indicators for AFp event detection



(a) (b)

Figure 6. (a) The scoring matrix of *Ur*. (b) The scoring criteria of *Ue*.

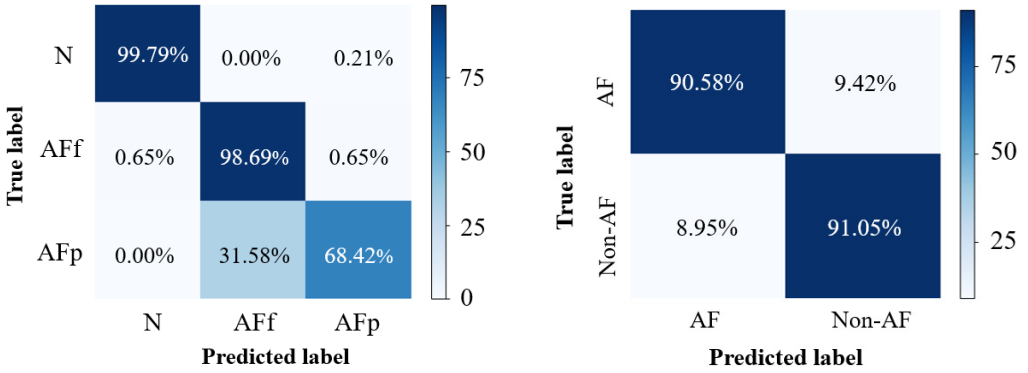
The indicators for AFp event detection are defined by CPSC 2021. *Ur* is the score of rhythm detection, and *Ue* is the score of the location of starting and ending points for AF episodes of AFp rhythms. The scoring criteria of *Ur* and *Ue* is shown in Figure 6. The score of each record is the sum of *Ur* and weighted *Ue*, and the final score *U* is the mean score of all records in a test set, which is descripted in Eq (6).

$$U = \frac{1}{N}\sum_{i=1}^{N}\left(Ur_i + \frac{Ma_i}{\max\{Mr_i, Ma_i\}} \times Ue_i\right) \tag{6}$$

where N is the number of records in a test set; Ma is the number of true AF episodes in each record; Mr is the number of predicted AF episodes in each record.

5.2. Test result

The realization of rhythm classification algorithm and heartbeat classification algorithm is the premise of starting and ending point detection for AF episodes. In Figure 7, two confusion matrices are the results of rhythm classification and heartbeat classification on their respective test sets. Table 5 shows calculated assessment metrics. Both algorithms have good classification performance, and achieve the overall Acc of 98.62% and 90.81%, respectively. However, there is room for optimization to distinguish AFf rhythms from AFp rhythms.



(a) (b)

Figure 7. Confusion matrices in the two steps. (a) The confusion matrix of rhythm classification. (b) The confusion matrix of heartbeat classification.

Table 5. Metrics of rhythm classification and heartbeat classification.

Algorithm	Class	Metric				Overall
		Sen	PPV	F1	Acc	Acc
Rhythm classification	N	0.998	0.998	0.997	0.998	0.986
	AFp	0.684	0.867	0.988	0.765	
	AFf	0.987	0.962	0.988	0.974	
Heartbeat classification	Non- AF beat	0.906	0.913	0.908	0.909	0.908
	AF beat	0.910	0.903	0.908	0.907	

We are honored to participate in CPSC 2021 and test our proposed two-step method on the unpublished test set. In the first step, the starting and ending points are none for non-AF rhythms, and for AFf rhythms, the starting and ending points are respectively the position of R points of the first beat and the last beat. If the target record is predicted as an AFp rhythm, the deep learning-based model classifies beats of the AFp rhythm in the second step, and the starting and ending points are located after postprocessing. Finally, our proposed two-step method achieves a final score U of 1.9310 on the unpublished test set and ranks the eighth among all competing algorithms in CPSC 2021.

6. Discussion

6.1. Comparison between two-step method and one-stem method

Comparing just one-stem algorithm for heartbeat classification to realize AF episode location, we discuss the advantages of the two-step method. For the two-step method, records are classified into three rhythm types based on ventricular responses in the first step. The SVM classifier can achieve high accuracy in classifying the records with variable lengths. For the records predicted as non-AF rhythms, the location of AF starting and ending points in non-AF rhythms is None. For those

predicted as AFf rhythms, the starting point of AF is 1, and the end point is the value equal to the length of the corresponding record. In the second step, the records that are predicted as AFf and non-AF rhythms in the first step do not participate in the second step. The 1-d CNN only analyzes AFp rhythms obtained in the first step beat by beat.

If we just propose the algorithm for heartbeat classification to realize AF episode location. The 1-d CNN analyzes all non-AF, AFf and AFp rhythms. When AF beats and non-AF beats are classified beat by beat, false positives and false negatives may occur in non-AF and AFf rhythms respectively. Therefore, the accuracy of location will be lower than that of the two-step method.

6.2. Effect of feature selection

Table 6. Feature groups after feature selection.

Feature group	Num	Feature description	Symbol
Time domain	1	Percentage of differences greater than 20ms between RR intervals.	pRR20
	2	Median absolute values of successive differences between RR intervals.	mavsd
	3	Ratio of the standard deviation divided by the mean of RR intervals.	rsdm
	4	Maximum heart rate.	maxHR
	5	Minimum heart rate.	minHR
Frequency domain	6	Low-frequency power.	LF
	7	Normalized high-frequency power.	HFnorm
Non-linear domain	8	Modified cardiac sympathetic Index.	mdcsi

The classifier for rhythm classification has room for optimization. Some redundant features negatively affect the performance of rhythm classification. Feature selection is a valid approach to abandon useless features. Thus, we apply genetic algorithm to select features to optimize the rhythm classifier. A SVM classifier is used as a prediction classifier to select the optimal features in the way of 5-fold cross validation. Finally, we get eight features, which are introduced in Table 6. The selected features represent records effectively, and the optimized SVM classifier has a good performance for all kinds of rhythms. Figure 8 shows the confusion matrix of the optimized SVM classifier on the rhythm test set.

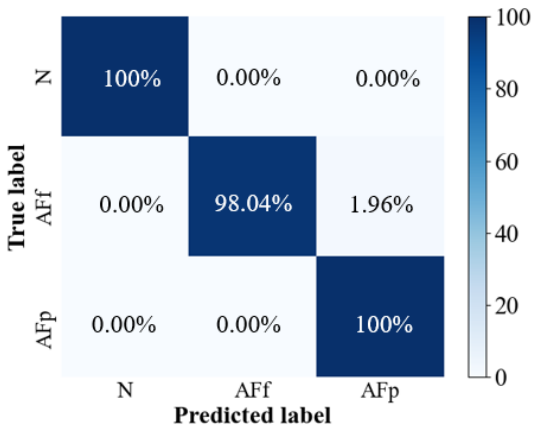


Figure 8. The confusion matrix of optimized SVM classifier on the rhythm test set.

6.3. Effect of classifiers for rhythm classification

Table 7. The comparation of different classifiers before and after feature selection.

Classifier	Before feature selection			After feature selection		
	Sen	PPV	Acc	Sen	PPV	Acc
LR	0.985	0.845	0.974	0.989	0.871	0.982
Bayes	1	1	1	1	1	1
KNN	0.961	0.784	0.946	0.988	0.869	0.980
SVM	0.890	0.942	0.986	0.990	0.920	0.991
DT	1	1	1	1	1	1
RF	0.999	0.983	0.998	1	1	1

We also compare the performance of other common classifiers before and after feature selection, including logistic regression (LR) classifier, naive Bayes (Bayes) classifier, K nearest neighbour (KNN) classifier, decision tree (DT) classifier and random forest (RF) classifier. In Table 7, decision tree achieves 100% Acc both before and after feature selection, and is the best rhythm classifier. To facilitate comparison, we test the DT after feature selection and deep model as well as the original SVM classifier and deep model on all records in the two public databases, and the score U is improved to 1.6632 from 1.6539.

7. Conclusions

This paper proposes a two-step method for the starting and ending point location of AF episodes as well as the detection of variable-length AFp rhythms. It is proposed for the first time that AFp recognition and AF episode location have been jointly considered for AFp event detection based on machine learning. In the first step, feature extraction and a classifier are combined to identify the rhythm type of a single record, which is either non-AF rhythm, or AFp rhythm, or AFf rhythm. In the second step, we use a CNN and take a single heartbeat as input for automatic deep feature extraction and heartbeat classification. In order to adapt to training on a large beat training set, a phased training method is proposed to change the size of the training set and the learning rate adaptively. In the end, our two-step method obtains a final score U of 1.9310 on the unpublished test set and ranks eight in CPSC 2021.

At present, we have optimized the algorithm for rhythm classification. (1) The genetic algorithm is applied for feature selection, and eight useful features are retained. (2) After multiple classifiers are compared, the SVM classifier are replaced by a decision tree classifier. In order to make the optimized method comparable with the original method, we test on all the records in the two public CPSC 2021 databases. Finally, the score U of the optimized method is improved by 0.0093. Our optimized two-step method is suitable for the AFp analysis of long-term recordings and helpful for AF burden index measurement.

Acknowledgments

This work was supported in part by Shanghai Municipal Special Project of Industry Transformation and Upgrading under Grant GYQJ-2020-1-31, Medical Scientific Research Key

Project of Jiangsu Commission of Health under Grant ZDB2020025 and Medical Scientific Research Instructional Project of Jiangsu Commission of Health under Grant Z2020075.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. M. Young, Atrial fibrillation, *Crit. Care. Nurs. Clin.*, **31** (2019), 77–90. https://doi.org/10.1016/j.cnc.2018.11.005
2. A. Margulescu, L. Mont, Persistent atrial fibrillation vs paroxysmal atrial fibrillation: Differences in management, *Expert. Rev. Cardiovas.*, **15** (2017), 601–618. https://doi.org/10.1080/14779072.2017.1355237
3. J. Imberti, W. Y. Ding, A. Kotalczyk, J. Zhang, G. Boriani, G. Lip, et al., Catheter ablation as first-line treatment for paroxysmal atrial fibrillation: A systematic review and meta-analysis, *Heart*, **107** (2021), 1630–1636. https://doi.org/10.1136/heartjnl-2021-319496
4. S. Hong, Y. Zhou, J. Shang, C. Xiao, J. Sun, Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review, *Comput. Biol. Med.*, **122** (2019), 103801. https://doi.org/10.1016/j.compbiomed.2020.103801
5. E. K. Wang, L. Xi, R. P. Sun, F. Wang, L. Y. Pan, C. X. Cheng, et al., A new deep learning model for assisted diagnosis on electrocardiogram, *Math. Biosci. Eng.,* **16** (2019), 2481–2491. https://doi.org/10.3934/mbe.2019124
6. G. Hindricks et al., 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC, *Eur. Heart J.*, **42** (2021), 373–498. https://doi.org/10.1093/eurheartj/ehaa612
7. Q. Li, B. Su, J. Liu, Diagnostic values of different ECG durations in paroxysmal AF diagnosis, *Ann. Noninvas. Electro.*, **27** (2022), e12921. https://doi.org/10.1111/anec.12921
8. M. Liu, X. Meng, P. Xiong, X. Liu, Detection of paroxysmal atrial fibrillation based on kernel sparse coding, *J. Elec. Info. Technol.*, **42** (2020), 1743–1749. https://doi.org/10.11999/JEIT190582
9. A. Petrėnas, L. Sörnmo, A. Lukoševičius, V. Marozas, Detection of occult paroxysmal atrial fibrillation, *Med. Biol. Eng. Comput.*, **67** (2020), 978–986. https://doi.org/10.1007/s11517-014-1234-y
10. N. Ganapathy, D. Baumgartel, T. M. Deserno, Automatic detection of atrial fibrillation in ECG using co-occurrence patterns of dynamic symbol assignment and machine learning, *Sensors (Basel)*, **21** (2021), 3542. https://doi.org/10.3390/s21103542
11. Y. Xin, Y. Z. Zhao, Y. H. Mu, Q. Li, C. C. Shi, Paroxysmal atrial fibrillation recognition based on multi-scale Rényi entropy of ECG, *Technol Health Care.*, **25** (2017), 189–196. https://doi.org/10.3233/THC-171321

12. E. Sabeti, M. B. Shamsollahi, F. Afdideh, Prediction of paroxysmal atrial fibrillation using empirical mode decomposition and RR intervals, in *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, (2012) 750–754. https://doi.org/10.1109/IECBES.2012.6498147
13. L. Y. Chen, M. K. Chung, L. A. Allen, M. Ezekowitz, K. L. Furie, P. McCabe, et al., Atrial fibrillation burden: moving beyond atrial fibrillation as a binary entity: A scientific statement from the American Heart Association, *Circulation*, **137** (2018), E623–E644. https://doi.org/10.1161/CIR.0000000000000568
14. X. Wang, C. Ma, X. Zhang, H. Gao, G. Clifford, C. Liu, Paroxysmal atrial fibrillation events detection from dynamic ECG recordings: The 4th China physiological signal challenge 2021, 2021, *PhysioNet*, **2021** (2021), 1–83. https://doi.org/10.13026/ksya-qw89
15. N. Larburu, T. Lopetegi, I. Romero, Comparative study of algorithms for atrial fibrillation detection, *Comput. Cardiol.*, **38** (2011), 265–268.
16. B. Chen, W. Chen, J. Liu, L. H. Zhu, The research of electrophysiological data normalization, in *2010 5th International Conference on Computer Science & Education*, (2010), 149–151. https://doi.org/10.1109/ICCSE.2010.5593672
17. B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, (1992), 144–152. https://doi.org/10.1145/130385.130401
18. P. de Chazal, M. O'Dwyer, R. B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, *IEEE Trans Biomed Eng.*, **51** (2004), 1196–1206. https://doi.org/10.1109/TBME.2004.827359
19. A. A. Almazroi, Survival prediction among heart patients using machine learning techniques, *Math. Biosci. Eng.*, **19** (2022), 134–145. https://doi.org/10.1109/TBME.2004.827359
20. J. Jiang, H. F. Zhang, D. C. Pi, C. L. Dai, A novel multi-module neural network system for imbalanced heartbeats classification, *Exp. Syst. Appl. X*, **1** (2019), 100003. https://doi.org/10.1016/j.eswax.2019.100003
21. Association for the Advancement of Medical Instrumentation, *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, 1998.
22. A. Saif, A. Garba, J. Awwalu, H. Arshad, L. Zakaria, Performance comparison of min-max normalisation on frontal face detection using haar classifiers, *Pertanika J. Sci. Technol.*, **25** (2017), 163–171.
23. C. He, H. kang, T. Yao, X. Li, An effective classifier based on convolutional neural network and regularized extreme learning machine, *Math. Biosci. Eng.*, **16** (2019), 8309–8321. https://doi.org/10.3934/mbe.2019420
24. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv: 1409.1556
25. K. Kosaka, T. Itoh, A visualization method for training data comparison, in *2021 25th International Conference Information Visualisation (IV)*, (2021), 205–210. https://doi.org/10.1109/IV53921.2021.00040
26. K. Weimann, T. Conrad, Transfer learning for ECG classification, *Sci. Rep.*, **11** (2021), 1–12. https://doi.org/10.1038/s41598-021-84374-8
27. C. Yu, X. Qi, H. Ma, X. He, C. Wang, Y. Zhao, LLR: Learning rates by LSTM for training neural networks, *Neurocomputing*, **394** (2020), 41–50. https://doi.org/10.1016/j.neucom.2020.01.106

Appendix

Table A1. The introduction of methods for rhythm classification.

Existing method	Liu [8]	Petrenas [9]	Ganapathy [10]	Xin [11]	Our method	Our optimized method
Feature	RR intervals	RR intervals, f wave presence, P wave absence, and noise level	RR intervals	HRV signal	RR intervals	
Lead	Single lead	Two leads	Single lead			
FT	Covariance representation and kernel mapping	Fuzzy logic	Dynamic symbol assignment and covariance representation	Wavelet analysis and scale entropy calculation	Time domain, frequency domain, and non-linear domain representation	Time domain, frequency domain, non-linear domain representation, and feature selection
Classifier	Kernel sparse coding and dictionary learning	Sliding window and threshed	KNN, SVM, RF, ROF, EL	SVM	SVM	DT
length	33 RR intervals	Finite length (about 2 min)	Finite length (1 min~5 min)	Fixed length (5 min)	Variable length (≥ 4 RR intervals)	Variable length (≥ 4 RR intervals)
Class	AFp rhythm	Occult AFp rhythm	AFp and non-AF rhythm	Non-AF, AFp and other rhythm	AFp, AFf and non-AF rhythm	AFp, AFf and non-AF rhythm

Note: FT represents feature transformation; HRV represents heart rate variability; ROF represents rotation forest; And EL represents ensemble learning.