ECG SIGNAL CLASSIFICATION BASED ON ADAPTIVE MULTI-CHANNEL WEIGHTED NEURAL NETWORK

F.J. Qiao[*][†]*, B. Li*[*]*, M.Q. Gao*[*]*, J.J. Li*[*]

Abstract: The intelligent diagnosis of cardiovascular diseases is a topic of great interest. Many electrocardiogram (ECG) recognition technologies have emerged, but most of them have low recognition accuracy and poor clinical application. To improve the accuracy of ECG classification, this paper proposes a multi-channel neural network framework. Concretely, a multi-channel feature extractor is constructed by using four types of filters, which are weighted according to their importance, as measured by kurtosis. A bidirectional long short-term memory (BLSTM) network structure based on attention mechanism is constructed, and the extracted features are taken as the input of the network, and the algorithm is optimized by attention mechanism. An experiment conducted on the MIT-BIH arrhythmia database shows that the proposed algorithm obtains excellent results, with 99.20% specificity, 99.87% sensitivity, and 99.89% accuracy. Therefore, the algorithm is practical and effective in the clinical diagnosis of cardiovascular diseases.

1. Introduction

Cardiovascular diseases have become common, especially among the elderly (over 60 years of age). These diseases bring great harm, and may even cause sudden death. Among them, cardiovascular diseases represented by arrhythmia are receiving increased attention from experts and scholars. The electrocardiogram (ECG) can record the electrical signal activity of the body surface, thereby reflecting heart activity, and is an important tool for clinical analysis and diagnosis of arrhythmia diseases [1, 2].

Doctors have scarce time to examine the increasing number of ECGs. Therefore, intelligent diagnosis technology of ECGs is necessary. The intelligent diagnosis of

---

[*]Fengjuan Qiao; Bin Li – Corresponding author; Mengqi Gao; Jiangjiao Li; School of Mathematics and Statistics, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China, E-mail: `ribbenlee@126.com`

[†]Fengjuan Qiao; School of Information and Communication Engineering, Communication University of China, Beijing 100024 China

ECGs has three stages: QRS detection, feature extraction, and classification. The normal ECGs are composed of P waves, QRS complex waves, T waves, and U waves. The QRS complex wave is the most intense amplitude variation band in an ECG, and it contains very obvious but important information. QRS detection locates QRS waves from ECG records over a long period of time. Robust QRS detection is helpful for subsequent classification. There have been many studies on QRS wave detection technology. The classic Pan-Tompkins algorithm [3] had a QRS detection rate reaching 99.3 % on the MIT-BIH arrhythmia database. Zhang et al. [4] used a Kalman filter to detect QRS waves, with 99.30 % detection sensitivity and 99.31 % positive prediction.

Feature extraction and classification are also important steps in the intelligent diagnosis of arrhythmia. There are two categories: the traditional method is often used for feature extraction, and machine learning classifiers are used to classify these features. Commonly used feature extraction methods include threshold-based methods [5], wavelet transform-based methods [6], digital filter-based methods [7, 8], higher order cumulant methods [9]. Commonly used classification methods include support vector machine (SVM) [10], extreme learning machine (ELM) [11] and random forest (RF) [12]. However, this type of method needs much manual intervention, and has poor classification results.

The second kind method is gradually evolving with the development of deep learning, whose methods in ECG intelligent diagnosis have been researched and have shown good results. Commonly used deep neural networks include convolutional neural networks (CNNs) [?] and recurrent neural networks (RNNs) [14]. Acharya et al. [15] used CNN to identify five types of arrhythmia with high accuracy of 94.03 %. Chauhan et al. [16] analyzed an ECG signal by LSTM and realized 96.45 % classification accuracy on the MIT-BIH arrhythmia database. Yildirm [17] used a wavelet transform (WT) to analyze an ECG signal before input to LSTM, which further improved classification accuracy to 99.32 %. Zhou et al. [18] introduced an attention mechanism (AM) to BLSTM and proved that it can improve the feature extraction ability of LSTM. Although good results have been achieved, deep learning methods have their disadvantages. They usually converge slowly, and easily fall into local optima. To train a deep neural network can take several days in real applications.

In our previous study [19], a hybrid method (ELM-LRF-BLSTM) was proposed, combining local receptive field based extreme learning machine (ELM-LRF) and bidirectional long short term memory network (BLSTM). This algorithm uses stacked convolution and pooling layers to extract features for use as input to BLSTM to learn the sequence representation and output the classification results. However, this algorithm uses a single random filter, and multiple filters have been proved [20] to improve the feature learning ability of the network. Therefore, this paper constructs a multi-channel feature extractor through different types of filters, which are weighted according to their importance, as measured by kurtosis. Combining the feature extractor with BLSTM and attention mechanism (AM), an algorithm called LSTM and AM-based multi-channel weighted ELM (MCW-CELM-LSTM-AM) is proposed. Applying this network to ECG classification remarkably improves the ability of feature extraction and classification.

We summarize the main contributions of this work as follows:

1) Four filters are used to extract features, and the weight of each filter is calculated based on kurtosis. The adaptive weighted combination of different filters is realized. Multiple filters can extract various features, so MCW-CELM-LSTM-AM can fully extract the implicit information of ECG signals.

2) MCW-CELM-LSTM-AM combines a BLSTM network with AM to mine the deep temporal information of the signal, and further enhances feature-extraction ability. Experiments conducted on the MIT-BIH Arrhythmia database demonstrate the algorithm's good performance.

The rest of this paper is organized as follows. Related work is reviewed in Section 2. The proposed MCW-CELM-LSTM-AM algorithm is described in Section 3. Experiments are discussed in Section 4. Section 5 provides conclusions and looks into the future.

2. Related work

We discuss related work on the convolutional extreme learning machine, bidirectional long short-term memory network, and attention mechanism.

2.1 Convolutional extreme learning machine

Convolutional extreme learning machine (CELM) is the general name of a class of algorithms, which originated from the algorithm of local receptive fields-based extreme learning machine (ELM-LRF) proposed by Huang et al. [21]. The ELM-LRF is composed of feature extraction and classification stages. The first stage uses orthogonal random convolution filters and pooling operations to achieve feature extraction, and the second stage uses least squares to calculate the output weights. This neural network achieved higher accuracy with a shorter training time than existing methods when applied to the NORB and MINIST datasets.

Experts and scholars have proved the good performance of ELM-LRF and have proposed a series of improved models, which are collectively known as CELM. Pang et al. [22] propose a deep extreme learning machine for the recognition of handwritten digits. This work is pioneer in evaluating the depth increase in CELM networks. The authors found that the ideal depth may vary with the dataset. Ding et al. [23] proposed CKELM, which mined feature information at different levels by continuous stacking of convolution and pooling layers, and evaluated the effect of network depth on performance. The above studies focused on a single random filter, but multiple filters enhance identification ability and improve generalization performance. McDonnell et al. [20] proposed a shallow convolutional fast neural network for image classification tasks based on random filters, using multiple filters, including patch, heuristic, and transferred learning filters, to improve recognition accuracy. Dos Santos et al. [24] realized a simple combination of random, patch, PCA, and Gabor filters, and proved that different types of filters can extract multifarious features.

Previous studies failed to statistically compare the importance of different types of convolution filters. Four types of filters are used in this paper to construct a multi-channel convolution layer. The filter weight in each channel is calculated based on kurtosis, and the adaptive weighted combination of filters is realized.

2.2 Bidirectional long short-term memory network

The long short-term memory (LSTM) network was proposed by Hochreiter and Schmid-Huber in 1997 [25]. It is an improved form of RNN, which solves its problem of long-term dependence.

It improves upon RNN by solving its problem of long-term dependence. LSTM introduces gating units to deal with the problems of memory, forgetting, and the input and output degrees of the memory unit, as shown in Fig. 1. The basic memory unit consists of one memory cell and three gating units. The memory cell stores the current network state. The three gating units are the input, output, and forgetting gates, which control the information flow in the memory block. In the forward propagation process, the input gate controls the information flow into the memory cell, and the output gate controls the information flow from the memory cell to the other structural units of the network. During backpropagation, the input gate controls the iteration error flowing out of the memory cell, and the output gate controls the iteration error flowing into the memory cell. The forgetting gate controls the internal circulation of memory cells, deciding whether to choose or forget information.
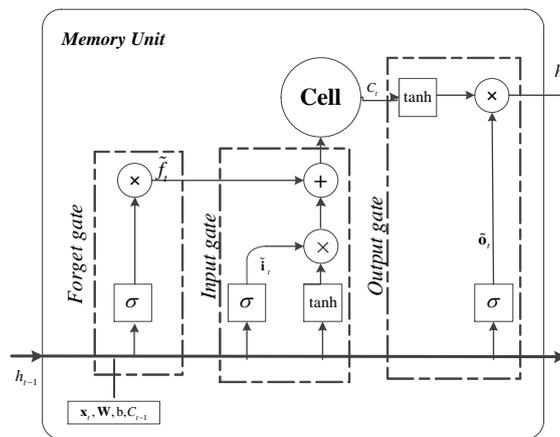

Fig. 1 *Structure of memory unit.*

Through this gating mechanism, the LSTM network can control the information flow in the unit, enabling it to preserve long-term information. That is, the LSTM network has the "memory" ability, which can eliminate external interference with the internal gradient in the training process, and avoid gradient disappearance and explosion.

Let the input vector of a single LSTM memory block be x_t. Then the update process of long-term memory unit c_t can be expressed as

$$f_t = \sigma \left(W_f \left[x_t, h_{t-1} \right] + b_f \right), \tag{1}$$

$$i_t = \sigma \left(W_i \left[x_t, h_{t-1} \right] + b_i \right), \tag{2}$$

$$g_t = \tanh \left(W_g \left[x_t, h_{t-1} \right] + b_g \right), \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t. \tag{4}$$

The update process of short-term memory unit is

$$o_t = \sigma \left(W_o \left[x_t, h_{t-1} \right] + b_o \right), \tag{5}$$

$$h_t = o_t \odot \tanh \left(c_t \right), \tag{6}$$

where f_t, i_t, o_t and c_t respectively, are the outputs of the forgetting, input, and output gates and cell state at time t. W_f, W_i and W_o are the weights of the forgetting, input, and output gates, respectively; b_f, b_i and b_o are the corresponding bases; c_t is the cell state at time t. σ is the sigmoid function; tanh is the hyperbolic tangent function and $\odot$ is the Hadamard product.

The connection between the hidden layers of the LSTM network is unidirectional, i.e., the cell state of the current moment depends on the input of the current moment and the hidden layer output of historical moments. However, in some specific tasks, the cell state at the present moment is also related to the state at a future time. Therefore, a bidirectional LSTM (BLSTM) network is proposed, whose structure is shown in Fig. 2. BLSTM takes into account both historical and future information, which can make up for the deficiency of LSTM.
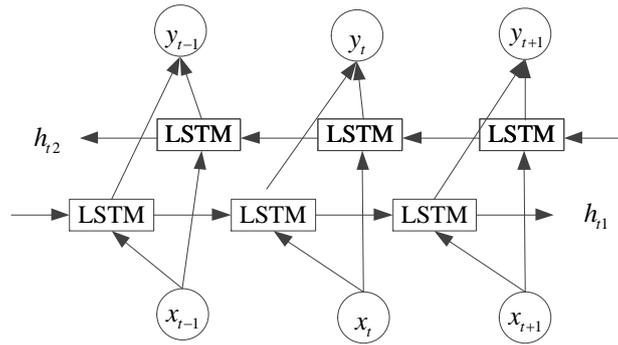

Fig. 2 *Structure of BLSTM.*

As can be seen from Fig. 2, for the each input x_t, two LSTM networks in opposite directions will be connected to it, and the output at the current moment is a combination of two LSTM networks, i.e., the forward and reverse LSTM networks jointly determine the output at the current moment. The formulas of BLSTM are

$$h_t = o_t \odot \tanh \left(f_t \odot c_{t-1} + i_t \odot g_t \right), \tag{7}$$

$$h_t{}' = o_t{}' \odot \tanh \left(f_t{}' \odot v'_{t+1} + i_t{}' \odot g'_t \right), \tag{8}$$

$$y_t = [h_t, h'_t], \tag{9}$$

where h_t and $h_t{}'$ arethe hidden layer outputs of the forward and reverse LSTM, respectively, as determined by Eq. (7)-(9). BLSTM can be regarded as forward and reverse LSTMs, which respectively learn the forward and backward features of the input [26]. Both LSTM networks are connected to an output layer that can only be updated if both forward and reverse features are introduced. Therefore, the output at the current time contains past and future information.

2.3 Attention mechanism

The attention mechanism (AM) was proposed by Bahdanau in 2014 [27]. It can selectively allocate attention according to specific tasks, strengthen the correlation between sampling points, and improve the ability of feature expression and recognition accuracy. In this paper, the AM is introduced into BLSTM. The weight of the feature vectors of the hidden layer is calculated and normalized according to the importance degree, and the attention resources are allocated according to the weight. The calculation process of AM is as follows.

1) Calculate the weight. Common methods to calculate the weights of feature vectors include the dot product, addition, and splicing. The weights of hidden layer output h_t at moment t can be expressed as

$$a_t = \text{score} \left(h_t \right). \tag{10}$$

2) Normalize the weights. The softmax function is used to normalize the weights to $[0, 1]$, so as to judge the importance of the features. If the weight of moment t moment is 1, then the features at this moment are important, and are given a large weight. If the weight of the moment t is 0, then the characteristic information at this moment is not important, and it receives a small weight. The normalized weight is calculated as

$$p_t = \exp \left(a_t \right) / \sum_{t=1}^{n} a_t. \tag{11}$$

3) Attention resources are allocated according to weights. We multiply the hidden layer output by the weights of attention to realize the allocation of attention resources. The weights of features differ, and the weights of more

important features are larger, so they are allocated more attention resources. The allocated attention resources y_t can be expressed as

$$y_t = \sum_{i=1}^{n} p_t h_t. \tag{12}$$

3. Methodology

CELM can be used for fast classification tasks, but its performance in processing complex signals is limited because only random filters are used. To solve this problem, this paper proposes MCW-CELM, which uses four types of filters to extract features, and the weight of each filter is calculated based on kurtosis. MCW-CELM realize adaptive weighted combination of different filters and is capable of extracting various features. BLSTM and AM are existing methods, and previous study has proved that the hybrid model (called BLSTM-AM) of the two methods has better performance [28]. To get better classification accuracy, MCW-CELM-LSTM-AM is proposed by combining the superiority of MCW-CELM and BLSTM-AM.

In this section, we introduce the proposed MCW-CELM-LSTM-AM algorithm. Four convolutional filters are discussed in Section 3.1. We explain the filter weighting process in Section 3.2, and the MCW-CELM-LSTM-AM structure in Section 3.3. Section 3.4 describes the loss and training algorithm.

3.1 Four filters

The filter can be regarded as a weight matrix to convolve the input data. Huang achieved good classification results on the NORB and MNIST datasets. However, random filters have been shown to degrade feature-extraction ability in other datasets, perhaps because of randomness of the weight matrix [29, 30]. We will improve upon this with four filters for feature extraction, as described below.

1) *Random filter*

   Each point in a random filter is randomly selected from a continuous probability density distribution and orthogonalized. Feature maps in a layer have different input weights.

2) *PCA filter*

   The PCA filter design comes from principal component analysis (PCA), as follows. If the input sample is X, the filter size is $k_1 \times k_2$, and the number of output mappings is K, the filter matrix is expressed as:

   $$W_k^{PCA} = \mathrm{mat}_{k_1,k_2}\left(q_k\left(XX^{\mathrm{T}}\right)\right) \in R^{k_1 \times k_2}, k = 1, 2, \ldots, K,$$

   where W_k^{PCA} is the weight matrix of the kth feature graph, and $\mathrm{mat}_{k_1,k_2}(v)$ can convert vector $v \in R^{k_1 k_2}$ to matrix $W \in R^{k_1 \times k_2}$. $q_k\left(XX^{\mathrm{T}}\right)$ represents the kth principal component vector.

3) *Patch filter*

Patch filters select sub-regions from the original input [20]. We randomly select signal fragments, whose elements are normalized and used as the patch filter matrix. It can be expressed as

$$W_k^{Patch} = norm(select(X, k_1 \times k_2)) \in R^{k_1 \times k_2}, k = 1, 2, \ldots, K,$$

where $select(X, k_1 \times k_2)$ indicates that a matrix of size $k_1 \times k_2$ is randomly selected from the sample X. $norm(\cdot)$ represents normalization, which is Z-score method in this paper.

4) *WT filter*

The WT filter is inspired by the wavelet transform (WT). It is useful when analyzing nonstationary time series. The input signal passes through high- and low-pass filters to obtain detail and approximate coefficients. The obtained approximate coefficients are filtered to obtain the second-layer component. This process is repeated until a specified level is reached, so as to obtain features that contain important information about the input signal. The detail coefficients represent frequency-related detail characteristics, and are selected as the values of the weight matrix of the WT filter.

3.2 Multi-channel weighting

The multi-channel weighting (MCW) refers to the process of weighting multiple filters based on kurtosis and feature fusion. Kurtosis is a statistic to reflect the property of a signal, which is particularly sensitive to impact signal and can reflect the strength of the impact feature. Kurtosis is a normalized fourth-order central moment, calculated as follows. For continuous variables,

$$K = \frac{\int_{-\infty}^{+\infty} [x(t) - \bar{x}]^4 p(x)dx}{\sigma^4} \tag{13}$$

and for discrete variables,

$$K = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{\sigma_t} \right)^4, \tag{14}$$

where $x(t)$ and x_i represent signal values at time t and the ith sampling, respectively. $\bar{x}$ is the mean value, $p(x)$ is the probability density function, σ is the standard deviation, N is the sample length.

The weights of each channel are calculated according to the kurtosis. Suppose there are m channels, and K_i represent the kurtosis value of the characteristic signal on ith $(i \leq m)$ channel. The weights α_i can be expressed as

$$\alpha_i = K_i / \left(\sum_{i=1}^{m} K_i \right). \tag{15}$$

3.3 Network architecture

Filters can extract different features of various importance. We use multiple filters to adaptively extract features, and combine LSTM and an attention mechanism in the proposed MCW-CELM-LSTM-AM algorithm. The algorithm uses multiple filters to construct a multi-channel feature extractor, and filters are weighted according to importance. We use random, PCA, patch, and WT filters to extract discriminative features, which are used as the input of BLSTM to learn and classify the sequence features. We introduce the AM between the hidden and output layers of BLSTM. The flowchart of the MCW-CELM-LSTM-AM network can be seen in Fig. 3.
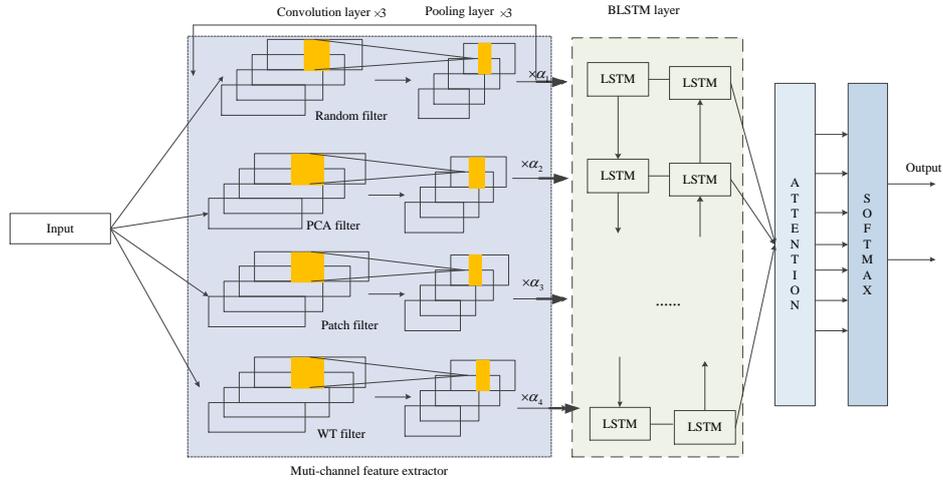
Fig. 3 *Flowchart of MCW-CELM-LSTM-AM network.*

3.4 Loss and training algorithm

The loss function is obtained by calculating the cross-entropy between the network output and the actual labels,

$$\text{Loss} = -\sum_i y_i' \log(y_i), \tag{16}$$

where y_i' represents the real labels of ith sample, and y_i is the output of the ith sample.

In the backpropagation stage, the Adam algorithm is used to optimize the parameters, and the learning rate of each parameter is dynamically adjusted using the first and second moment estimation of parameters, so that the weight is updated to obtain the optimal solution.

4. Experiment

4.1 Data description and preprocessing

We evaluated the performance of the proposed network through experiments on the commonly used MIT-BIH Arrhythmia (MIT-BIH-AR) Database [31], which consists of 48 dual-lead ECG records, each containing more than 30 minutes of ECG data. ECG records are taken from 47 subjects in patients at Beth Israel Hospital in Boston. The subjects include 25 men and 22 women aged between 32 and 89. Each record has been annotated by cardiovascular specialists in units of heartbeats. Most of the existing comparative studies have used only five types of ECGs. To facilitate comparison with existing studies, five types of ECGs from different records in the MIT-BIH-AR database are used in this study. The five ECG types are normal sinus rhythm (NSR), left bundle branch block (LBBB), right bundle branch block (RBBB), ventricular premature contraction (VPC), and atrial premature beat (APB). In Fig. 4, the waveforms of the beat types used in the study are given.
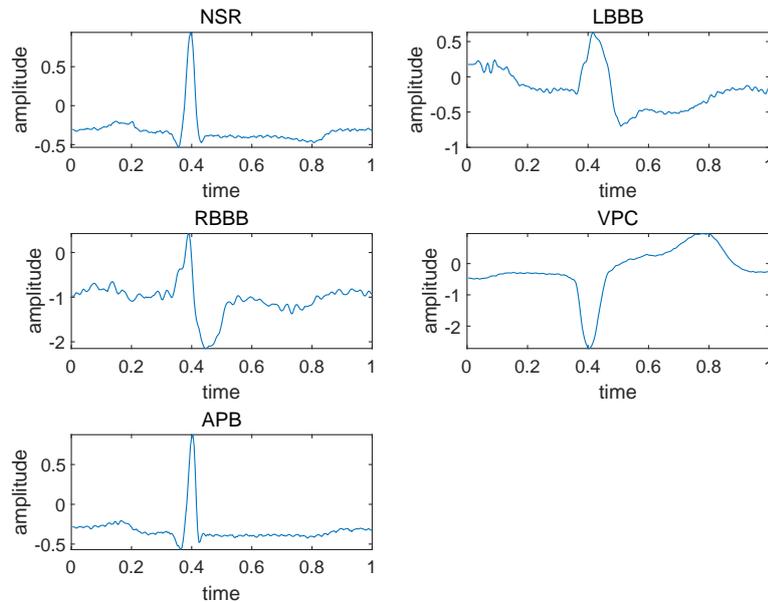

Fig. 4 *Waveforms of five different heartbeat types.*

Segmented signals from these beat types are used for a period that includes 250 samples. The records were preprocessed as follows before being fed into the network.

1) Use the Pan-Tompkins algorithm [3] to locate the R peak, and select ECG of 30 millisecond before and after R peak, with a total of 1 second, as a complete heartbeat cycle. Subsample the data to 250 Hz.

2) Normalize the heartbeat data x by

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}},\tag{17}$$

where $x_{\min}$ and $x_{\max}$ represent the maximum and minimum values of each sample, respectively.

3) Divide all data into training and testing sets at a 4:1 ratio. The number of heartbeats of each type in the training and testing sets are shown in Tab. I.

Beat Types	Number of training set	Number of testing set	Total
NSR	59970	14992	74962
LBBB	6454	1614	8068
RBBB	5803	1451	7254
VPC	5627	1407	7034
APB	2036	509	2545
Total	70890	19973	99863

Tab. I *Number of heartbeats in training and testing sets.*

4.2 Network structure design and parameter setting

The network structure was designed as shown in Tab. II.

The other parameters were set as follows. The initial learning rate was set to 0.01, the optimizer was selected as Adam, the batch size was 128, and number of iterations was 50.

We calculated the kurtosis values of the MIT-BIH-AR data according to Eq. (14), and then calculated the weights of each channel according to Eq. (15). The calculation results are shown in Tab. III

5. Experimental Results and analysis

5.1 Experimental Results

Fig. 5 shows the loss curve of MCW-CELM-LSTM-AM during training process.

As can be seen from Fig. 5, the loss values of MCW-CELM-LSTM-AM on the MIT-BIH-AR database decrease with the increase of iterations. After about 30 iterations, the algorithm tends to be stable and the loss value is close to zero. Tab. IV shows the confusion matrix of MCW-CELM-LSTM-AM on the testing set, where the vertical direction represents the true label of the sample and the horizontal direction represents the predicted label.

The classification results can be seen from Tab. IV. There were 14992 NSR samples, all but one correctly classified. The recognition accuracy of the NSR category reached 99.99 %. The recognition accuracies of LBBB, RBBB, VPC and APB are 99.81 %, 99.72 %, 98.64 % and 98.23 % respectively. Among all categories, 19972

Number	Layer Name	Parameters	Output Size
0	Input	–	250×1
1	Convolution layer 1	Random filter 4@17 × 1 PCA filter 4@17 × 1 Patch filter 4@17 × 1 WT filter 4@17 × 1	$4 \times 234 \times 4$
2	Pooling layer 1	2	$4 \times 117 \times 4$
3	Convolution layer 2	Random filter 8@6 × 1 PCA filter 8@6 × 1 Patch filter 8@6 × 1 WT filter 8@6 × 1	$4 \times 112 \times 8$
4	Pooling layer 2	2	$4 \times 56 \times 8$
5	Convolution layer 3	Random filter 3@5 × 1 PCA filter 3@5 × 1 Patch filter 3@5 × 1 WT filter 3@5 × 1	$4 \times 52 \times 3$
6	pooling layer 3	2	$4 \times 26 \times 3$
7	Weighting	–	$4 \times 78 \times 1$
8	BLSTM	64 Unit	1×128
9	BLSTM	64 Unit	1×128
10	Attention	–	128×1
11	Dense	–	5

Tab. II *Network structure design of MCW-CELM-LSTM-AM.*

	Kurtosis values	channel 1	channel 2	channel 3
Kurtosis	$K_1 \approx 3.77$	$K_2 \approx 3.56$	$K_3 \approx 4.11$	$K_4 \approx 7.00$
Weights	$\alpha_1 \approx 0.20$	$\alpha_2 \approx 0.19$	$\alpha_3 \approx 0.22$	$\alpha_4 \approx 0.38$

Tab. III *Kurtosis values and weights of each channel.*

	NSR	LBBB	RBBB	VPC	APB
NSR	14991	0	1	0	0
LBBB	2	1611	0	0	1
RBBB	1	2	1447	0	1
VPC	2	1	0	1402	2
APB	6	0	2	1	500

Tab. IV *Confusion matrix of MCW-CELM-LSTM-AM.*

samples were collected, and 19551 correctly classified. The average recognition accuracy of MCW-CELM-LSTM-AM on the MIT-BIH-AR database was 99.89 %.

We also conducted experiments with single filters, i.e., the PCA, patch, and WT filters in the convolution stage. The loss curves based on single filters are shown in Fig. 6.
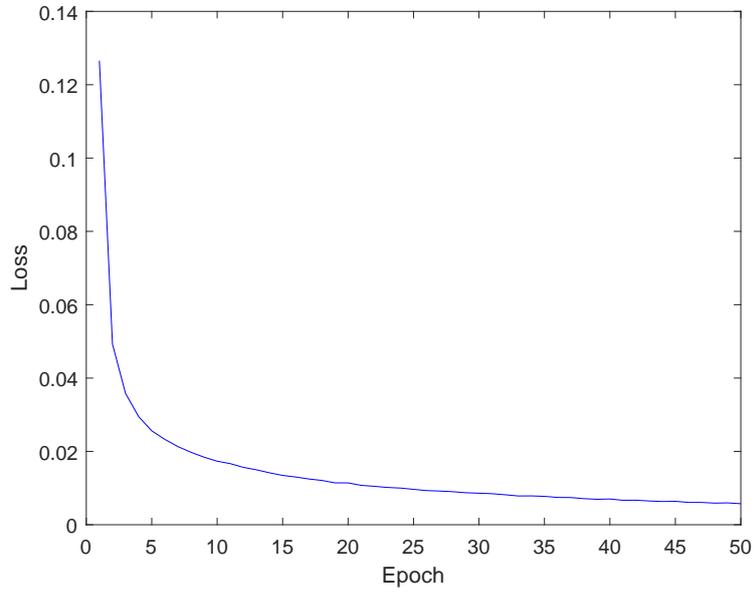
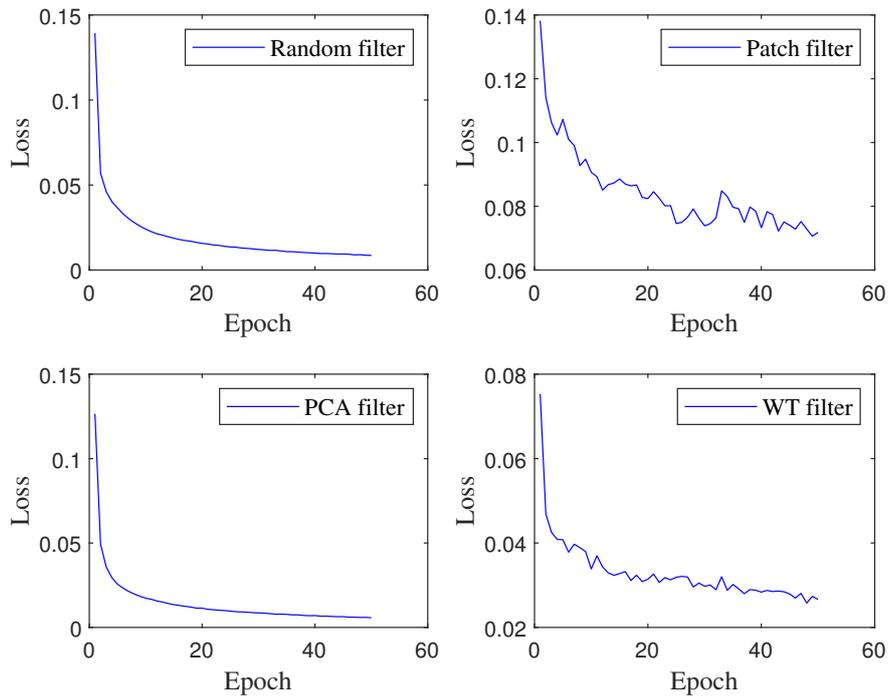**Fig. 5** *Loss curve of AFLRF-BLSTM-AM algorithm in MIT-BIH-AR database.*

Fig. 6 *Loss curve of MCW-CELM-LSTM-AM based on single filter.*

It can be seen from Fig. 6, the loss function curves based on different filters all decrease with increasing iterations. However, compared to MCW-CELM-LSTM-AM, the single filter-based algorithm has a slower convergence speed and larger final value. Experimental results of MCW-CELM-LSTM-AM and single filter - based algorithms are shown in Fig. 7.
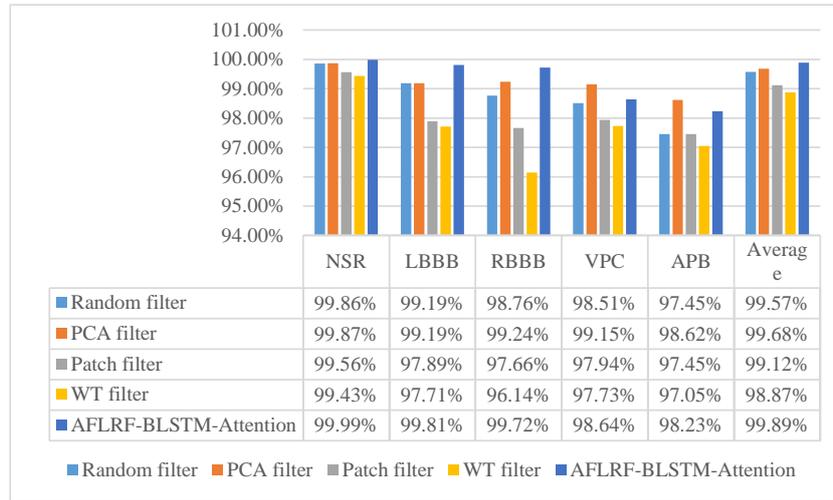

	NSR	LBBB	RBBB	VPC	APB	Average
■ Random filter	99.86%	99.19%	98.76%	98.51%	97.45%	99.57%
■ PCA filter	99.87%	99.19%	99.24%	99.15%	98.62%	99.68%
■ Patch filter	99.56%	97.89%	97.66%	97.94%	97.45%	99.12%
■ WT filter	99.43%	97.71%	96.14%	97.73%	97.05%	98.87%
■ AFLRF-BLSTM-Attention	99.99%	99.81%	99.72%	98.64%	98.23%	99.89%

■ Random filter ■ PCA filter ■ Patch filter ■ WT filter ■ AFLRF-BLSTM-Attention

Fig. 7 *Classification accuracy based on different filters on MIT-BIH-AR database.*

To evaluate the generalization performance of the proposed algorithm, we conducted comparative experiments of CELM (using only random filters), BLSTM, CELM-BLSTM (only random filters), CELM-BLSTM-AM (only random filters), and MCW-CELM-LSTM-AM. The specificity, sensitivity, and accuracy of these algorithms are shown in Tab. V.

Algorithm	Specificity [%]	Sensitivity [%]	Accuracy [%]
CELM(random filter)	89.64	92.71	93.02
BLSTM	92.25	96.45	97.26
CELM-BLSTM	92.72	99.30	99.32
CELM-BLSTM-AM	99.20	99.51	99.57
MCW-CELM-LSTM-AM	99.20	99.87	99.89

Tab. V *Performance of different algorithms.*

From Tab. V, the specificity, sensitivity, and accuracy of CELM-BLSTM are 92.72 %, 99.30 %, and 99.32 %, respectively, better than those of the single CELM or BLSTM algorithm. This is because CELM-BLSTM combines the advantages of CELM and BLSTM, thereby improving feature-extraction and classification capabilities. CELM-BLSTM-AM shows further improvements, with 99.20 % specificity,

99.51 % sensitivity, and 99.57 % accuracy. This illustrates the necessity of AM. Compared to CELM-BLSTM, the specificity of CELM-BLSTM-AM is greatly improved, which shows its improved ability in diagnosing disease-free patient. The specificity, sensitivity, and accuracy of CELM-BLSTM-AM reach 99.20 %, 99.87 %, and 99.89 %, respectively. This is because the use of four types of filters enhances feature-extraction ability, and discriminative features are fully extracted.

5.2   Discussion

Feature extraction is an important step in classification, whose quality can greatly affect its results. We proposed the MCW-CELM-LSTM-AM algorithm, using four types of filters (random, PCA, patch, WT) to construct a multi-channel feature extractor, so that various local spatial features were fully extracted. The introduction of kurtosis realized the adaptive combination of filters of different types according to their importance degree, to further improve the feature-extraction capability.

BLSTM can consider long-time dependence and analyze the influence of historical and future information on the current state. The combination of CELM and BLSTM is necessary to learn local spatial and temporal characteristics. The introduction of an attention mechanism strengthens the correlation between sampling points, allocates limited attention resources to important features, and further improves feature-extraction ability.

Experiments on the MIT-BIH arrhythmia database demonstrated the algorithm's high recognition accuracy. The algorithm was compared to CELM, BLSTM, CLEM-BLSTM, and CELM-BLSTM-AM, and was shown to have higher recognition accuracy. The combination of CELM, BLSTM, and AM can improve feature-extraction and classification ability.

To better evaluate the performance of the proposed algorithm, we compared it to several current algorithms, with results as shown in Tab. VI.

Literature	Methods	Accuracy [%]
Chauhan et al. (2015) [16]	LSTM	96.45
Kiranyaz et al. (2016) [32]	1D-CNN	99.00
Acharya et al. (2017) [15]	CNN	94.03
Sahoo et al. (2017) [33]	DWT + SVM	98.39
Yang et al. (2018) [34]	PCANet + SVM	97.08
Shu et al. (2018) [35]	CNN + LSTM	98.10
Yildirim et al. (2020) [17]	DWT + BLSTM	99.39
This paper	MCW-CELM-LSTM-AM	99.89

Tab. VI *Comparison results with the State-of-the-art.*

As can be seen from Tab. VI, MCW-CELM-LSTM-AM has higher recognition accuracy than the other methods, which demonstrates its competitiveness and practicability. The superiority of the proposed algorithm is summarized as follows:

1) The algorithm can extract a variety of discriminative features through adaptive combination of filters.

2) The proposed algorithm is fully automatic and does not require denoising or manual feature extraction.

The algorithm has some potential limitations:

1) The algorithm is based on the premise that there is only one label per heartbeat.

2) This study only uses five types of ECGs for training and testing to compare with the existing literature. In addition, the number of other types of ECG samples in the MIT-BIH-AR database is too small to carry out corresponding training and testing. It has not been proved whether the model can be used in other types of ECGs. We will explore the application of this algorithm to multi-classification problems in future work.

6. Conclusion

We proposed a multi-channel neural network algorithm, MCW-CELM-LSTM-AM. A multi-channel feature extractor was constructed using four filters (random, patch, PCA, WT) to extract spatiotemporal features of the signals. To measure the importance of the filter, it was weighted according to kurtosis, so as to realize an adaptive weighted combination of filters. The extracted features were used as the input of a BLSTM layer to learn the temporal features of the sequences. An attention mechanism further optimized the algorithm. Output categories were obtained by a softmax function. The proposed algorithm was used for ECG classification tasks, with excellent results of 99.20 % specificity, 99.87 % sensitivity, and 99.89 % accuracy. The recognition accuracy of this algorithm is higher than that of some current algorithms.

The next step of our research will include the suppression and removal of noise in ECG to further improve classification accuracy. We will further explore the possibilities of deep-learning methods in the field of medical diagnosis to aid in the development of intelligent healthcare.

Acknowledgement

References

[1] ESMAILI A., KACHUEE M., SHABANY M. Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time. *IEEE Transactions on Instrumentation and Measurement.* 2017, 66(12), pp. 3299–3308, doi: 10.1109/TIM.2017.2745081.

[2] DASTJERDI A.E., KACHUEE M., SHABANY M. Non-invasive blood pressure estimation using phonocardiogram. *IEEE International Symposium on Circuits & Systems.* 2017, doi: `10.1109/ISCAS.2017.8050240`.

[3] PAN J., TOMPKINS W.J. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering.* 1985, 32(3), pp. 230–236, doi: `10.1109/TBME.1985.325532`.

[4] ZHANG Z., YU Q., ZHANG Q. A Kalman filtering based adaptive threshold algorithm for QRS complex detection. *Biomedical Signal Processing and Control.* 2020, pp. 101827, doi: `10.1016/j.bspc.2019.101827`.

[5] CHOUHAN V.S., MEHTA S.S. Threshold-based Detection of P and T-wave in ECG Using New Feature Signal. *IEEE Transactions on Biomedical Engineering.* 2008, 8(2), pp. 144–153, doi: `10.1080/00045603909357325`.

[6] SARITHA C., SUKANYAE V., MURTHY Y.N. ECG Signal Analysis Using Wavelet Transforms. *Bulg. J. Phys. Chapter.* 2008, 35(1), pp. 68–77, doi: `10.1007/s12064-011-0124-1`.

[7] MAHMOUD S.A., BAMAKHRAMAH A., Al-TUNAIJI S.A. Six Order Cascaded Power Line Notch Filter for ECG Detection Systems with Noise Shaping. *Circuits, Systems, and Signal Processing.* 2014, 33(8), pp. 2385–2400, doi: `10.1007/s00034-014-9761-1`.

[8] PASOLLI E., MELGANI F. Active Learning Methods for Electrocardiographic Signal Classification. *IEEE Transactions on Information Technology in Biomedicine.* 2010, 14(6), pp. 1405–1416, doi: `10.1109/TITB.2010.2048922`.

[9] AFKHAMI R.G., AZARNIA G., TINATI M. A. Cardiac Arrhythmia Classification Using Statistical and Mixture Modeling Features of ECG Signals. *Pattern Recognition Letters.* 2016, 70(Jan.15), pp. 45–51, doi: `10.1016/j.patrec.2015.11.018`.

[10] REZGUI D., LACHIRI Z. ECG biometric recognition using SVM-based approach. *IEEJ Transactions on Electrical and Electronic Engineering.* 2016, 11(S1), pp. 94–100, doi: `10.1002/tee.22241`.

[11] SADR N., CHAZAL P.D., SCHAIK A.V. Sleep apnoea episodes recognition by a committee of ELM classifiers from ECG signal. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2015, pp. 7675–7678, doi: `10.1109/EMBC.2015.7320170`.

[12] GANESHKUMAR R., YSKUMARASWAMY P.D., SCHAIK A.V. Investigating Cardiac Arrhythmia in ECG using Random Forest Classification. *International Journal of Computer Applications.* 2012, 37(4), pp. 31–34, doi: `10.5120/4599-6557`.

[13] BALOGLU U.B., TALO M., YILDIRIM O. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognition Letters.* 2019, 122, pp. 23–30, doi: `10.1016/j.patrec.2019.02.016`.

[14] SALLOUM R., KUO C.J. ECG-based biometrics using recurrent neural networks. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 2017, 122, pp. 23–30, doi: `10.1109/ICASSP.2017.7952519`.

[15] ACHARYA A.B., OH S.L., HAGIWARA Y. A deep convolutional neural network model to classify heartbeats. *Computers in Biology Medicine.* 2017, 89, pp. 389–96, doi: `10.1016/j.compbiomed.2017.08.022`.

[16] CHAUHAN S., VIG L. Anomaly detection in ECG time signals via deep long short-term memory networks. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA).* 2015, pp. 1–7, doi: `10.1109/DSAA.2015.7344872`.

[17] YILDIRIM O. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine.* 2020, 96, pp. 189–202, doi: `10.1016/j.compbiomed.2018.03.016`.

[18] ZHOU P., SHI W., TIAN J. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.* 2016, 12, doi: `10.18653/v1/P16-2034`.

[19] QIAO F.J., LI B., ZHANG Y.M. A Fast and Accurate Recognition of ECG Signals Based on ELM-LRF and BLSTM Algorithm. *IEEE Access.* 2020, 8, pp. 71189-71198, doi: `10.1109/ACCESS.2020.2987930`.

[20] MCDONNELL M.D., VLADUSICH T. Enhanced image classification with a fast-learning shallow convolutional neural network. *2015 International Joint Conference on Neural Networks (IJCNN)*. 2015, doi: `10.1109/IJCNN.2015.7280796`.

[21] HUANG G.B., BAI Z., KASUN L.L. Local Receptive Fields Based Extreme Learning Machine. *IEEE Computational Intelligence Magazine* . 2015, 10(2), pp. 18–29, doi: `10.1109/MCI.2015.2405316`.

[22] SHAN P., YANG X.Y. Deep Convolutional Extreme Learning Machine and Its Application in Handwritten Digit Classification. *Multidimensional Systems Signal Processing*. 2016, 2016, pp. 1–10, doi: `10.1155/2016/3049632`.

[23] DING S., GUO Y. Extreme learning machine with kernel model based on deep learning. *Neurocomputing*. 2017, 28(8), pp. 1975–84, doi: `10.1007/s00521-015-2170-y`.

[24] DOS SANTOS M.M., DA S.F. Deep Convolutional Extreme Learning Machines: filters combination and error model validation. *Neurocomputing*. 2019, 329, pp. 359–69, doi: `10.1016/j.neucom.2018.10.063`.

[25] HOCHREITER S., SCHMIDHUBER J. Long Short-Term Memory. *Neural Computation*. 1997, 9(8), pp. 1735–1780, doi: `10.1162/neco.1997.9.8.1735`.

[26] HSU W.N., ZHANG Y., GLASS J. A prioritized grid long short-term memory RNN for speech recognition. *2016 IEEE Spoken Language Technology Workshop (SLT)*. 2016, pp. 467-473, doi: `10.1109/SLT.2016.7846305`.

[27] CHEN K., WANG R., UTIYAMA M. Neural Machine Translation With Sentence-Level Topic Context. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019, 27(12), pp. 1970–1984, doi: `10.1109/TASLP.2019.2937190`.

[28] WANG Y.Q., HUANG M.L., ZHU X.Y. Attention-based LSTM for aspect-level sentiment classification. *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016, pp. 606–615, doi: `10.18653/v1/D16-1058`.

[29] CHAN T.H., JIA K., GAO S. PCANet: A Simple Deep Learning Baseline for Image Classification?. *IEEE Transactions on Image Processing*. 2015, 24(12), pp. 5017-32, doi: `10.1109/TIP.2015.2475625`.

[30] JARRETT K., KAVUKCUOGLU K. RANZATO M.A., What is the best multi-stage architecture for object recognition?. *2009 IEEE 12th International Conference on Computer Vision*. 2009, doi: `10.1109/ICCV.2009.5459469`.

[31] MOODY G.B., MARK R.G. The Impact of The MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*. 2001, 20(3), pp. 45-50, doi: `10.1109/51.932724`.

[32] KIRANYAZ S., INCE T., GABBOUJ M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*. 2016, 63(3), pp. 664-675, doi: `10.1109/TBME.2015.2468589`.

[33] SAHOO S., KANUNGO B., BEHERA S. Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities. *Measurement*. 2017, 108, pp. 55–66, doi: `10.1016/j.measurement.2017.05.022`.

[34] YANG W.Y., SI Y.J., DI W. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Computers in Biology and Medicine*. 2018, 101, pp. 22-32, doi: `10.1016/j.compbiomed.2018.08.003`.

[35] SHU L.O., EDDIE Y.K.NG., RY S.T. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Computers in Biology and Medicine*. 2018, pp. 278-287, doi: `10.1016/j.compbiomed.2018.06.002`.