IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON SCALABLE DEEP LEARNING FOR BIG DATA

Received July 14, 2019, accepted August 1, 2019, date of publication August 7, 2019, date of current version August 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933473

Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network

RUNCHUAN LI [1,2], XINGJIN ZHANG [2,3], HONGHUA DAI [2,4], (Member, IEEE), BING ZHOU [1,2], AND ZONGMIN WANG [1,2]

[1] School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China
[2] Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou 450000, China
[3] State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450003, China
[4] Institute of Intelligent Systems, Deakin University, Burwood, VIC 3125, Australia

Corresponding author: Zongmin Wang (zmwang@ha.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1401200, in part by the Key Research, Development, and Dissemination Program of Henan Province (Science and Technology for the People) under Grant 182207310002, in part by the Key Science and Technology Project of Xinjiang Production and Construction Corps, under Grant 2018AB017, and in part by the National Natural Science Foundation of China under Grant 61872324.

ABSTRACT Arrhythmia is a disease that threatens human life. Therefore, timely diagnosis of arrhythmia is of great significance in preventing heart disease and sudden cardiac death. The BiLSTM-Attention neural network model with heartbeat activity's global sequence features can effectively improve the accuracy of heartbeat classification. Firstly, the noise is removed by the continuous wavelet transform method. Secondly, the peak of the R wave is detected by the tagged database, and then the P-QRS-T wave morphology and the RR interval are extracted. This feature set is heartbeat activity's global sequence features, which combines single heartbeat morphology and 21 consecutive RR intervals. Finally, the Bi-LSTM algorithm and the BiLSTM-Attention algorithm are used to identify heartbeat category respectively, and the MIT-BIH arrhythmia database is used to verify the algorithm. The results show that the BiLSTM-Attention model combined with heartbeat activity's global sequence features has higher interpretability than other methods discussed in this paper.

INDEX TERMS Heartbeat activity's global sequence features, BiLSTM-attention neural network, interpretability, heartbeat classification.

I. INTRODUCTION

In recent years, with the improvement of people's living standards, the prevalence of cardiovascular diseases has increased significantly. The number of deaths caused by cardiovascular diseases has also increased year by year. As an important organ of the human body, the heart plays an important role in the function of the human body. Arrhythmia is a disorder of heart rate or a rhythm conduction and it is a critical manifestation of ECG abnormalities [1], [2]. Other heart diseases always come in hand with arrhythmia, such as myocardial infarction and heart failure. Therefore, accurate detection of the patients' arrhythmia plays an important role in preventing heart disease and sudden cardiac death. Electrocardiography is an important basis for preliminary diagnosis of arrhythmia [3]. Electrocardiography can be used to judge whether the arrhythmia is sinus or ectopic. By analyzing the nature and source of the early or delayed heartbeats one by one, ECG intelligent analysis can help doctors diagnose the nature of patients' arrhythmia. However, patients have to wear holter for a long time to monitor health and safety, because abnormal ECG signal is difficult to capture during some arrhythmic events.

The dynamic electrocardiogram record during arrhythmia attack is an important basis for the diagnosis of arrhythmia. Although sometimes patients have done a routine resting

The associate editor coordinating the review of this manuscript and approving it for publication was Liangxiu Han.

ECG, it is necessary to analyze the ECG record obtained by the 24-hour holter monitor according to different condition of patient. The dynamic electrocardiogram can continuously record about 100,000 heartbeats signal in 24 hours. 24-hour continuous ECG records make the diagnosis more accurate. Howeverčthe traditional dynamic ECG analysis is done manually. In this situation, the medical staff commonly observe the patient's ECG signal and then make a final diagnosis based on relevant rules and personal experience. Due to the large amount of ECG data and the shortage of medical staff, doctors who have been engaged in ECG classification and identification for a long time will inevitably have fatigue. In this case, mistakes, missed inspections or misdetections are easily occur. It is very troublesome for the doctor to judge and identify a large number of ECG one by one. Therefore, automated intelligent diagnosis is important in daily medicine [4]. It can help individuals make better judgments on the symptoms of arrhythmia. In addition, it can provide good health care in areas where medical resources are scarce.

In this paper, a new explanatory deep learning method is used for heartbeat classification, which based on heartbeat activity's sequence features and BiLSTM-Attention neural network. This method improves the accuracy of heartbeat classification. The contributions of this work are as follows:

1.In the case of unbalanced ECG data set, a novel neural network learning algorithm based on BiLSTM-Attention model is proposed for heartbeat classification.

2. Heartbeat activity's global sequence features proposed in this paper demonstrate the heartbeat category information more comprehensively.

3. The BiLSTM-Attention deep neural network model and heartbeat activity's global sequence features are used to classify various kinds of arrhythmias of different patients.

4. According to the heartbeat classification process, a general supervised learning framework that has high ability of automatic feature extraction and scalability is designed.

The rest of this paper is organized as follows: Section II gives a brief introduction to related works. Section III introduces the formation of the problem. Section IV introduces ECG signal preprocessing and heartbeat features extraction. Section V introduces the BiLSTM-Attention deep neural network model in detail. The electrocardiogram classification experiment is performed in Section VI. Section VII summarizes the full text and discusses the future work.

II. RELATED WORK

Traditionally, the diagnosis of early arrhythmia mainly depends on doctors' analysis of the ECG waveform. However, this method mainly relies on doctor's experience. At the same time, due to the diversity of arrhythmia and the complexity of the corresponding ECG waveform, manual analysis can not meet the needs of patients. With the development of artificial intelligence, the use of intelligent processing technology to classify arrhythmia has become a hot topic in recent years. However, due to the imbalance of data sets and individual patient differences, there are still some difficulties in achieving accurate heartbeat classification.

In the past few decades, several researches were developed to produce automatic ECG signal intelligent analysis. Researchers have proposed various methods for heartbeat classification, which can be divided into two categories: feature extraction and deep learning based methods.

A. HEARTBEAT CLASSIFICATION METHODS BASED ON FEATURE EXTRACTION

Feature extraction phase is the key to successful classification of arrhythmic heartbeat. Any information extracted from the heartbeat can be regarded as a feature as long as it can identify ECG heartbeat classification. Researchers design various features based on ECG signals and input them into machine learning models for decision making. The quality design of the feature depends on doctors and experts' clinical experience. Proposed features include morphological features [5], [6], temporal features [5]–[7], hermite basis function(HBF) [8], higher order statistics(HOS) [8], [9], and personalized features [10]. Many literatures of arrhythmia classification had been developed by using machine learning algorithms, such as k-nearest neighbor [11], support vector machine [8]–[11], conditional random fields [12]. Those machine learning models had been used to learn the difference between different heartbeats and to achieve the automatic classification of heartbeat [1], [13].

The features of manual designs mainly depend on the designer's prior knowledge, and it is difficult to take advantage of big data. Although these effective features can greatly improve the performance of the arrhythmia recognition system, these analysis methods are confined to the doctor's clinical experience. In clinical data, arrhythmia has many categories, great variability, and complex waveforms, which makes it difficult to accurately detect and locate waveforms and set classification features.

B. HEARTBEAT CLASSIFICATION METHODS BASED ON END-TO-END MANNER

The biggest difference between deep learning and traditional feature extraction method is that deep learning automatically learns the characteristics of ECG big data through multi-layer nonlinear transformation, and it replaces the features of manual design. Deep learning can quickly learn from training data and automatically obtain valid features, mining the information hidden behind ECG big data. Various hidden factors are often associated with complex non-linear approaches, but deep learning can separate these factors. The deep structure of deep learning makes it highly expressive and learning. It is especially good at extracting complex global features and making heartbeat classification simple and effective. Recently, some researchers [14]–[18] analyzed the application of deep learning methods of heartbeat classification to improve the accuracy. For example, some works [19]–[23] used convolutional neural networks to detect abnormal ECG signals. Zubair *et al.* [20] introduced an

ECG beat classification system using convolutional neural networks. By using a small and patient-specific training data, the classification system efficiently classified ECG beats into five different classes. Acharya *et al.* [23] developed a 9-layer deep convolutional neural network to automatically identify five different categories of heartbeats in ECG signals. Kiranyaz *et al.* [24] used an one-dimensional convolutional neural network to classify ECG signals. Hannun *et al.* [25] developed a deep neural network that consisted of 33 convolutional layers followed by a linear output layer into a softmax.

A few works [26]–[28] used recurrent neural network (RNN) model to classify heartbeat. On the basis of morphological information and temporal information, Wang *et al.* [26] applied a single RNN for automatic feature learning and classification. Zhang *et al.* [27] proposed a novel patient-specific ECG classification algorithm based on recurrent neural network and density clustering technique. Hochreiter and Schmidhuber [29] proposed a long-term short-term memory (LSTM) recurrent neural network model, which can effectively alleviate the long-distance dependence on RNN. Literature [30] used recurrent neural network (RNN), gated recurrent unit (GRU) and long short-term memory (LSTM) neural networks to classify heartbeat, and concluded that LSTM has better performance than RNN and GRU in detecting arrhythmia.

ECG diagnosis algorithm based on deep learning can identify and judge the arrhythmia event more effectively. It is important for modern medical treatment. It can better assist the medical staff to make a diagnosis and treatment program, effectively prevent the damage or even death caused by the heart disease, and improve the quality of patients' health. An important issue in ECG intelligent diagnosis is the accurate classification of each heartbeat, which directly affects the performance of the analysis system. Therefore, accurate classification of heartbeat is one of the important contents of ECG intelligent diagnosis, especially supraventricular ectopic beat and ventricular ectopic beat. In order to improve the accuracy of heartbeat classification, It is necessary to establish an interpretable and automated deep learning algorithm framework.

III. HEARTBEAT CLASSIFICATION PROBLEM FORMULATION

The heart has four physiological functions: self-discipline, excitability, conductivity, and contractility. Arrhythmia is caused by the self-discipline, excitability and abnormal conduction of the myocardium. In view of the continuous change of heart beat waveform, the computer can read ECG accurately and effectively, and gradually give the diagnosis results. Heartbeat classification results are the first level of computer intelligence diagnosis.

A. PROBLEM STATEMENT AND FORMULATION

The ECG heartbeat classification is a sequence of tasks that sort inputs that are ECG signals B = $[b_1, b_2, \ldots, b_n]$ and outputs a sequence of labels C = $[c_1, c_2, \ldots, c_5]$, each of which can take one of C different heartbeat classes. This problem is solved through designing a new neural network model in an end-to-end manner in the supervised learning framework. The *loss* function is the cross-entropy error of the ECG heartbeat classification, it can be expressed as (1). Where B is the training data, C is the number of ECG heartbeat categories, b means a beat, $p_c(b)$ is the probability of predicting b as class c given by the softmax layer, and $\hat{p}_c(b)$ indicates whether class c is the correct ECG heartbeat category, whose value is 1 or 0.

$$Loss = -\sum_{b \in B} \sum_{c=1}^{C} \hat{p}_c(b) \cdot \log(p_c(b)) \tag{1}$$

B. ARRHYTHMIA DATABASE AND CLASSIFICATION CRITERIA

The MIT-BIH arrhythmia database is a recognized authoritative ECG database for ECG heartbeat classification algorithms [35]. This paper evaluates the performance of the proposed ECG classification algorithm by using MIT-BIH arrhythmia database. Each record includes a half-hour 2-lead dynamic ECG segment with a sampling rate of 360 Hz. In this paper, non-redundant 2-lead ECG signals are used to learn different characteristics from different leads.

The ANSI/AAMI EC57: 2012 specific taxonomy developed by the Association for the Advancement of Medical Instruments (AAMI), which stipulates that ECG waveforms can be divided into five categories: N (normal or bundle branch block), S (supraventricular ectopic beat), V (ventricular ectopic beat), F (fusion beat) and Q (unassigned beat) [36]. As there are kinds of arrhythmia diseases, the computer can read ECG accurately and effectively and provide the diagnosis results, which needs to be carried out step by step. The classification result is the first step of computer intelligent diagnosis. However, the categories of ECG beat that are prescribed by AAMI and those annotated by the MIT-BIH arrhythmia database are two clinically relevant and differentiated classification methods. There are fifteen recommended classes for arrhythmia that are classified into five super classes. Based on this, the pathological changes of different heart parts of a wide variety of cardiovascular diseases can be judged, and it can help doctors in making specific clinical decisions on the next step. The ECG beat categories standard conversion relationship is shown in Table 1. Fig.1 shows annotation(L,V) in a MIT-BIH arrhythmia database. The common heartbeat category example is shown in Fig.2.

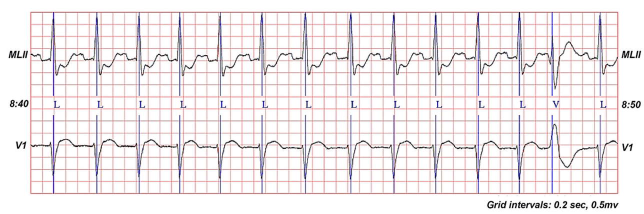

FIGURE 1. Example of annotation (L,V) in a MIT-BIH arrhythmia database record 109 (10 sec).

TABLE 1. **MIT database arrhythmia annotation and N, S, V, F, Q correspondence.**

MIT Database Heartbeat Comment Information		
AAMI provision of the heartbeat category	*N*	Normal beat (1), Left bundle branch block beat(2), Right bundle branch block beat(3), Nodal(junctional)escape beat(11), Atrial escape beat(34)
	S	Aberrated atrial premature beat(4), Nodal(junctional)premature beat(7), Atrial premature beat(8), Premature or ectopic supraventricular beat(9)
	V	Premature ventricular contraction (5), Ventricular escape beat(10)
	F	Fusion of ventricular and normal beat(6)
	Q	Paced beat (12), Unclassifiable beat(13), Fusion of paced and normal beat(38)

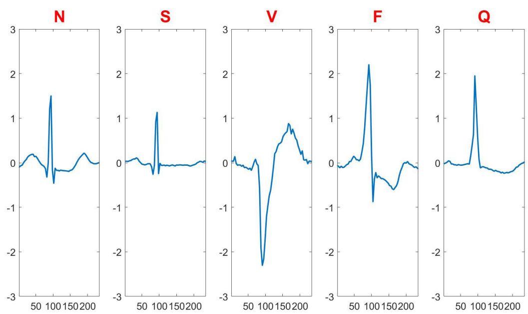

FIGURE 2. **Common heartbeat category example.**

IV. ECG SIGNAL PREPROCESSING AND HEARTBEAT FEATURES EXTRACTION

Each record of MIT-BIH arrhythmia database contains two leads, one of which (lead A) is lead II and the other lead (lead B) is lead V1. However, in some records, lead B is known to be V2, V5 or V4. In the process of collecting ECG signals, noise signals such as myoelectric interference and baseline drift are often mixed, which affects the analysis of ECG signals. In this paper, the optimal processing of ECG signal noise is realized based on continuous wavelet transform. Generally, lead A is used to detect heartbeats, since the QRS complex is more prominent in this lead. Lead B favors the arrhythmic classification of the types S and V [1]. According to the literature [1], [45], the prior knowledge extracted by combined with lead A and lead B.

A. ECG SIGNAL DENOISING

Wavelet transform is a method of time-frequency analysis of a signal [37]. Its biggest feature is the nature of multi-resolution analysis. It is a time-frequency localization analysis method

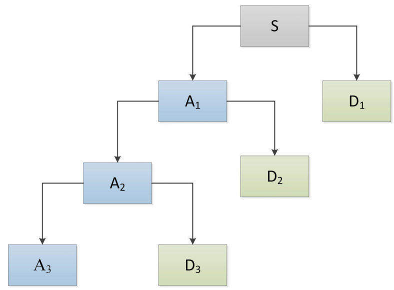

FIGURE 3. **The example of continuous wavelet transform decomposition (S is the original signal, A1, A2, A3 are the first order, second order, third order approximation signal, D1, D2, D3 are the first order, second order and third order).**

in which the window size is fixed. However its shape can be changed, and both the time window and the frequency window can be changed. That is to say, the method of wavelet analysis has the ability to have higher time resolution and lower frequency resolution to the high frequency part, higher frequency resolution and lower time resolution to the low frequency part. Therefore, wavelet transforms is also known as a mathematical microscope for analyzing signals. This characteristic makes the wavelet transform has adaptability to the signal and has its unique superiority in the time-varying signal. In this paper, the function of time-frequency local analysis and the multi-resolution analysis of wavelet transform are used to remove the noise of ECG signals.

$$\psi_{a,\tau}(t) = a^{-\frac{1}{2}}\psi\left(\frac{t-b}{a}\right), \quad a > 0,\ \tau \in R. \tag{2}$$

$$W_f(a,\tau) = a^{-\frac{1}{2}}\int_{-\infty}^{+\infty} f(t)\psi(\frac{t-b}{a})dt \tag{3}$$

The basic definitions of wavelet basis functions and continuous wavelet transform are shown in (2) and (3). Here, a is the scale factor, b is the transformation factor. Because a and b are continuously transformed values [38]. The original signal continuously passes through the filter bank. The filter bank has multiple levels, and each level obtain a set of wavelet components. The next layer filter to get the next layer of coefficients, the signal stratified filtering and the spectrum is divide into two equal segments: the low-pass and the high-pass. High-pass part contains only a small amount of signal details to be retained, the low-pass part still contains a lot of information details to continue filtering subdivision, and each level of output sampling rate can be further halved.

In this paper, continuous wavelet transform is used to denoise ECG signals. Firstly, noise signals is decomposed into multi-scale components, and threshold functions are set on small and high scales. Secondly, wavelet coefficients that smaller than threshold functions on minimum and maximum scales are removed. Finally, in order to remove the noise, the remaining wavelet coefficients are used to reconstruct the ECG signal. Fig.3 shows the example of continuous wavelet transform decomposition. Fig.4 shows the ECG record before denoise. Fig.5 shows the ECG record after denoise.

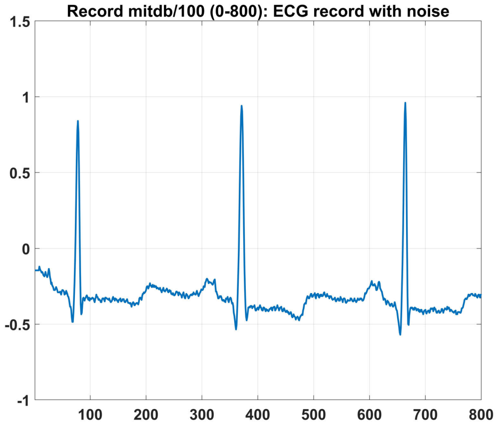

FIGURE 4. Record mitdb/100 (0-800): ECG signal with baseline drift and myoelectric interference.

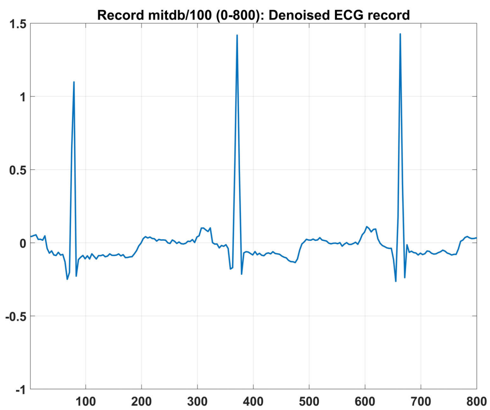

FIGURE 5. Record mitdb/100 (0-800): ECG signal after denoise.

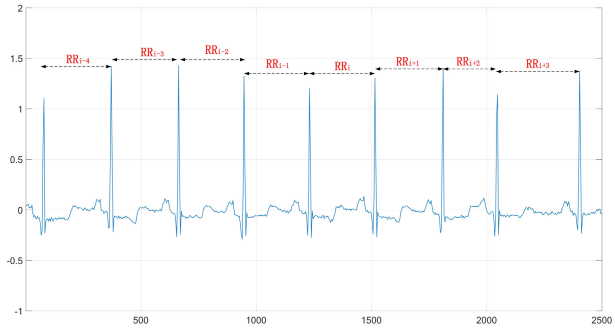

FIGURE 6. Heartbeat RR interval position diagram.

B. HEARTBEAT ACTIVITY'S GLOBAL SEQUENCE FEATURES

This paper uses the position of R peak that is determined in the annotation file to extract 235 points near R peak from ECG records according to the position of R peak [44]. There are 90 sampling points in front of R peak and 144 sampling points after R peak. If there are less than 235 sampling points before and after the first or last QRS complex wave detected in the ECG record file, the corresponding heartbeat is ignored. On this basis, the 235 sample points extracted as single heartbeat morphology features.

RR interval is a time interval between two consecutive R peaks. The normal distance from the RR interval is 0.6 to 1.0 seconds. The corresponding ventricular rate is 60 to 100 beats per minute. RR interval generally fluctuates in a certain range. Its sequence not only reflects the situation of heart rate variability, but also the fluctuation of abnormal interval can characterize some arrhythmic diseases. RR intervals are usually different in patients with arrhythmia, which are common in sinus arrhythmia, atrial premature beats, ventricular premature beats and atrioventricular block. Therefore, the RR interval is one of the important characteristics of the response to cardiac activity.

The physiological activities of cardiovascular system are regulated by autonomic nervous system. When the ability of autonomic nervous system to regulate the heart decreases, the malignant arrhythmia will occur. Literature [46] showed that the nature of malignant arrhythmia in patients with cardiovascular disease is the imbalance of sympathetic-vagus nerve. In 2012, the literature [47] proposed heart rate decelerations(DRs) based on the heart rate deceleration of the autonomic nerve monitoring method. If there are 2 to 10 cardiac cycles with sustained slow beats, the prognosis of patients is better. Based on this clinical conclusion, more comprehensive boundary factors are considered in the experiment. Therefore, using the 10 RR cycles before and after the current RR interval, a total of 21 consecutive RR cycles are selected. In summary, the 21 consecutive RR intervals extracted as continuous heartbeat interval activity features.

Heartbeat activity's global sequence features contain both intra-cardiac and inter-cardiac features, which are single heartbeat morphology and 21 consecutive RR intervals. Fig.6 shows heartbeat RR interval position diagram. Fig.7 shows RR interval time series.

V. BILSTM-ATTENTION NEURAL NETWORK MODEL DESCRIPTION

The processing of time-series tasks is important in the field of natural language processing and speech recognition. To accommodate this need, a new recurrent neural network learning architecture is proposed, which can increases the temporal structure of the structure [39]. The output can be directly applied to itself at the next time stamp. RNN can be regarded as a deep neural network that passes through in time. Its depth is the length of time. As the time interval increases, the gradient of the RNN unit disappears and only short-term memory can be maintained. LSTM unit combines short-term memory with long-term memory by introducing memory unit and gate control unit, which solves the problem of gradient disappearance to a certain extent. Therefore, literature [29] proposes a Long Short-Term Memory that implements temporal memory through a gate switch that selectively passes information and prevents gradients from disappearing. LSTM is a time recurrent neural network that effectively preserves historical information and analyzes sequence data [32], [40].

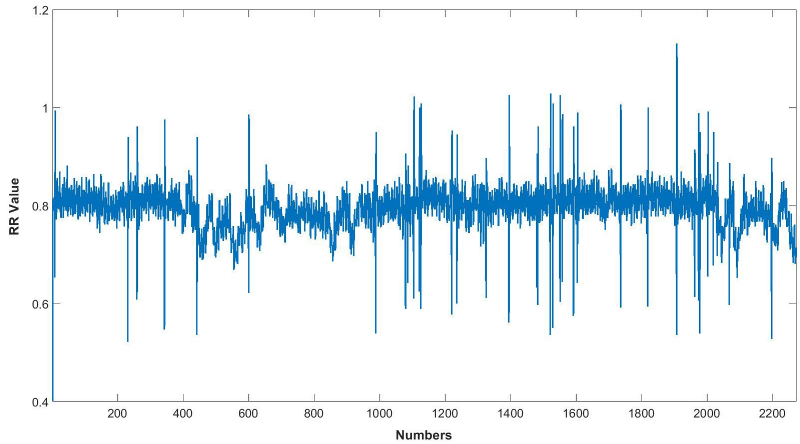

FIGURE 7. Record mitdb/100 (1-2273 numbers): RR interval time series.

A. BILSTM NEURAL NETWORK STRUCTURE

LSTM consists of three gates (input gate, forget gate and output gate) and a cell unit. It aims at achieving the historical information updates and reservations. It can add and delete information on cells through the gate unit. The input gate decides how much new information to add to the state, forget gate determines the information that needs to be retained and discarded, and the output gate determines which part of the information will be output. The gate can selectively determine whether the information passes, it has a sigmoid neural network layer and a pairwise multiplication operation. The structure of bidirectional LSTM (Bi-LSTM) model provides complete past and future context information for each point in the output layer input sequence. Among them, C_{t-1} and C_t respectively represent the state value of the memory unit at the previous moment and the current time, $\overline{C}_t$ represents the current heartbeat state candidate value, h_{t-1} and h_t respectively represent the output of the previous time and the current LSTM network, and x_t represents the current input, f_t, i_t, and o_t denotes the forget gate, the input gate and the output gate, respectively, μ denotes the sigmoid function, and denotes the hyperbolic tangent function. Fig.8 shows the LSTM cell structure.

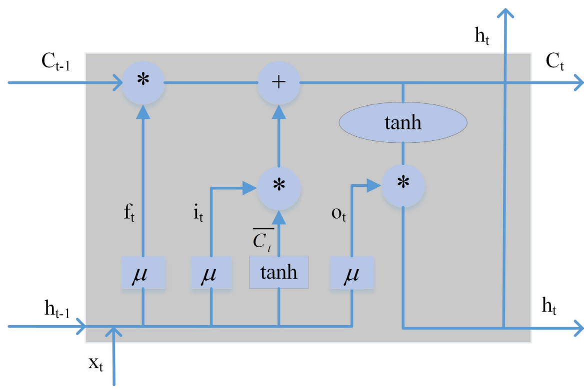

FIGURE 8. The LSTM cell structure.

The LSTM cell calculation process is shown in the following formula. At times t, the input gate is input according to the output result h_{t-1} of the cell at the previous moment. The input x_t at the current moment, determines whether to update the current information into the cell through calculation. Forget gate based on the last moment hidden layer output h_{t-1} and the current time input as input to decide the need to retain and discard the information to achieve the storage of historical information. The current candidate memory cell value is determined by the current input data x_t and the output result h_{t-1} of the LSTM hidden layer cell at the previous moment. In the current moment, the memory cell state value C_t is adjusted by both the current candidate cell $\overline{C}_t$ and its own state C_{t-1} and input gate and forget gate. Calculate the output gate o_t, the output used to control the memory cell status value. The output of the last cell is h_t, which can be expressed as (9). Character $*$ is the element-wise matrix multiplication, while character $\cdot$ denotes point multiplication. W is the weight, and b is the bias of neuron, where W and b are both obtained through training. (4)-(9) refer to the literature [40].

$$i_t = sigmoid(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$
$$f_t = sigmoid(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$
$$\overline{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$
$$C_t = f_t * c_{t-1} + i_t * \overline{c}_t \tag{7}$$
$$o_t = sigmoid(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$
$$h_t = o_t * \tanh(c_t) \tag{9}$$

If the algorithmic model can access the future contexts as it did the past context information, it has far-reaching implications for sequence learning. The standard LSTM cell is used to process sequence data. Since data is processed in time series, it often ignores future context information. The basic idea of the Bi-LSTM is that each training sequence consists of forward and backward LSTM neural network layers.

The forward LSTM layer encodes the heartbeat from beginning to end, and the backward LSTM layer encodes the opposite direction. Therefore, the hidden layer state of BiLSTM at times t is obtained by weighted summation of the forward hidden layer state $\overrightarrow{h_t}$ and the backward hidden layer state $\overleftarrow{h_t}$, and the specific formula is as follows:

$$\overrightarrow{h_t} = LSTM(x_t, \overrightarrow{h_{t-1}}) \tag{10}$$

$$\overleftarrow{h_t} = LSTM(x_t, \overleftarrow{h_{t-1}}) \tag{11}$$

$$H_t = w_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t \tag{12}$$

w_t, v_t respectively represent the weights corresponding to the forward hidden layer state $\overrightarrow{h_t}$ and the backward hidden layer state $\overleftarrow{h_t}$ corresponding to the BiLSTM hidden layer state, and b_t represents the bias corresponding to the hidden layer state at time t.

B. BI-DIRECTIONAL LSTM BASED ON ATTENTION MECHANISM NEURAL NETWORK MODEL

Mnih et al. [31] proposed a model simulates the mechanism of human brain attention. It highlights key inputs by weight and thus optimizes the traditional model. The principle is to selectively focus on the corresponding information about the input when the model is output. The method using the attention mechanism is widely used in time series classification [32], which includes automatic text generation [33], text summarization [34], and so on. The attention mechanism breaks the limitation of the traditional encoder-decoder structure that relies on a fixed-length vector internally during encoding and decoding. It is implemented by retaining the intermediate output of the input sequence by the Bi-LSTM encoder, and then training a model to selectively learn these inputs. The attention mechanism simulates the characteristics of human brain attention. The core idea is to assign more attention to what it considers important and less attention to other parts. The input of the attention mechanism layer is the output vector of the upper layer activated by the Bi-LSTM neural network layer. (13)-(15) refer to the literature [40], the formula for the attention mechanism layer is as follows:

$$u_t = \tanh(W_w H_t + b_w) \tag{13}$$

$$a_t = \frac{\exp u_t^T u_w}{\sum_t \exp u_t^T u_w} \tag{14}$$

$$v_t = \sum\nolimits_t a_t H_t \tag{15}$$

where H_t is the output vector of the upper layer of the Bi-LSTM neural network layers, W_w is the weight coefficient, b_w is the bias coefficient, and u_t is the energy value determined by H_t. The a_t is the weighting coefficient of the specific gravity of each hidden layer state in the new hidden layer state. The u_w is an attention matrix indicating random initialization, which is continuously learned during the training process. The v_t is output vector through the attention mechanism. After the weight coefficient is calculated by the attention layer, the vector v_t is output to the dense layer, and the dense layer is received and processed by the rectified linear unit (*Relu*) function [41]. Because the linear model has insufficient expressive power, the activation function is used to add nonlinear factors. The *Relu* function is the most commonly used activation function in neural networks. The *Relu* function causes the output of some neurons to be 0, whose results in the sparseness of the network and the interdependence of parameters are reduced, which alleviates the occurrence of over-fitting and reduces the training time of deep networks. Finally, the input is calculated by the *softmax* function, and the final result are output.

In the Bi-LSTM model, the output vector of the last time series is usually used as the feature vector for the next layer and then input to the classified *softmax* function. However, this method of feature extraction method only uses the features of the last step, discarding other feature information. Therefore, the BiLSTM-Attention model is used in this experiment. The model adds the attention layer to the Bi-LSTM model. The attention mechanism first calculates the weight of each time series, then weights the vectors of all time series, and then inputs the weighted average vector as a new feature vector into the *softmax* function for classification. The BiLSTM-Attention deep neural network hybrid model designed in this paper considers the information above and below the heartbeat and the key position of the heartbeat. The Bi-directional LSTM model based on attention mechanism is shown in Fig.9.

VI. EXPERIMENTS AND RESULTS

ECG signals are crucial to the treatment of patients. This paper focus on the three main steps: preprocessed data, P-QRS-T wave group extraction, RR interval calculation, and heartbeat classification. Every heartbeat is categorized into N (Normal or bundle branch block), S (Supraventricular abnormal beat), V (Ventricular abnormal beat), F (Fusion beat), Q (Unclassified beat) according to ANSI / AAMI EC 57. Therefore, this classification result can be used as preliminary results of computer diagnosis developed by the American Medical Instrument Promotion Association. Based on this, it is better to help doctors make the next specific clinical decision for this part. The framework of the heartbeat classification algorithm in Fig.10.

A. EXPERIMENTAL DATA

As can be seen from Table 2, 109,454 heartbeats in the MIT-BIH arrhythmia database are classified. This paper randomly used 90% of the data for the training dataset and 10% of the data for the testing dataset. 90,595 are recorded by the experts as N category heartbeat, among them, 81,560 heartbeats are used for training dataset and 9,035 heartbeats are used for testing dataset. 2,781 heartbeats are recorded as S category, among them, 2,528 heartbeats are used for training dataset and 253 heartbeats are used for testing dataset. 7,235 heartbeats are recorded as V category, among them, 6,450 heartbeats are used for training and 785 heartbeats are used for testing. Only 802 heartbeats are recorded as

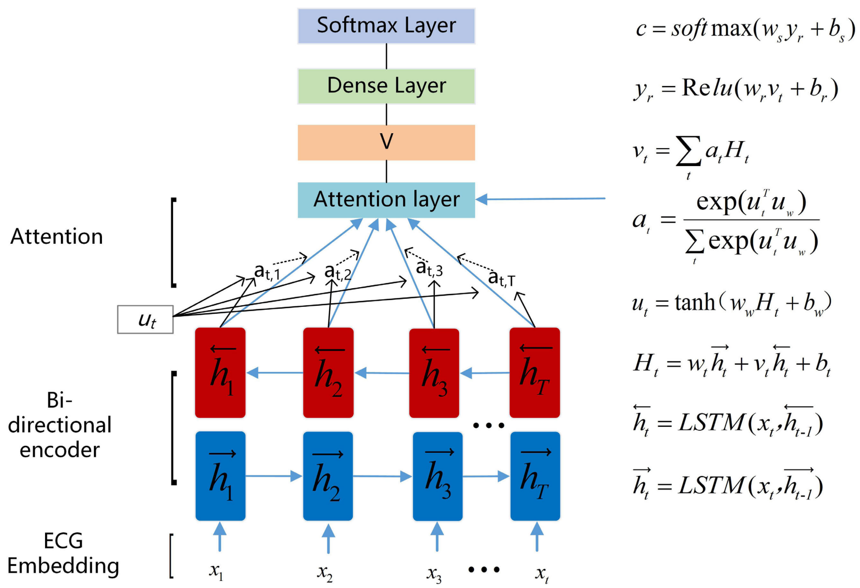

FIGURE 9. Bi-directional LSTM model based on attention mechanism.

TABLE 2. Experimental data statistics.

		Training dataset	Testing dataset	Total
category	N	81,560	9,035	90,595
	S	2,528	253	2,781
	V	6,450	785	7,235
	F	723	79	802
	Q	7,248	793	8,041

F category, among them, 723 heartbeats are for training and 79 heartbeats are for testing. And 8,041 heartbeats are recorded as Q category, among them, 7,248 heartbeats are used for training and 793 heartbeats are used for testing. It can be seen that the experiments are carried out under an unbalanced ECG dataset.

B. MODEL TRAINING AND PARAMETERS SELECTION

The experiment is trained as a PC workstation with an i7-7700K processor and 32GB of RAM. The algorithm is developed in tensorflow-gpu V1.11.0, keras V2.2.4 and matlab@R2018b software platform. In order to get a better model, the selection of model parameters is very important. Literature [43] analyzed the effects of parameters on model performance and suggestd that other scholars refer to the values are given by them.

In the model proposed in this paper, according to the experience of setting the same parameters in the deep learning algorithm, the learning rate, batch size, cell size and epoch parameters are initialized. Table 3 shows all parameters that are used in the BiLSTM-Attention model. The training process of the neural network model is performed by a

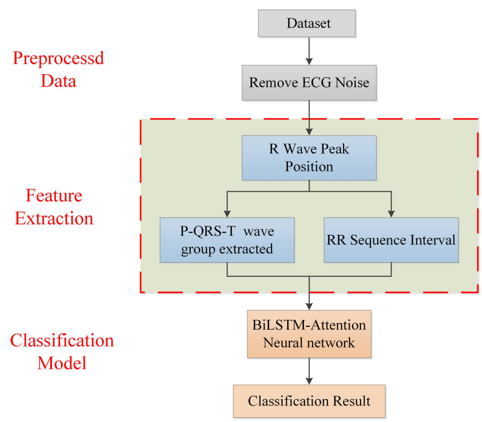

FIGURE 10. The framework of the heartbeat classification algorithm.

TABLE 3. The parameters of model.

Parameter	Value
Learning rate	0.001
Batch size	64
Input size	1
Time step	491
Cell size	128
Epoch	20
Output size	5

back propagation method. Learning rate is one of the most parameters affecting performance. Adam is an optimization algorithm used to replace the random gradient in the deep learning model. The recommended learning rate parameter is 0.001 [42]. The cell size indicates how many LSTM cells are to be placed in the hidden layer of the network. The experimental results show that the cell size is 128, the accuracy is

TABLE 4. Parameter cell size selection(batch size = 64, epoch = 20).

Cell size	8	16	32	64	**128**	256
Acc(%)	99.16	99.14	99.24	99.32	**99.49**	99.31

TABLE 5. Parameter batch size selection(cell size = 128, epoch = 20).

Batch size	8	16	32	**64**	128	256
Acc(%)	99.44	99.44	99.41	**99.49**	99.39	99.28

the highest. The experimental results are shown in Table 4. The batch size represents the number of pieces trained in each batch. The experimental results are fairly accurate and the batch size is 64. The experimental results are shown in Table 5. The time step is the data of how many time points to read. The time step of this experiment is 491. The input size represents the number of reads per row with a value of 1 and the output size represents the length of the classification result with a value of 5. Epoch represents the number of iterations. It is experimentally found that the epoch value is 20 with the highest accuracy, the experimental results are shown in Fig.11.

Algorithm 1 Training of BiLSTM-Attention Model Based Classfier

Input: Training set D = (b_1,c_1), (b_2,c_2),…,(b_n,c_n) ;
Base learning algorithm BiLSTM-Attention model.
Output: The parameters of the model and result.
Process:
begin
Build the BiLSTM-Attention model with a softmax output layer
while training **do**
begin
Calculate the loss on the training set according to (1)
Train the BiLSTM-Attention using the adam back propagation method
Evaluate the training loss on training set
Record the prediction of classifier on training set
if training loss stop decreasing **then**
Store the model and break
end
end

C. EVALUATION METRICS

As shown in (16)-(19), in order to achieve a category heartbeat classification results, the following formulas are needed: N category true positive heartbeat (TP_N), N category false positive heartbeat (FP_N), N category true negative heartbeat (TN_N), N category false negative heartbeat (FN_N). In the same way, the classification results of other category are calculated. Table 6 shows the confusion matrix of

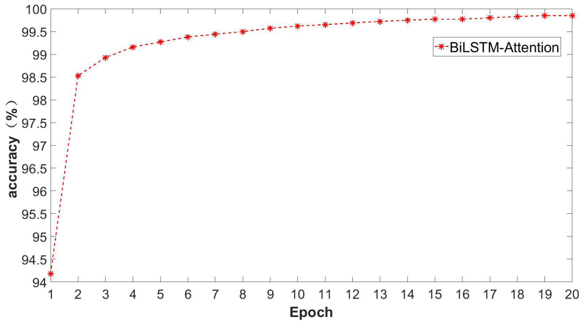

FIGURE 11. Epoch value parameter selection.

TABLE 6. The confusion matrix of classification results.

		Forecast Category				
Label category		**n**	**s**	**v**	**f**	**q**
	N	Nn	Ns	Nv	Nf	Nq
	S	Sn	Ss	Sv	Sf	Sq
	V	Vn	Vs	Vv	Vf	Vq
	F	Fn	Fs	Fv	Ff	Fq
	Q	Qn	Qs	Qv	Qf	Qq

classification results.

$$TP_N = Nn \tag{16}$$
$$FN_N = Ns + Nv + Nf + Nq \tag{17}$$
$$TN_N = Ss + Sv + Sf + Sq + Vs + Vv + Vf + Vq + Fs + Fv + Ff + Fq + Qs + Qv + Qf + Qq \tag{18}$$
$$FP_N = Sn + Vn + Fn + Qn \tag{19}$$

In this paper, sensitivity, specificity, positive predictivity, and accuracy are used as indicators to evaluate the performance of classifiers. Sensitivity (se) refers to the proportion of actual positive samples that judged to be positive. The higher the sensitivity, the greater the proportion of correct prediction. Specificity (sp) refers to the proportion of samples that are actually negative and are judged to be negative. The higher the specificity, the more accurate the prediction. Positive predictivity ($+p$) is also known as precision. Accuracy (Acc, also known as efficiency) is the ratio of the sum of true positives to true negatives to the number of subjects, which reflects the consistency between test results and actual results. The calculation formulas for the above four evaluation indicators are as follows:

$$Se = TP/(TP + FN) \tag{20}$$
$$Sp = TN/(TN + FP) \tag{21}$$
$$+p = TP/(TP + FP) \tag{22}$$
$$Acc = (TP + TN)/(TP + TN + FP + FN) \tag{23}$$

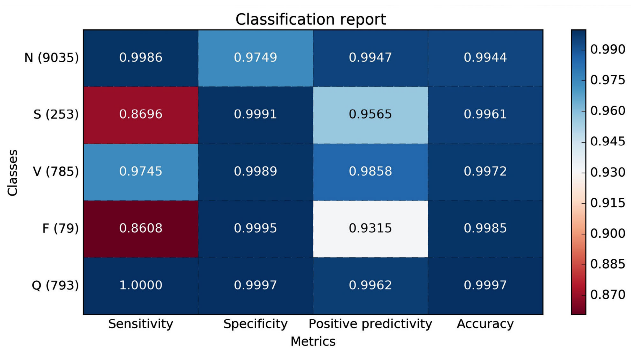

FIGURE 12. Confusion matrix based on classification results of BiLSTM-Attention model with single heartbeat activity features.

TABLE 7. BiLSTM-attention model with single heartbeat morphology features of classification results specific category statistics.

		Forecast Category					
		n	s	v	f	q	total
Labelcategory	N	9,022	7	3	2	1	9,035
	S	31	220	2	0	0	253
	V	13	3	765	3	1	785
	F	4	0	6	68	1	79
	Q	0	0	0	0	793	793

D. EXPERIMENTAL RESULTS AND ANALYSIS

In order to compare the differences between the designed features and to have a better understand of the calculation process of the BiLSTM-Attention model, the following four groups of experiments are conducted for comparative analysis.

Experiment I only collects 235 feature points in a single heartbeat. The experiment is performed on this single heartbeat morphology features of the average classification accuracy of 99.30%, but the single heartbeat morphology features have certain locality. The shortcoming of this experimental method is that the single heartbeat morphology features are not comprehensive enough. Fig.12, Table 7 and Table 8 show BiLSTM-Attention model with single heartbeat morphology features of classification results.

In experiment II, only 21 RR intervals are collects as continuous heartbeat interval activity features. The classification accuracy of N, V, F and Q classes is obviously lower. The average classification accuracy reached 96.94%. Fig.13, Table 9 and Table 10 show BiLSTM-Attention model with continuous heartbeat interval activity features of classification results. The shortcoming of this experimental method is that the continuous heartbeat interval activity features are not sufficiently specific.

Experiment III is the result of heartbeat classification using Bi-LSTM model without attention mechanism. The experiment collects heartbeat activity's global sequence features. Considering the overall time series relationship between heartbeat and heartbeat, this information is sufficient and comprehensive enough. The experimental results show that

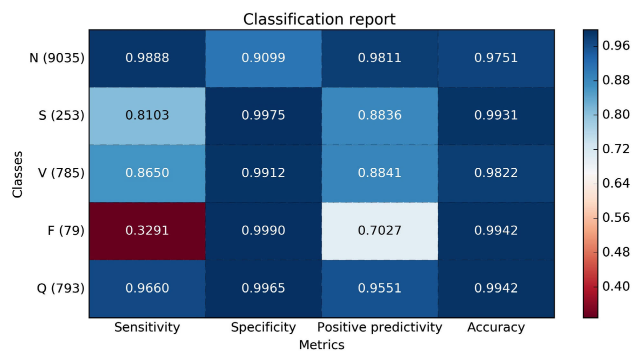

FIGURE 13. Confusion matrix based on classification results of BiLSTM-attention model with continuous heartbeat interval activity features.

TABLE 8. BiLSTM-attention model with single heartbeat morphology features of classification results.

		Evaluation Metrics (%)							
		TP	TN	FP	FN	Se	Sp	+p	Acc
Labelcategory	N	9,022	1,862	48	13	99.86	97.49	99.47	99.44
	S	220	10,682	10	33	86.96	99.91	95.65	99.61
	V	765	10,149	11	20	97.45	99.89	98.58	99.72
	F	68	10,861	5	11	86.08	99.95	93.15	99.85
	Q	793	10,149	3	0	100	99.97	99.62	99.97

TABLE 9. BiLSTM-attention model with continuous heartbeat interval activity features of classification results specific category statistics (21 consecutive RR intervals).

		Forecast Category					
		n	s	v	f	q	total
Labelcategory	N	8,934	15	51	6	29	9,035
	S	21	205	27	0	0	253
	V	82	12	679	5	7	785
	F	42	0	11	26	0	79
	Q	27	0	0	0	766	793

TABLE 10. BiLSTM-attention model with continuous heartbeat interval activity features of classification results.

		Evaluation Metrics (%)							
		TP	TN	FP	FN	Se	Sp	+p	Acc
Labelcategory	N	8,934	1,738	172	101	98.88	90.99	98.11	97.51
	S	205	10,665	27	48	81.03	99.75	88.36	99.31
	V	679	10,071	89	106	86.50	99.12	88.41	98.22
	F	26	10,855	11	53	32.91	99.90	70.27	99.42
	Q	766	10,116	36	27	96.60	99.65	95.51	99.42

the accuracy of the method is 99.32%, the accuracy of S category is 99.70%, the accuracy of V category is 99.71%, and the accuracy of F category is 99.84%. Fig.14, Table 11 and Table 12 show Bi-LSTM model with heartbeat activity's global sequence features of classification results. In the

TABLE 11. Bi-LSTM model with heartbeat activity's global sequence features of classification results specific category statistics (single heartbeat morphology and 21 consecutive RR intervals).

	Forecast Category						
		n	s	v	f	q	total
Labelcategory	N	9,020	6	6	3	0	9,035
	S	21	227	4	1	0	253
	V	12	1	769	3	0	785
	F	3	0	6	69	1	79
	Q	7	0	0	0	786	793

TABLE 12. Bi-LSTM model with heartbeat activity's global sequence features of classification results.

	Evaluation Metrics (%)								
		TP	TN	FP	FN	Se	Sp	+p	Acc
Labelcategory	N	9,020	1,867	43	15	99.83	97.75	99.53	99.47
	S	227	10,685	7	26	89.72	99.93	97.01	99.70
	V	769	10,144	16	16	97.96	99.84	97.96	99.71
	F	69	10,859	7	10	87.34	99.94	90.79	99.84
	Q	786	10,151	1	7	99.12	99.99	99.87	99.93

TABLE 13. BiLSTM-attention model with heartbeat activity's global sequence features of classification results specific category statistics (single heartbeat morphology and 21 consecutive RR intervals).

	Forecast Category						
		n	s	v	f	q	total
Labelcategory	N	9,022	6	5	0	2	9,035
	S	18	233	2	0	0	253
	V	9	1	773	2	0	785
	F	3	0	7	69	0	79
	Q	0	0	1	0	792	793

TABLE 14. BiLSTM-attention model with heartbeat activity's global sequence features of classification results.

	Evaluation Metrics (%)								
		TP	TN	FP	FN	Se	Sp	+p	Acc
Labelcategory	N	9,022	1,880	30	13	99.86	98.43	99.67	99.61
	S	233	10,685	7	20	92.09	99.93	97.08	99.75
	V	773	10,145	15	12	98.47	99.85	98.10	99.75
	F	69	10,864	2	10	87.34	99.98	97.18	99.89
	Q	792	10,150	2	1	99.87	99.98	99.75	99.97

experiment, the heartbeat activity's global sequence features are considered, which is sufficient improves the classification accuracy of the model.

Experiment IV is the result of heartbeat classification using the BiLSTM-Attention model, which uses the same features set as Experiment III. The experimental results verify the feasibility and effectiveness of the proposed model. The average accuracy of the classification reached 99.49%. The results show that the accuracy of the S category is 99.75%, the accuracy of V category is 99.75%, and the accuracy of F category is 99.89%. Compared with Experiment III, the overall

TABLE 15. Comparisons of the classification results with the previous studies.

Work	Feature set	Classifier	Performance
Lin et al. (2014) [5]	RR-interval, Morphological feature	Linear Discriminant	Acc=93.0%; Se_n=94.6%; $+P_n$=98.9%; Se_s=78.2%; $+P_s$=24.1%; Se_v=77.8%; $+P_v$=79.7%
Kiranyaz et al. (2016) [24]	ECG morphology	1-D CNN	Se_s=60.3%; Sp_s=99.2%; $+P_s$=63.5%; Se_v=93.9%; Sp_v=98.9%; $+P_v$=90.6%
Acharya et al. (2017) [23]	ECG morphology	9-Layer CNN	Acc=94.03%; Se_n=91.54%; Sp_n=96.71%; $+P_n$=87.43%; Se_s=90.59%; Sp_s=98.63%; $+P_s$=94.30%; Se_v=94.22%; Sp_v=98.84%; $+P_v$=95.30%
He et al. (2018) [17]	ECG morphology	LiteNet, AlexNet, GoogleNet, SqueezeNet, MobileNets	Acc=98.80%; Acc=98.83%; Acc=99.01%; Acc=98.29%; Acc=98.20%
Shraddha et al. (2018) [30]	ECG morphology	LSTM	Acc=88.1%
Proposed	ECG morphology, 21 RR intervals	BiLSTM-Attention	Acc=99.49%; Se_n=99.86%; Sp_n=98.43%; $+P_n$=99.67%; Se_s=92.09%; Sp_s=99.93%; $+P_s$=97.08%; Se_v=98.47%; Sp_v=99.85%; $+P_v$=98.10%

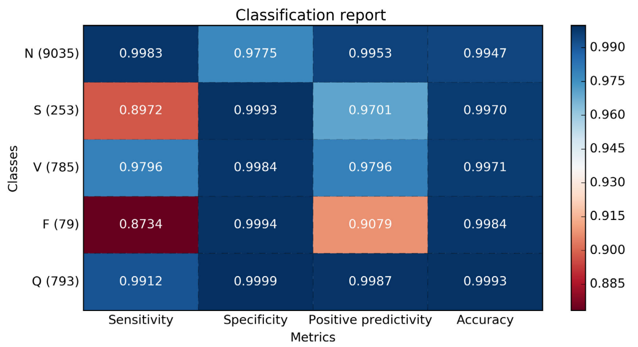

FIGURE 14. Confusion matrix based on classification results of Bi-LSTM model with heartbeat activity's global sequence features.

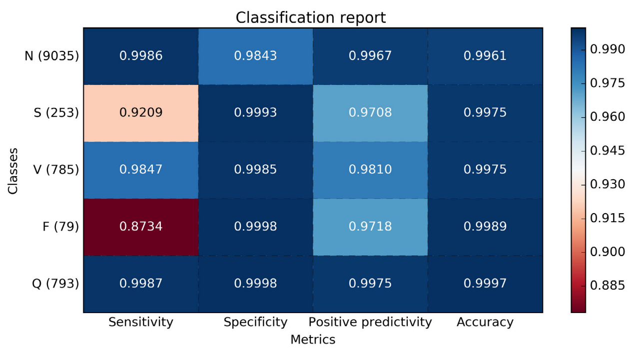

FIGURE 15. Confusion matrix based on classification results of BiLSTM-attention model with heartbeat activity's global sequence features.

accuracy of Experiment IV is increased by 0.17%. Compared with experiment I and experiment II, it is concluded that making full use of the prior knowledge of ECG is the key to further improve the performance of the model. Fig.15, Table 13 and Table 14 show BiLSTM-Attention model with heartbeat activity's sequence features of classification results.

Compared with all the experiments in Table 15, the identification results of the five categories of heartbeat are relatively stable. Total accuracy of our proposed method for this scheme is 99.49%. Literature [30] classified normal and abnormal heartbeats using LSTM model in MIT-BIH arrhythmia database. However, it does not take into account the influence of heartbeat activity's global sequence features on the category of rhythm. The LSTM model is used to process sequence data, which often ignoring future contextual information.

In this paper, BiLSTM-Attention model is used to merge heartbeat activity's global sequence features under unbalanced samples, which improves the accuracy of heartbeat classification. The experimental results show that the scheme has the advantages of distinguishing normal or bundle branch block, supraventricular abnormal beat, ventricular abnormal beat, fusion beat. The network connection among neurons is realized by deep learning algorithm, and then a interpretable, accurate and objective model for calculating cardiac activity is established. Therefore, it has obvious clinical significance and practicability for arrhythmia.

VII. CONCLUSION

This paper proposes a new framework about ECG heartbeat classification. This framework can simulate the thinking process of medical experts in diagnosing diseases, and automatically learn the characteristics of heartbeat categories. The significance of this study is to provide better clinical monitoring, diagnosis and treatment for heart disease patient.

Highlights of this paper are listed as follows:

(1) The BiLSTM-Attention model is used to extract and describe the feature of ECG automatically to learn the potential correlation between a individual heartbeat internal data and the relationship of the different individual heartbeats in massive ECG data.

(2) The interpretability analysis of the learning content of the BiLSTM-Attention algorithm model is carried out by constructing feature sets of different ECG prior knowledge.

(3) The experimental results show that the BiLSTM-Attention model with heartbeat activity's global sequence features can effectively simplify the feature extraction procedure and improve the accuracy of heartbeat classification.

However, in order to achieve higher accuracy, this method needs a large amount of ECG data. Because of the complexity of BiLSTM-Attention model, the time complexity of the algorithm is greatly increased. To ensure the real-time performance of the BiLSTM-Attention algorithm, future research will focus on cloud computing and parallel programming technology.

VIII. COMPETING INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, Apr. 2016.

[2] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, Sep. 2018.

[3] R. E. Mason, I. Likar, R. O. Biern, and R. S. Ross, "Multiple-lead exercise electrocardiography: Experience in 107 normal subjects and 67 patients with angina pectoris, and comparison with coronary cinearteriography in 84 patients," *Circulation*, vol. 36, no. 4, pp. 517–525, 1967.

[4] H. Li, D. Yuan, X. Ma, D. Cui, and L. Cao, "Genetic algorithm for the optimization of features and neural networks in ECG signals classification," *Sci. Rep.*, vol. 7, Jan. 2017, Art. no. 41011.

[5] C. C. Lin and C. M. Yang, "Heartbeat classification using normalized RR intervals and wavelet features," in *Proc. IEEE Comput. Soc. Int. Symp. Comput.*, Jun. 2014, pp. 650–653.

[6] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, pp. 79–89, Mar. 2014.

[7] H. Huang, J. Liu, Q. Zhu, R. Wang, and G. Hu, "A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals," *Biomed. Eng. Online*, vol. 13, no. 1, p. 90, Dec. 2014.

[8] K. S. Park, B. H. Cho, D. H. Lee, S. H. Song, J. S. Lee, Y. J. Chee, I. Y. Kim, and S. I. Kim, "Hierarchical support vector machine based heartbeat classification using higher order statistics and Hermite basis function," in *Proc. IEEE Comput. Cardiol.*, Sep. 2008, pp. 229–232.

[9] G. De Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted SVMs and feature relevance assessment in supervised heart beat classification," in *Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2011, pp. 212–223.

[10] C. Ping and X. Dong, "Life-threatening ventricular arrhythmia detection with personalized features," *IEEE Access*, vol. 5, pp. 14195–14203, 2017.
[11] E. Alickovic and A. Subasi, "Effect of multiscale PCA de-noising in ECG beat classification for diagnosis of cardiovascular diseases," *Circuits, Syst., Signal Process.*, vol. 34, no. 2, pp. 513–533, Feb. 2015.
[12] G. De Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 241–247, Oct. 2011.
[13] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey," in *Proc. Comput. Eng. Appl.*, Mar. 2015, pp. 714–721.
[14] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Comput. Biol. Med.*, vol. 99, pp. 53–62, Aug. 2018.
[15] A. Majumdar and R. Ward, "Robust greedy deep dictionary learning for ECG arrhythmia classification," in *Proc. Int. Joint Conf. Neural Netw.*, 2017, pp. 4400–4407.
[16] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky, "Deep learning for ECG classification," *J. Phys., Conf. Ser.*, vol. 913, no. 1, 2017, Art. no. 012004.
[17] Z. He, X. Zhang, Y. Cao, Z. Liu, B. Zhang, and X. Wang, "LiteNet: Lightweight neural network for detecting arrhythmias at resource-constrained mobile devices," *Sensors*, vol. 18, no. 4, p. 1229, 2018.
[18] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Jun. 2016.
[19] 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," *Inf. Sci.*, vol. 405, pp. 81–90, Sep. 2017.
[20] M. Zubair, J. Kim, and C. Yoon, "An automated ECG beat classification system using convolutional neural networks," in *Proc. IEEE 6th Int. Conf. IT Converg. Secur. (ICITCS)*, Sep. 2016, pp. 1–5.
[21] X. Fan, R. Chen, C. He, Y. Cai, P. Wang, and Y. Li, "Toward automated analysis of electrocardiogram big data by graphics processing unit for mobile health application," *IEEE Access*, vol. 5, pp. 17136–17148, Aug. 2017.
[22] Y. Wu, F. Yang, Y. Liu, X. Zha, and S. Yuan, "A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification," in *Proc. Conf. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 2018, pp. 324–327.
[23] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
[24] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Aug. 2016.
[25] Y. H. Awni, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, Jan. 2019.
[26] G. Wang, C. Zhang, Y. Liu, H. Yang, D. Fu, H. Wang, and P. Zhang, "A global and updatable ECG beat classification system based on recurrent neural networks and active learning," *Inf. Sci.*, vol. 501, pp. 523–542, Oct. 2019.
[27] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang, "Patient-specific ECG classification based on recurrent neural networks and clustering technique," in *Proc. IASTED Int. Conf. Biomed. Eng.*, Feb. 2017, pp. 63–67.
[28] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9393–9400, Sep. 2018.
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[30] S. Shraddha, S. K. Pandey, U. Pawar, and R. R. Janghel, "Classification of ECG arrhythmia using recurrent neural networks," *Procedia Comput. Sci.*, vol. 132, pp. 1290–1297, Jan. 2018.
[31] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2204–2212.
[32] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.
[33] H.-T. Zheng, W. Wang, W. Chen, and A. K. Sangaiah, "Automatic generation of news comments based on gated attention neural networks," *IEEE Access*, vol. 6, pp. 702–710, 2017.
[34] K. Al-Sabahi, Z. Zuping, and M. Nadher, "A hierarchical structured self-attentive model for extractive document summarization (HSSAS)," *IEEE Access*, vol. 6, pp. 24205–24212, 2018.
[35] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.
[36] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, document ANSI/AAMI EC38, Association for the Advancement of Medical Instrumentation, 1998.
[37] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
[38] P. Du, W. A. Kibbe, and S. M. Lin, "Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching," *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006.
[39] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, Dec. 2009.
[40] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the risk of heart failure with EHR sequential data modeling," *IEEE Access*, vol. 6, pp. 9256–9261, 2018.
[41] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[43] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 437–478.
[44] W. Li and J. Li, "Local deep field for electrocardiogram beat classification," *IEEE Sensors J.*, vol. 18, no. 4, pp. 1656–1664, Feb. 2018.
[45] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.
[46] T. Laitio, J. Jalonen, T. Kuusela, and H. Scheinin, "The role of heart rate variability in risk stratification for adverse postoperative cardiac events," *Anesthesia Analgesia*, vol. 105, no. 6, pp. 1548–1560, 2007.
[47] P. Guzik, J. Piskorski, P. Barthel, A. Bauer, A. Müller, N. Junk, K. Ulm, M. Malik, and G. Schmidt, "Heart rate deceleration runs for postinfarction risk prediction," *J. Electrocardiol.*, vol. 45, no. 1, pp. 70–76, 2012.

RUNCHUAN LI received the B.S. degree in computer science from Zhengzhou University, China, in 2015, where he is currently pursuing the Ph.D. degree. He is also a Researcher with the Cooperative Innovation Center of Internet Healthcare, Zhengzhou University. His research interests include machine learning, data mining, artificial intelligence, and intelligence healthcare.

XINGJIN ZHANG received the B.E. and M.E. degrees in computer science from the School of Information Engineering, Zhengzhou University, Zhengzhou, China, in 1997 and 2005, respectively. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Mathematical Engineering and Advanced Computing. Since 1997, he has been a Faculty Member of the School of Information Engineering, Zhengzhou University, where he is currently a Lecturer in computer science. He is also a Researcher with the Cooperative Innovation Center of Internet Healthcare, Zhengzhou University. His research interests include machine learning and healthcare big data.

HONGHUA DAI received the M.Sc. degree from the Graduate School, Chinese Academy of Sciences, in 1986, and the Ph.D. degree from the Department of Computer Science, RMIT University, in 1994. After he received the Ph.D. degree, he was a Research Fellow of the Department of Computer Science, Monash University, from 1994 to 1997. He joined Deakin University, in 1999. He is also a specially appointed Professor with the Cooperative Innovation Center of Internet Healthcare, Zhengzhou University. His recent research interests include minimum message length principle-based causal discovery, inexact field learning approach, reliable knowledge discovery, big data intelligence, machine learning for weather forecasting, data mining, artificial intelligence, and intelligence healthcare. He is a member of the ACM and the IEEE Computer Society.

BING ZHOU was born in Shangqiu, China. He received the master's degree in computer science from Xi'an Jiaotong University, in June 1989, and the Ph.D. degree in computer science from the Beijing University of Aeronautics and Astronautics, in March 2003. From 2006 to 2007, he was a Visiting Scholar with the University of Washington. Since 2003, he has been served as a Professor and a Doctoral Tutor with Zhengzhou University. His main research interests include the intelligent diagnosis of ECG, computer vision, and internet medical and computer networks.

ZONGMIN WANG was born in Zhengzhou, China. He received the Ph.D. degree in engineering from Tsinghua University, in March 1996. From 1996 to 1997, he was with the University of Hong Kong for collaborative scientific research. From 1997 to 2017, he was a Professor and a Doctoral Tutor with Zhengzhou University. From August 2008 to August 2017, he was the Vice President of Zhengzhou University. From January 2010 to July 2010, he was a Senior Research Scholar with the University of Memphis. Since 2017, he has been the President of the Zhongyuan University of Technology. He is currently the Director of the Cooperative Innovation Center of Internet Healthcare, Zhengzhou University. His main research interests include the intelligent diagnosis of ECG, internet medical, computer networks, and intelligence healthcare.