

A cascaded step-temporal attention network for ECG arrhythmia classification

Yanyun Tao, Guoqi Yue, Kaixin Wang
School of rail transportation, Soochow university,
Suzhou, China, 215005

Yuzhen Zhang (*), Bin Jiang
The First Affiliated Hospital of Soochow University,
Suzhou, China, 215006
zhangyuzhen@suda.edu.cn

Abstract—To improve the accuracy of arrhythmia diagnosis and reduce the recheck time, we design a cascaded step-temporal attention network called ArrhythmiaNet to classify 15 categories of arrhythmias on electrocardiogram (ECG) signals. In ArrhythmiaNet, the first level contains a convolution layer with step-attention, which recognizes abnormal heartbeat and provides morphological feature expression. The second level is composed of a gated recurrent unit (GRU) with temporal attention, which mines the temporal correlation of long-term rhythm for abnormal rhythm judgement. To share the feature expression, ArrhythmiaNet was trained by end-to-end multitask learning. In the experiment, ArrhythmiaNet and the comparison algorithms were tested on a dataset (819 training samples and 264 test samples) from MIT-BIH arrhythmia database. The results showed that the accuracy of ArrhythmiaNet was 20.3% higher than that of support vector machine (SVM), Naive Bayesian, gradient boost decision tree (GBDT) and random forest (RF), and 8.2% higher than that of long-term memory network (LSTM) and recurrent neural network (RNN). Compared with 1-dimension convolutional neural network (1D-CNN), ArrhythmiaNet obtained similar overall accuracy, higher recall and precision. Compared to the genetic ensemble of SVM classifiers and evolutionary neural system, ArrhythmiaNet has much lower complexity than them with competitive accuracy. Besides, ArrhythmiaNet has higher interpretability in arrhythmias diagnosis.

Keywords—arrhythmia; ECG; cascaded neural network; attention network; convolutional layer; gated recurrent unit; multitask learning

I. INTRODUCTION

Arrhythmia is a serious cardiovascular disease, which can lead to palpitation, dizziness, blackness, syncope and even sudden death. Therefore, its diagnosis and treatment have important clinical significance. Arrhythmia is mainly caused by abnormal cardiac electrical activity, which can be measured by surface electrodes and stored as time-series pulse data, that is, body surface electrocardiogram (ECG). Most of the arrhythmias are paroxysmal and would appear in ECG signal at irregular intervals. With the dynamic ECG machines, i.e. Holter, the clinicians can analyze the heart rate, heart rhythm and diagnose arrhythmias diseases by detecting the patients' routine or dynamic monitoring of their ECG. Clinically, the analysis software of Holter can automatically establish templates of normal heartbeat, atrial premature beat

and ventricular premature beat. According to the templates, the initial recognition of each heartbeat of the patient can be given, and the clinicians further review the recognition. However, because of the baseline drift and the poor recognition of software, the accuracy of the instrument diagnosis is always not high. Moreover, the data volume of ECG signals recorded in the Holter is so huge, each record is about 100 thousand beats, that it takes a lot of time for clinicians to review. It is inevitable to misread and miss important information in ECG signals [1].

Using machine learning technology to assist ECG signal recognition and arrhythmia diagnosis can greatly improve the efficiency and reduce the misdiagnosis rate. They include the following categories: a) multi-layer perceptron (MLP) [2,3]. For MLP, the setting of initial parameters has a great influence on the final results; b) hybrid neural network [4]. It employed fuzzy clustering layer in a neural network, reducing the training complexity of MLP; c) support vector machine (SVM) [5,6]. SVM classifier has a good classification ability on a small dataset. For a large dataset, the quadratic programming problem becomes more complex to solve, and the choice of kernel function is still a difficult issue; d) hidden Markov model [7]. It can find the hidden states under different probability density distribution, and then used them for signal recognition; e) deep learning networks [8-11] have been successfully applied in ECG signal recognition. Compare to the other methods, deep learning method, which was an expert in extracting complex features and discovering time dependency in ECG sequence, has better potential in clinic application.

To accomplish the task of arrhythmias recognition, both morphological and temporal features, such as QRS, P waves and $R-R$ time interval in ECG signal, must be fully extracted. In this study, we proposed a cascade neural network called ArrhythmiasNet to improve the accuracy of arrhythmias recognition. In ArrhythmiasNet, a step-temporal attention mechanism is designed to concentrate on the critical area of the ECG signal. Due to the difference between recognition on heartbeat and heart rhythm, the recognition task is divided into two sub-tasks for complexity reduction, and a joint loss function is defined for end-to-end multitask learning.

II. RELATED WORK

In literature [12,13], due to the high beat variability of

these characteristics among individuals of arrhythmia signal, the accuracy of the current methods for ECG diagnosis on different patients were mostly around 85%, which cannot get satisfactory results. Paweł Pławiak [5] designed the genetic ensemble of SVM classifiers, which used cross-validation, integrated learning, hierarchical learning, and genetic optimization on classifier parameters. It obtained the classification sensitivity of seventeen heart diseases at 91.40% (64 errors per 744 samples), the accuracy of 98.9% and the specificity of 99.4%. Paweł Pławiak [6] proposed an SVM classifier-based evolutionary neural system based for seventeen cardiac health recognition. This system obtained the recognition accuracy of 98.85% and the specificity of 99.39%.

In reference [3], a block-based neural network (BBNN) with particle swarm optimization (PSO) was developed for the personalized classification of ECG signals. The simulation results show that the average detection accuracy of ventricular ectopic beats (VEBs) demonstrated by MIT-BIH arrhythmia database was 98.1%, and the average detection accuracy of supraventricular ectopic beats (SVEBs) was 96.6%. Pan et al. [7] developed a hidden Markov model with adaptive features. The model recognized four abnormal heartbeats including the left bundle branch block (LBBB), the right bundle branch block (RBBB), the ventricular premature contractions (VPC), and the atrial premature contractions (APC) with a total accuracy 89.25%.

Literature [8] used a convolutional neural network (CNN) to perform the classification on five typical cardiac arrhythmia signals, which obtained an accuracy of 97.8%. In reference [11], an automatic classification method of abnormal ECG based on deep neural network (DNN) was proposed. They made the classification on normal ECG and abnormal ventricular premature beat, supraventricular premature beat, and fusion of ventricular and normal beat. The overall accuracy was 99.68%. Yang et al. [14] used stacked sparse self-encoder and softmax to identify six categories of heartbeats with an accuracy of 99.22%.

In reference [9], a deep CNN classifier was proposed for heartbeat recognition, which converted ECG signals into dual beat coupling matrix as two-dimensional input to the classifier, can capture the beat shape and correlation of ECG signals at the same time. This CNN classifier improved the sensitivity and positive prediction rate of S-beats, which were increased by 12.2% and 11.9% respectively. In reference [10], a 9-layer CNN was developed to automatically recognize five types of heartbeat in ECG signals. Its average accuracy was 94.03%. Literature [15] attempted to use CNN to directly identify thirteen classes of arrhythmias. The average accuracy was 99.24%. Currently, Ozal Yildirim et al. [16] investigated a deep bidirectional long short-term memory (LSTM) network with wavelet sequences (WS) process, called DBLSTM-WS, for ECG signal classification. Although DBLSTM-WS got a high accuracy on ECG classification, it can deal with only five classes of arrhythmias. Ozal Yildirim et al. [17] further gave a deep 1-dimension CNN (1D-CNN) to study the long-term ECG signals for seventeen arrhythmias. The calculation effort was small and the overall accuracy was 91.33%. To address the problem of declined accuracy of the trained model on new patients, multitask learning (MTL) network [18] was explored

for ECG analysis by inserting parameter-sharing neural layers to the neural network. The average accuracy of multitask learning network improved up to 5.1%.

III. CASCADED STEP-TEMPORAL ATTENTION NETWORK

A. Framework

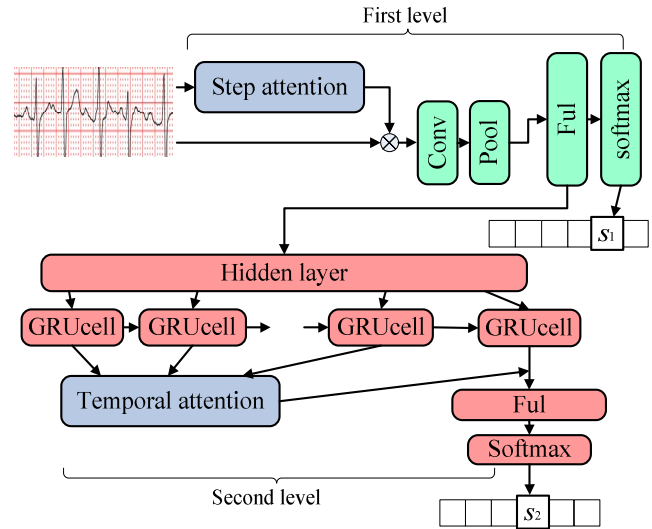


Figure 1 two levels of ArrhythmiaNet

Some arrhythmias can be recognized by a single heartbeat signal recognition, e.g. bundle branch block and ventricular premature beat, while other ones need the recognition on a period of heart rhythm, e.g. atrial fibrillation, ventricular flutter and tachycardia. Correspondingly, the heartbeat recognition only required a short-term data of a single heartbeat, while the heart rhythm recognition required both the shape of heartbeat signal and the time dependency in ECG sequence. In this work, a cascaded neural network with step-temporal attention called ArrhythmiaNet is designed for arrhythmia classification on ECG data. It can be adaptive to both heartbeat and heart rhythm data. As shown in Figure 1, the cascade structure contains two levels:

1) *The first level:* step attention, based on opening-closing operation, focuses on the most important information of ECG signal, which plays the pivotal role in the heartbeat recognition. The step attention scores integrated with the input data were fed to the convolutional layer (Conv) with max-pool layer, which extracts the morphological features of heartbeats. Fully connected layer (Ful) realizes the non-linear mapping, transforming feature maps to a feature expression. Finally, the softmax layer conducts on the feature expression to calculate the probabilities of heartbeats belonging to diseases categories. Meanwhile, the feature expression was provided to *the second level* for rhythm recognition.

2) *The second level:* the hidden layer receives the feature expression of the first level and performs non-linear mapping. Gated recurrent unit (GRU) cell process the time-series information (k time steps) from $t-k+1$ to t . The switch function in the updated gate and reset gate retains the most important information and the time dependency between the heart rhythm data. The temporal attention concentrates on the

critical rhythmic information in the ECG signal sequence. The softmax layer processes the k time-step outputs of GRU cell with temporal attention and calculates the probability of the heart rhythm belonging to categories.

B. Step-temporal attention

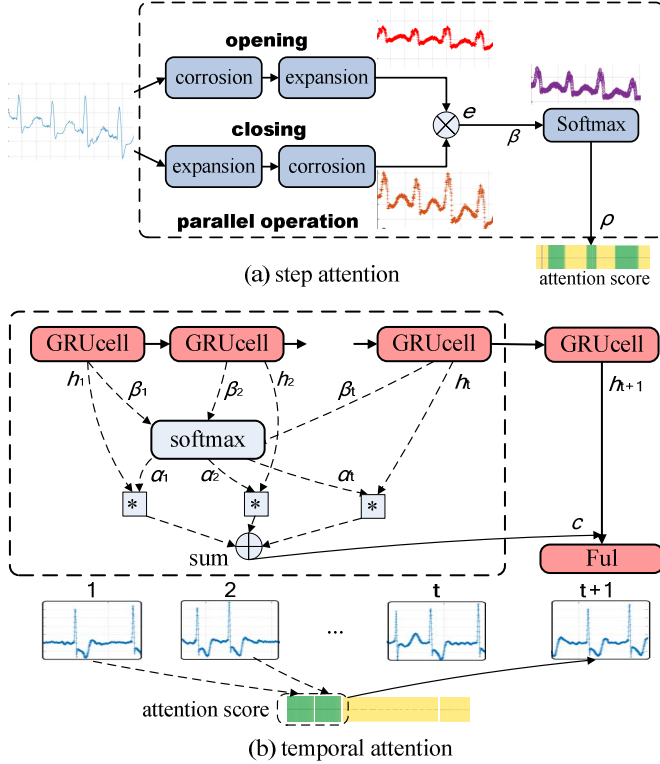


Figure 2 step-temporal attention

Step attention: the essence of ECG variation is the step of the signal. In ECG, QRS and P waves involve the information of arrhythmia. It was difficult for common filtering methods to extract step features of ECG signal with serious baseline drift. To capture the feature of step signal, we proposed the step attention as a component of ArrhythmiaNet. In this component, step perceptron is designed based on mathematical morphology, which captures the specific signals of ECG.

Corrosion (\ominus) and expansion (\oplus) are the most basic operation in mathematical morphology. In this work, one-dimension ECG data is the study object. Corrosion can eliminate the useless noise from ECG data, while expansion is able to enhance the step signal of QRS and P waves in ECG.

Assumed a sequence ECG data $X=\{0,1,\dots, N-1\}$, and the structure element $B=\{0,1,\dots,M-1\}$, $N>M$, then, the corrosion operation is defined as follow.

$$X\ominus B(m)=\min_{n=0,1,\dots,M-1}\{X(m+n)-B(n)\} \quad (1)$$

Where, $m=0, 1, \dots, N-M$. Structure element B determines the effect of morphology computation. For computational convenience, the linear element structure (32 sample points) is used.

The expansion operation is defined as follow.

$$X\oplus B(m)=\max_{n=n-M+1,\dots,1,0}\{X(m+n)-B(n)\} \quad (2)$$

Here, $m=M-1, M, \dots, N-1$.

The corrosion and expansion operations can be integrated as opening and closing operation.

$$\begin{aligned} \text{opening: } y_{op} &= (X\ominus B)\oplus B \\ \text{closing: } y_{cl} &= (X\oplus B)\ominus B \end{aligned} \quad (3)$$

The opening and closing operation can be combined in different signal processing processes, such as opening-closing operation, closing opening operation and parallel opening-closing operation.

The details of step-attention were shown in Figure 2 (a). In this attention, the parallel opening-closing operation was used. The opening operation is responsible to eliminate the baseline drift and noise from ECG, while the closing operation keeps the pivotal information of QRS and P waves as possible. Through this parallel operation, attention can be focused on the key information of QRS and P waves without being affected by the noise and baseline drift.

The score of step attention is indicated by ρ_i , and the input data with step attention score is denoted as X' . Their forms are as follow.

$$\begin{aligned} \rho_i &= \frac{\exp(\beta_i e_i)}{\sum_j \exp(\beta_j e_j)} \\ e_j &= y_{op} \cdot y_{cl} \\ X' &= X * \rho_i \end{aligned} \quad (4)$$

Here, e indicates the output of opening-closing operation on input X , β indicates the weight of e , and j denotes the sample point of X .

Temporal attention: for heart rhythm classification, to discover the time dependency in ECG sequence is necessary. This requires GRU cells to be strong enough for encoding the heart rhythm data. It must ensure that the rhythm information at the $t+1$ time step obtained from the ECG sequence can retain the information of previous time steps (1, 2, ..., t) as much as possible. As the size of ECG sequence data continuously grows, GRU cells cannot completely retain most of the information of previous time steps. To acquire the information of heart rhythm at previous time steps, the temporal attention mechanism is adopted.

At the $t+1$ time step, the hidden state (i.e. h_1, h_2, \dots, h_t) of GRU cells of previous t time steps will be weighted and averaged as the context vector of t time steps. The $t+1$ time step can acquire the most relevant information on previous time steps from the context vector. We take an ECG sequence of atrial fibrillation, in which the distance between heartbeats was irregular, as the example shown in Figure 2(b). An attention score vector of previous time steps of this sequence can be calculated, and the attention scores at the time step 1 and 2 (two green grids) are larger than other time steps. As a

result, the states at time step 1 and 2 are more likely to be integrated with the state at the $t+1$ time step.

The score α_t corresponds to the state h_t of GRU cell at the t time step, and h_c denotes the output state of the GRU cell at the $t+1$ time step with temporal attention score vector. The forms of α_t and h_c are as follow.

$$\begin{aligned}\alpha_t &= \frac{\exp(\beta_t h_t)}{\sum_j \exp(\beta_j h_j)} \\ c &= \sum_T \alpha_t h_t \\ h_c &= \tanh(W_c [c; h_{t+1}])\end{aligned}\quad (5)$$

Here, β_t indicates the weight of h_t , and T denotes time steps; c indicates the context vector of previous t time steps. W_c denotes the weight connecting the combination of output states to the classification layer.

C. Multitask loss function

In this work, the arrhythmia classification task can be divided into two sub-tasks, i.e. heartbeat recognition and rhythm recognition. They have a strong correlation, sharing common feature expression and improving each other. For example, an abnormal heartbeat probably leads to an abnormal rhythm. To accomplish two sub-tasks simultaneously, multitask learning is a suitable way. The cross-entropy loss functions for heartbeat and heart rhythm recognition are as (6):

$$\begin{aligned}\text{loss}_1 &= -[s_1 \log(\hat{s}_1) + (1-s_1) \log(1-\hat{s}_1)] \\ \text{loss}_2 &= -[s_2 \log(\hat{s}_2) + (1-s_2) \log(1-\hat{s}_2)] \\ L_2 &= \|W_{\text{Conv}}\|_2 + \|W_{\text{Ful}}\|_2 + \|W_{\text{GRU}}\|_2 + \|W_{\text{H}}\|_2 \\ \text{dist} &= \|W_{\text{Ful}}^{\text{heartbeat}} - W_{\text{Ful}}^{\text{rhythm}}\|_2\end{aligned}\quad (6)$$

Here, \hat{s}_1 indicates the probability vector of heartbeat category, and s_1 is the label vector (one hot coding). \hat{s}_2 denotes the probability vector of the rhythm category, and s_2 is the label vector (one hot coding). L_2 is a regularization term, which is used to balance the classification accuracy and the complexity of network. In L_2 , W_{GRU} and W_{H} are GRU cell and hidden layer parameters, and W_{Conv} and W_{Ful} are convolutional kernel and parameters of fully connected layer, respectively. dist denotes the difference between parameters of fully connected layers for sub-tasks.

In order to realize multitask learning, we combine two loss functions and L_2 regularization term into a joint loss function (7).

$$\text{loss} = \text{loss}_1 + \text{loss}_2 + \lambda \cdot L_2 + \mu \cdot \text{dist}\quad (7)$$

Here, λ was the weight of L_2 , which balances the loss terms and regulation; μ was the weight of dist . Adaptive moment estimation (Adam) is used for ArrhythmiaNet optimization.

IV. EXPERIMENTS

A. Dataset

The dataset used in this study is selected from arrhythmia,

normal sinus rhythm, atrial fibrillation and malignant ventricular arrhythmia databases in MIT-BIH. Each sub-database contains ECG data with 2-leads and 250/360Hz sampling rate. The data of heartbeat mainly come from arrhythmia database (files "100", "101", "102" and "103"), normal sinus arrhythmia database (files "16265", "16272" and "16273") and supraventricular arrhythmia database (files "800", "892", "893" and "894"). The rhythm data are selected from atrial fibrillation database (files "04015" and "04043") and malignant ventricular arrhythmia database (files "418", "421", "428", "430", "612", "614" and "615").

TABLE I NUMBER OF SAMPLES ON CATEGORIES

	type	train	test
Heart beat	normal heart beat (N)	104	30
	ventricular premature beat (V)	83	24
	left bundle branch block (L)	50	16
	right bundle branch block (R)	70	18
	atrial premature beat (A)	12	6
	fusion of ventricular beat (F)	14	5
	pacing heart beat (/)	62	15
	supraventricular premature beat (S)	65	20
	Total	460	134
Heart rhythm	normal heart rhythm (N)	85	30
	ventricular tachycardia (VT)	37	16
	ventricular flutter (VFL)	29	11
	supraventricular tachyarrhythmia (SVTA)	46	17
	atrial fibrillation (AFIB)	65	23
	ventricular bigeminy (B)	16	7
	ventricular fibrillation (VF)	45	16
	second-degree 2° heart block (BII)	32	10
	Total	355	130

The number of samples in the dataset we used is 1083, i.e. 819 training samples and 264 test samples. Each sample is a signal with 2-leads, 250Hz sampling rate and 0.8s/4s time for heartbeat/heart rhythm recognition. The number of training and testing samples corresponding to categories is shown in Table I. The dataset was divided into 16 categories for the heartbeat and heart rhythm recognition.

B. Parameters of algorithm

The parameters of ArrhythmiaNet are shown in Table II. The hyper-parameter setting for ArrhythmiaNet training are as follow: the number of iterations and training samples were 800 and 819, respectively; the dataset is not large, so, we initially used full batch learning and #epoch is 800, and we tested whether the other batch size could be more effective in training ArrhythmiaNet. The initial learning rate was set to 0.01. λ in loss function (7) was initially set to 2 and other values of λ were also tested. μ was set to 2.5. For step attention, the structure element B is a vector of 32 random values in $[0, 2.5]$, which was the best one we have tested.

Two types of inputs, heartbeat data and heart rhythm data, were one-dimensional vectors. Heartbeat data is a segment of 200 sampling points with a depth of 2 (2 leads), while heart rhythm data is a segment of 1000 sampling points with a depth of 2 (2-leads). Both ArrhythmiaNet and the comparison algorithms were implemented in Tensorflow, and run on GPU 1060i. Each algorithm runs ten repetitions on the dataset, and the best result was given.

TABLE II PARAMETERS OF ARRHYTHMIANET

No.	layer	structure	function
1	Conv	8×2	ReLU, stride=2, size=2
2	Maxpooling	--	Poolsize =2, stride=2
3	Flatten	--	--
4	Dense	16	Dropout=0.5
5	Dense	8	Softmax
6	GRU cell	--	Input = 100, Time_step=10
7	Hidden	50	tanh
8	Dense	8	Softmax

C. Evaluations

The evaluations on algorithms are Accuracy, Precision and Recall, and their expressions are as the form.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$precision = \frac{TP}{TP + FP} \quad (9)$$

$$recall = \frac{TP}{TP + FN} \quad (10)$$

Here, True positive (TP): predict positive class as positive class number; true negative (TN): predict negative class as negative class number; false positive (FP): predict negative class as positive class number error (type I error); false negative (FN): predict positive class as negative class number → Missing Report (Type II error).

V. RESULTS AND DISCUSSION

We compared ArrhythmiaNet with different comparison

TABLE III ACCURACY, RECALL AND PRECISION OF COMPARISON ALGORITHM IN ARRHYTHMIA CLASSIFICATION

	Methods	samples*channels	#categories	types	Accuracy%	Recall%	Precision%
External-reported comparison	BBNN+PSO [3]	--	5	heartbeat and heart rhythm	82.91	--	--
	RNN with MTL [18]	--	4		81.29	--	--
	LSTM with MTL[18]	--	4		85.01	--	--
	Genetic ensemble of SVM classifiers [5]	3600	15		93.61	--	--
	Evolutionary neural system [6]	3600	15		94.01	--	--
	DBLSTM-WS [16]	variable	5		99.39	--	--
	1D-CNN [17]	3600	15		92.51	88.57	90.48
	ArrhythmiaNet	200*2(heart beat) 1000*2(rhythm)	15		92.6	98.3	100
Internal-tested comparison	RF	200*2(heart beat) 1000*2(rhythm)	15	heartbeat	64.5	97.6	67.5
				heart rhythm	47.6	71.0	76.9
	SMOTE-FFT-RF			heartbeat	72.5	76.6	91.6
				heart rhythm	57.6	64.0	76.6
	GBDT			heartbeat	74.6	74.1	71.1
				heart rhythm	65.3	10.3	100
	NB			heartbeat	23.1	13.1	21.3
				heart rhythm	23.6	13.8	44.8
	SVM			heartbeat	67.6	51.2	76.5
				heart rhythm	47.5	77.8	58.5
	ArrhythmiaNet			heartbeat	93.9	100	100
				heart rhythm	91.5	96.7	100

A. Accuracy of ArrhythmiaNet

The results in Table III showed that ArrhythmiaNet achieved the highest precision and recall, and an enough high recognition accuracy. Compared with the internal-tested

algorithms and gave the test results in Table III. Figure 3 shows the decline of loss in the training process of ArrhythmiaNet.

- The comparison algorithms such as random forest (RF), synthetic minority oversampling technology-fast Fourier transform-RF (SMOTE-FFT-RF), gradient boosting decision tree (GBDT), naive Bayesian (NB) and support vector machine (SVM), which were tested on our dataset, are called as internal-tested algorithms.
- The comparison algorithms such as 1D-CNN, BBNN+PSO, recurrent neural network (RNN) with MTL, LSTM with MTL, DBLSTM-WS, Genetic ensemble of SVM classifiers and Evolutionary neural system were not tested on our dataset. Their results were directly taken from the reports in corresponding works. These algorithms were denoted as external-reported algorithms. The comparison of ArrhythmiaNet with external-reported algorithms maybe not fair but can be a reference.

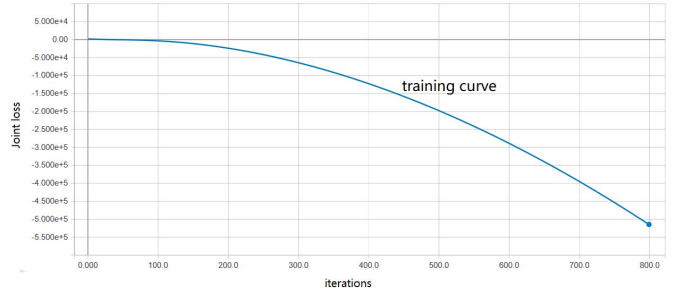


Figure 3 the joint loss in the training process of ArrhythmiaNet

comparison algorithms, ArrhythmiaNet outperformed more than 20.3% in terms of accuracy for heartbeat recognition, and also got a much higher precision and recall. Especially for the recognition of rhythm, ArrhythmiaNet yielded nearly 28.6%

higher accuracy than internal-tested comparison algorithms. Since ArrhythmiaNet extracted morphological features with drift elimination and kept the temporal information, it has more advantages than the internal-tested comparison algorithm. Step attention, masked the information not worthy of attention in ECG data, certainly have large positive impact on arrhythmia diagnosis. Meanwhile, ArrhythmiaNet's hierarchical processing of ECG signals also has a high interpretability, which reduced the difficulty of multiclass arrhythmia classification. Through multitask learning, the feature expression of the heartbeat and rhythm recognition were commonly shared. The abnormal features extracted from the short-term data, such as premature beat, can also help the recognition on long-term rhythm data, such as bigeminy, tachycardia and atrial fibrillation.

The recall and precision of ArrhythmiaNet on 15 categories of arrhythmias can be seen in Figure 4, respectively. Figure 4 (a) shows that the recall of ArrhythmiaNet on category F was lower than on the other categories. It is noted that ArrhythmiaNet wrongly classified samples of category F into category V. This is because the category F, fusion of ventricular beat, can be generated by the fusion of atrial premature beat (A) and ventricular premature beat (V). Figure 4 (b) shows that the recall of ArrhythmiaNet on categories (B and (BII were lower than on the other categories. Meanwhile, the precision of ArrhythmiaNet on category (B was also not high. Some samples of category (B were classified to (BII and (SVTA. This was arisen by the absence of association between two leads extracted by ArrhythmiaNet.

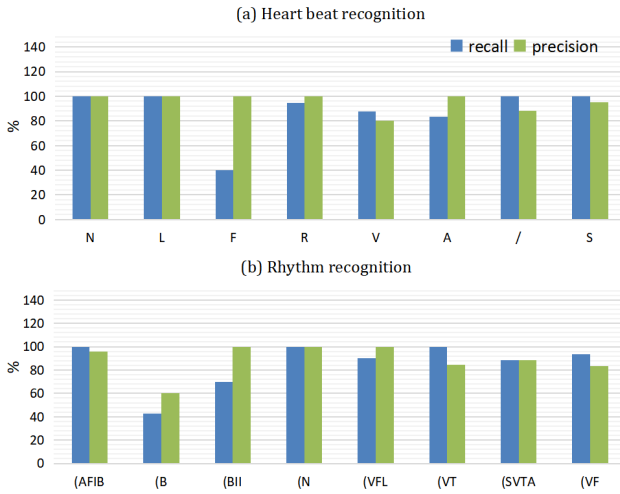


Figure 4 recall and precision on 15 categories (a) heartbeat recognition (b) heart rhythm recognition

The comparison of ArrhythmiaNet with external-reported comparison algorithms can be seen in Table III. Although the comparison is not completely equivalent, the dataset, on which these algorithms were tested, come from the same database with our dataset. Both LSTM and RNN, integrated with MTL, yielded a lower accuracy than that of ArrhythmiaNet, and they can only recognize 4 categories of arrhythmias. BBNN+PSO presents a network for 5 categories classification on ECG, requiring a pre-process for power-line noise and base-line drift elimination. It also obtained a lower accuracy than that of ArrhythmiaNet. DBLSTM-WS got a high accuracy of 99% for

five categories classification, but it still depended on frequency feature extraction by wavelet sequence.

Among external-reported algorithms, the genetic ensemble of SVM classifiers and Evolutionary neural system, both of which are based on evolutionary algorithm and SVM classifier, obtained the highest accuracy for 15 categories of arrhythmias classification, at the cost of much higher complexity of the model. Compared with 1D-CNN, designed for 15 categories of arrhythmias classification, ArrhythmiaNet got a higher precision and recall without any decrease in the accuracy. Meanwhile, the structural complexity of ArrhythmiaNet and the computational time spent on training ArrhythmiaNet were both greatly reduced. The first level of ArrhythmiaNet only used 1 convolutional layer with max-pooling, and the second level used GRU with 50 hidden nodes, while 1D-CNN uses 7 convolutional layers and 4 max-pooling layers. Additionally, ArrhythmiaNet only required an input data of 200/1000 sampling length, while 1D-CNN required a data of 3600 sampling length.

B. Analysis on internal comparison algorithm

Figure 5 shows the confusion matrix of ArrhythmiaNet, SMOTE-FFT-RF, GBDT and SVM tested on heartbeat and rhythm data, respectively. The efficiency of the tree structure models (RF series, GBDT), NB and SVM on a small dataset was high. It is noted that the accuracy of heartbeat recognition was higher than that of the rhythmic type. As given in Table III, the highest accuracy is up to 74.6%. The recognition of rhythmic type is generally poor, and the accuracy is only 65%. Some models have a high recall but low precision and the number of samples wrongly classified into normal rhythm are large; GBDT has the lowest recall, although the accuracy was high. There are serious errors in the recognition of normal rhythm.

Since the number of features was larger than that of samples, first we tried principal component analysis (PCA) to reduce the dimension on a RF model. However, the large difference between the number of features and samples lead to an unsatisfied accuracy, i.e. only 25~55%. Then, we attempted to use the oversampling technology to generate a more balanced dataset. The RF model built on the dataset obtained an improvement in accuracy but not large. Finally, we attempted to transform ECG time-domain to frequency-domain features by FFT, and then used these features to create a RF model. The results turned out only a slight improvement in accuracy. When FFT and the oversampling method were both used to build a RF model, the confusion matrix of classification becomes more reasonable and the accuracy was significantly improved.

GBDT combined complex features to classify samples to corresponding categories through iterative residual, but itself cannot extract features. It can be seen in the confusion matrix of Figure 5 that, GBDT gave the wrong classification on most samples of AFIB and N. Since the number of variables of ECG data is larger than the number of samples. It is very difficult for GBDT to find appropriate feature combination. As for the SVM model, it wrongly classified most of the samples into (N and (VT). For example, all (B samples were classified into (VT. This leads to the overall low accuracy of

SVM on normal rhythm recognition. This is because SVM performed classification without extraction on morphological and temporal features of the ECG signal.

NB algorithm was suitable to process the heart rhythmic data with such a large number of variables (1000*2). However, NB obtained an extremely low accuracy. This is because NB assumed variables independent. The correlation between morphological and temporal features is split. When calculating the category probability of the sample, the useless variables probably lead to the inaccurate estimation. Another reason is that the estimation on the conditional probability depends on the maximum likelihood estimation. If the samples are independent and identically distributed, there maybe a large deviation in the estimation of data distribution.

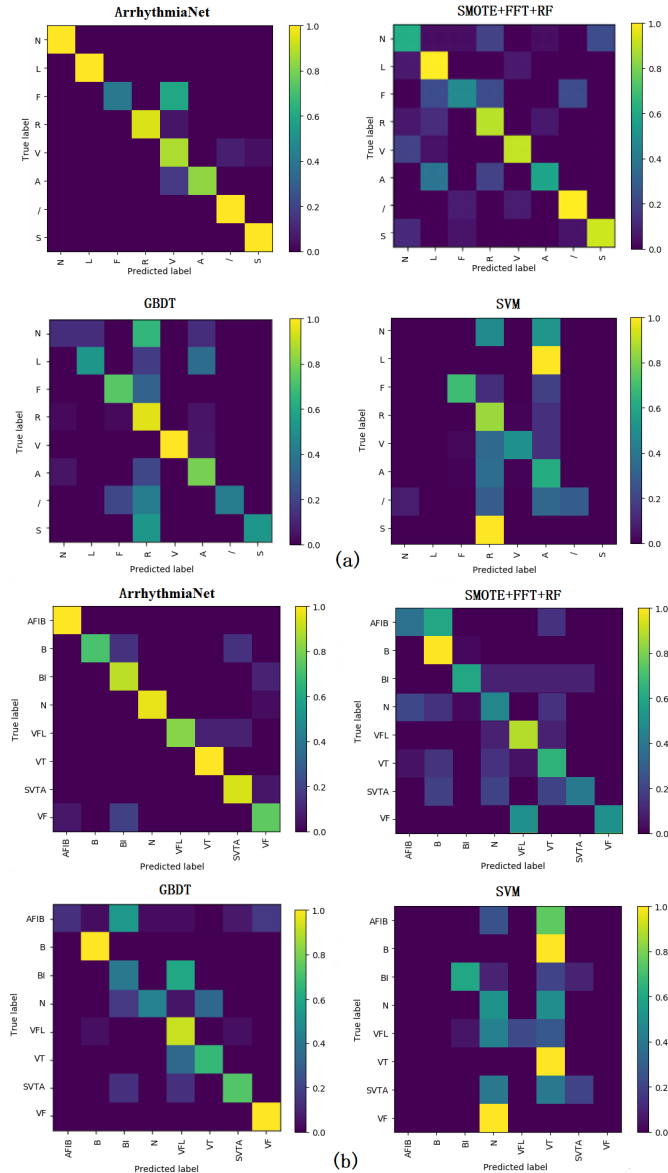


Figure 5 Confusion matrix of ArrhythmiaNet, SMOTE-FFT-RF, GBDT, SVM (a) heartbeat recognition (b) heart rhythm recognition

C. Ablation study

Table IV, V and VI show the results of ablation study on ArrhythmiaNet. In tables, F_1 indicates the balanced F score, which was Harmonic mean of precision and recall. It was noted in Table IV that when the batch size was less than 350, there was no significant improvement or decrease in the accuracy, precision and recall on the test set. Only when the batch size was up to 350 can the performance of ArrhythmiaNet be improved. Under different batch sizes, the mean of accuracy, precision and recall were around 81%, 90% and 87%, respectively. Full batch size learning was better than other batch sizes in the performance of classification. The advantage of full batch size learning can be also reflected by F_1 .

TABLE IV THE EFFECT OF BATCHING IN TRAINING ARRHYTHMIA NET

Batch size	Accuracy%	Recall%	Precision%	F_1
50	79.1	84.5	86.0	0.851
100	83.0	94.8	88.7	0.916
150	81.4	86.2	87.7	0.869
200	80.3	91.4	80.6	0.857
250	82.9	89.6	94.5	0.920
300	80.8	93.1	88.5	0.907
350	83.8	96.6	91.8	0.941
full batch size	92.6	98.3	100	0.991

TABLE V THE COMPARISON BETWEEN MTL AND STL IN ARRHYTHMIA CLASSIFICATION

	types	Accuracy%	Recall%	Precision%	F_1
MTL	heartbeat	93.9	100	100	1
	heart rhythm	91.5	96.7	100	0.983
STL	heartbeat	92.4	96.4	100	0.981
	heart rhythm	89.2	100	90.9	0.952

TABLE VI THE EFFECT OF L_2 REGULARIZATION TERM IN LOSS FUNCTION

loss function	λ	Accuracy%	Recall%	Precision%
regularization term	0.1	85.0	93.1	93.1
	0.5	87.7	96.6	94.9
	1	88.9	98.2	96.6
	1.5	91.1	100	96.7
	2	92.6	98.3	100
	2.5	92.3	100	100

Observed from Table V, both MTL and STL could achieve high accuracy, precision and recall on the test set for two sub-tasks, respectively. MTL slightly outperformed STL in accuracy but performed much higher in precision than STL. It was noted that MTL obtained higher F_1 than that of STL. This demonstrates that the shared feature expressions between two sub-tasks certainly improved the training quality of ArrhythmiaNet. Besides, we tested different weights of the regularization term in Table VI. In the joint loss function, the effect of the regularization term has been illustrated by the continuous improvement on the accuracy, precision and recall as λ increases. When λ was 2.5, both the recall and precision achieved 100% and the accuracy was 92.3%. So, the regularization term not only reduced the complexity of

ArrhythmiaNet but also improved the generalization.

VI. CONCLUSION

In order to improve the accuracy of arrhythmia diseases diagnosis and reduce the recheck time, we proposed ArrhythmiaNet to complete the arrhythmia classification on ECG signals. There are advantages of ArrhythmiaNet as follow.

- ECG data entering into ArrhythmiaNet does not need to be filtered and denoising. It can achieve automatic denoising, feature extraction and classification;
- There is no necessary to detect QRS and *P* waves in the ECG signal. In ArrhythmiaNet, step-temporal attention modules automatically focus on QRS and *P* waves of ECG signal and retain rhythm information of ECG sequence, which definitely have a large impact on arrhythmia recognition;
- ArrhythmiaNet realized the end-to-end training of neural network structure. The cascaded structure made the ArrhythmiaNet more interpretability on arrhythmias diagnosis;
- Multitask learning actually reduced the difficulty in recognizing 15 categories of arrhythmias and improved the recognizing accuracy.

In future, there are two tasks needed to be accomplished. The first one is that we would try to realize the recognition on 17 categories of arrhythmias and improve the recognition accuracy on the imbalance dataset. The second one is that the interpretability of neural network for arrhythmias diagnosis should be improved.

REFERENCES

- [1] J Liu, G Li. The comparison between computer automatic measurement and manual measurement of electrocardiogram. Chinese Journal of cardiovascular disease. 2001; 29(11): 691-692
- [2] Güler İ, Übeyli ED. ECG beat classifier designed by combined neural network model. Pattern Recognition 2005; 38(2): 199-208
- [3] Wei Jiang, Seong G,Kong. Block-based neural networks for personalized ECG signal classification. IEEE Transactions on Neural Networks, 2007,8(6):1750-1761.
- [4] Ozbay Y, Ceylan R, Karlik B. A fuzzy clustering neural network architecture for classification of ECG arrhythmias. Computers in Biology & Medicine 2006; 36(4): 376-388
- [5] Paweł Pławiak, Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals, Swarm and Evolutionary Computation 39 (2018) 192–208.
- [6] Paweł Pławiak, Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system, Expert Syst. Appl. 2018. 92. 334–349.
- [7] S.-T. Pan, Y.-H. Wu, Y.-L. Kung, H.-C. Chen, Heartbeat recognition from ECG signals using hidden Markov model with adaptive features, in: 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013, 2013, pp. 586–591
- [8] Zhai W. A study on ECG classification based on convolutional neural network. Chinese Journal of Modern Computer, 25 (2019):32-35.
- [9] XL.Zhai and C.Tin, Automated ECG classification using dual heartbeat coupling based on convolutional neural network, IEEE Access, 2018. 6,pp: 27465–27472 .
- [10] Acharya U R, Oh S L, Hagiwara Y, et al. A deep convolutional neural network model to classify heartbeats. Computers in Biology and Medicine, 2017, 89: 389-396.
- [11] G.Sannino and G.De Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, Future Gener. Comput.Syst., 2018 vol. 86, pp. 446–455
- [12] Eduardo Jose da S. Luz, William Robson Schwartz, Guillermo Camara Chavez,David Menotti, ECG-based heartbeat classification for arrhythmia detection: a survey, Comput. Meth. Progr. Biomed. 127 (2016) 144–164.
- [13] K. Padmavathi, K. Sri Ramakrishna, Classification of ECG signal during atrial fibrillation using autoregressive modeling, Procedia Computer Science 46 (2015) 53–59.
- [14] Yang J L, Yang B, Lin F, et al. A novel electrocardiogram arrhythmia classification method based on stacked sparse auto encoders and softmax regression. Int J Mach Learn Cyb, 2018, 9 (10): 1733-1740.
- [15] Wang Z, Liu H, Shi J, Wang W. ECG recognition based on convolution neural network. Chinese Journal of Medical Physics. 2019,36(8):938-944.
- [16] Ozal Yildirim, A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification, Computers in Biology and Medicine. 2018 .96. pp: 189–202.
- [17] Özal Yıldırima,, Paweł Pławiak, Ru-San Tan, U. Rajendra Acharya, Arrhythmia detection using deep convolutional neural network with long duration ECG signals. Computers in Biology and Medicine, 2018. 102. pp:411-420.
- [18] Jinlong Ji, Xuhui Chen, Changqing Luo, A Deep Multi-task Learning Approach for ECG Data Analysis. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2018