

Prediction of H-Type Hypertension Based on Pulse Wave MFCC Features Using Mixed Attention Mechanism

Jingdong Yang (✉ eerfriend@yeah.net)

University of Shanghai for Science and Technology <https://orcid.org/0000-0003-1296-8230>

Lei Chen

University of Shanghai for Science and Technology School of Optical-Electrical and Computer Engineering

Shuchen Cai

University of Shanghai for Science and Technology School of Optical-Electrical and Computer Engineering

Tianxiao Xie

Shanghai University of Traditional Chinese Medicine

Haixia Yan

Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: H-type Hypertension, MFCC, BiLSTM, Mixed attention mechanism

Posted Date: December 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1050432/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Prediction of H-type Hypertension based on Pulse Wave MFCC Features using Mixed Attention Mechanism

Jingdong Yang^{1*}, Lei Chen¹, Shuchen Cai¹, Tianxiao Xie², Haixia Yan²

1. Autonomous Robot Lab, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

2. Department of Traditional Chinese Medicine Diagnosis, Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Corresponding author: Jingdong Yan, E-mail: eerfriend@yeah.net

Abstract: H-type hypertension increases the risks of stroke and cardiovascular disease, posing a great threat to human health. Pulse diagnosis in traditional Chinese medicine (TCM) combined with deep learning can independently predict suspected H-type hypertension patients by analyzing their pulse physiological activities. However, the traditional time-domain feature extraction has a higher noise and baseline drift, affecting the classification accuracy. In this literature, we propose an effective prediction on frequency-domain pulse wave features. First, we filter time-domain pulse waves via removal of high-frequency noises and baseline shift. Second, Hilbert-Huang Transform is explored to transform time-domain pulse wave into frequency-domain waveform characterized by Mel-frequency cepstral coefficients (MFCC). Finally, an improved BiLSTM model, combined with mixed attention mechanism is built to applied for prediction of H-type hypertension. With 337 clinical cases from Longhua Hospital affiliated to Shanghai University of TCM and Hospital of Integrated Traditional Chinese and Western Medicine, the 3-fold cross-validation results show that sensitivity, specificity, accuracy, F1-score and AUC reaches 93.48%, 95.27%, 97.48%, 90.77% and 0.9676, respectively. The proposed model achieves better generalization performance than the classical traditional models. In addition, we calculate the feature importance both in time-domain and frequency-domain according to purity of nodes in Random Forest and study the correlations between features and classification that has a good reference value for TCM clinical auxiliary diagnosis.

Keywords: H-type Hypertension, MFCC, BiLSTM, Mixed attention mechanism

1. Introduction

Hypertension with plasma levels of homocysteine (HCY) greater than $10\mu\text{mol/L}$ is defined as "H-type" hypertension [1]. According to a study, H-type hypertension accounts for about 75% of Chinese adult hypertension patients (91% in males and 60% in females) [2]. Chinese guidelines for the prevention and treatment of hypertension indicate that the intensity of stroke caused by elevated blood pressure in Chinese population is 1.5 times that of western population. A large number of studies have shown that elevated plasma HCY level is an independent risk factor for cardiovascular and cerebrovascular diseases, and the plasma HCY level is positively correlated with the risk of cardiovascular and cerebrovascular events. However, like hypertension, it has no clear demarcation value [3]. Up to now, detection of homocysteine content in blood is the only way to clinically diagnose H-type hypertension [4]. Therefore, it is of great significance to use a noninvasive and economical method for rapid detection of H-type hypertension. The pulse condition of traditional Chinese medicine (TCM) contains abundant physiological and pathological information of the human body, and is a window for transmitting and observing the changes of internal functions [5-6]. A disease can often be identified by pulse waves, so pathological pulse has become an important basis for the diagnosis of diseases. With the development of machine learning, the fragmentation of TCM knowledge is no longer a problem in pathological diagnosis, and many artificial intelligence approaches have been applied in the field of pulse diagnosis of TCM.

2. Related Work

In recent years, many pulse wave acquisition instruments and many methods of pulse wave characteristic analysis have been applied to modern Chinese medicine. The earliest pulse wave analysis methods were to extract feature points via signal processing, such as zero-crossing based on wavelet transform [7], amplitude threshold and sliding window positioning main peak [8], and time-domain differential period ratio [9] etc. However, the global information of pulse wave could not be captured due to heavy workload and low recognition rate. Luo [10] applied AdaBoost on hypertension prediction based on time-domain pulse wave, and reached classification accuracy of 86.41%. Feng [11] used the fuzzy C-means clustering to classify the characteristics of frequency-domain pulse wave. Zhang [12] used random forest (RF) to reduce the feature dimension of pulse wave and applied SVM classification to improve the classification accuracy by 10%. However, the above methods all ignore the sequential characteristic of pulse wave.

With the development of deep learning, convolutional neural network (CNN) has been widely used in image processing. Zhang^[13] proposed a CNN extended dimension preprocessing, which adopted sample statistical features and Hilbert-Huang transform to extend the dimension and improve the speed of training. Liu^[14] extracted single-period and multi-period features of pulse wave based on CNN and combined with frequency-domain features for classification with the accuracy of 93%. Yan^[15] transformed pulse wave into threshold free recursive graph for classification via VGG-16 network, and the accuracy reached 98.14%, which provided a new idea for feature extraction of pulse wave. However, the above approaches only classify the time-domain waveform, but do not consider frequency-domain characteristics of pulse wave.

In this literature, we propose an effective model of frequency-domain pulse wave classification using mixed attention mechanism on H-type hypertension. The filtered time-domain pulse wave is transformed into frequency-domain Mel-scale Cepstral Coefficients, and the mixed attention mechanism is applied to extract local and global relevant features of pulse wave. Experiments show that the proposed model excels in classification accuracy and generalization performance.

3. MFCC Feature Extraction of pulse waveform

3.1 Filtering

In clinical pulse wave sampling, external interference, the collector's breathing, and slight body movements etc. will lead to the difference between the collected instances and the actual instances, which results in high-frequency noise and baseline drift^[16-18]. Wavelet transform is usually used to reduce high-frequency noise. The orthogonality, direction selectivity and variable resolution in time and frequency domain of wavelet transform can effectively identify the catastrophe point of signal on the time axis and filter the high-frequency noise of pulse wave. Common methods to remove baseline drift of pulse wave include Wavelet transform (WT), Empirical mode decomposition (EMD) and Smooth Prior Approach (SPA)^[19]. WT and EMD generally adopt multiple parameter adjustment. When the interference frequency is wide, the filtering parameters are difficult to be set. SPA only adjusts the frequency response by changing the smoothing parameters, which can effectively improve the filtering speed. The formula is as follows.

$$p = (I - (I + \lambda^2 D_2^T D_2)^{-1})y \quad (1)$$

Where, p is the effective pulse wave signal, I is the unit matrix of observation matrix, D_2 is the second-order differential matrix, and y is the original pulse wave signal. Different filtering properties depend on different regularization parameters λ . The baseline drift frequency is 0.2-0.3Hz, and the sampling frequency is 200Hz. Therefore, under the sampling frequency, the cut-off frequency of $\lambda=2500$ is $200 \times 0.0025\text{Hz} = 0.5\text{Hz}$, which can effectively remove baseline drift lower than 0.5Hz in original pulse waveform.

The signal to noise ratio (SNR) and root mean square error (RMSE) are considered as evaluation indicators of pulse wave denoising. The larger the SNR value, the smaller the RMSE value, and the better the pulse wave denoising effect.

$$SNR = 10 \log_{10} \left(\sum_{n=1}^N \frac{p^2(n)}{[p'(n) - p(n)]^2} \right) \quad (2)$$

$$RMSE = \sqrt{\left(\sum_{n=1}^N \frac{[p'(n) - p(n)]^2}{N} \right)} \quad (3)$$

Where $p'(n)$ represents the original signal, $p(n)$ denotes the signal after removal of baseline drift, and N represents the length of original signal.

3.2 MFCC Feature

Mel-scale Frequency Cepstral Coefficients (MFCC) are characterized by cepstral parameters extracted from Mel cepstral frequency domain^[20-22]. The pulse wave of each frame is composed of a Mel cepstral coefficient vector,

and the MFCC features of each frame are continuous in pulse wave timing and correlated. The physical meaning of MFCC features is a set of feature vectors obtained by encoding the physical information (spectrum envelope and details) of the signal, which represents the distribution of signal energy in different frequency ranges. Traditional MFCC feature extraction is to transform time-domain features into frequency-domain by Fourier transform. Fourier transform is a global transform, and the conversion effect is good in frequency-domain of stationary signals. However, the pulse wave is a non-stationary signal, and the local characteristics of the signal in frequency-domain cannot be described by frequency-domain of Fourier transform fully, which affects classification performance of pulse wave in patients with H-type hypertension. Therefore, an improved MFCC feature extraction is proposed in this literature. The detailed process is shown in Figure 1.

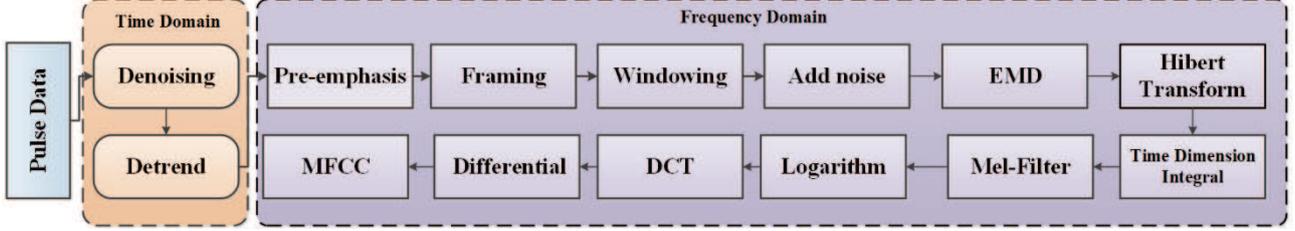


Fig.1 The processing of improved MFCC feature extraction

In the pre-emphasis process, a Gaussian filter is applied to framing (sampling) operation, and 256 sampling points are taken as an observation unit to ensure that an observation unit contains at least one pulse wave period. In the process of windowing, we select hamming window to add the continuity of the left and right ends of the "frame" to reduce the reconstruction error. The pulse wave signal shows leaping change on the time scale. The traditional EMD decomposition cannot effectively separate the different Intrinsic Mode Function (IMF) components according to the characteristic scale, nor clearly reflect the intrinsic characteristics of pulse wave. Direct screening of pulse wave will produce the phenomenon of mode mixing. In this literature, we add the adaptive white noise to pulse wave decomposition stage to make pulse wave components in different time scales superimposed and eliminated, and also eliminates the reconstruction error caused by adding white noise, which ensures the decomposition accuracy of pulse wave and reduces mode mixing effect. The pulse wave signal set with white noise can be expressed as the following formula.

$$p^i(t) = p(t) + g^i(t) \quad (4)$$

Where, $g^i(t) (i = 1, 2, \dots, I)$ represents Gaussian noise, and i is the total number of pulse wave instances.

The pulse waveform can be decomposed into various IMF components and corresponding residual of each order via EMD with adaptive noise and expressed as follows.

$$\begin{cases} r_k(t) = p(t) - IM\tilde{F}_1(t), k = 1 \\ r_k(t) = r_{k-1}(t) - IM\tilde{F}_k(t), k > 1 \end{cases} \quad (5)$$

$$\begin{cases} IM\tilde{F}_{k+1}(t) = \frac{1}{I} \sum_{i=1}^I IMF_{k+1}^i, k = 0 \\ IM\tilde{F}_{k+1}(t) = \frac{1}{I} \sum_{i=1}^I E_1\{r_k(t) + \varepsilon_k E_k[g^i(t)]\}, k > 0 \end{cases} \quad (6)$$

The marginal spectrum, the pulse wave frequency-domain feature, is obtained by Hilbert-Huang transform (HHT) and time integration according to the IMF components. The traditional MFCC extraction by Fourier transform cannot reflect the change of pulse wave in a certain period of time and the information of a certain frequency moment. The Fourier transform is only suitable for stationary signals, rather than non-stationary signals such as pulse waves. In this literature, the improved HHT is applied to obtain the non-stationary characteristics of pulse wave. Then, Mel triangle filter banks are applied to smooth frequency-domain features to eliminate harmonic effect and highlight the resonance peak. The frequency response of Mel triangle filter can be expressed as the following formula.

$$T_m(k) \begin{cases} 0, & k < f(m-1) \text{ or } k > f(m+1) \\ \frac{2(k-f(m-1))}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m-1) \leq k \leq f(m) \\ \frac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m)-f(m-1))}, & f(m) \leq k \leq f(m+1) \end{cases} \quad (7)$$

Where, $\sum_{m=0}^{M-1} T_m(k) = 1, M = 24$.

Then the logarithmic energy is calculated according to the output of each filter set. After discrete cosine transform, the MFCC features of pulse waveform can be obtained.

$$C(n) = \sum_{m=0}^{N-1} \ln \left(\sum_{k=0}^{N-1} |H_\alpha(k)|^2 T_m(k) \right) \cos\left(\frac{\pi n(m-0.5)}{M}\right), 0 \leq m \leq M, n = 1, 2, \dots, L \quad (8)$$

Where, $H_\alpha(k)$ is the marginal spectrum signal, and L is the order of MFCC, set as 12.

4. Improved BiLSTM Model with Mixed Attention Mechanism

In this literature, the time-domain of pulse waveform is transformed into frequency-domain for extraction of MFCC feature. Long short-term memory network (LSTM) [23-24] can learn and remember with long-dependent temporal pulse wave MFCC features via its unique gate mechanism. BiLSTM [25] is an improvement on the traditional LSTM, including two layers of LSTM for transmission of the forward and reverse input data respectively. BiLSTM stacks the results of the two layers and extracts the feature correlations from two directions, which can effectively enhance the feature extraction effect of LSTM. When the pulse wave data passes through the BiLSTM layer, the hidden state unit will increase the channel dimension of pulse wave from the initial one dimension to N dimension, and the data correlations generated after the channel expansion is often ignored. Therefore, this literature adds the channel attention mechanism on the basis of BiLSTM model to learn the feature correlations of pulse wave in the channel dimension.

However, when the input sequence is too long and the redundancy of input data from two directions increases, the vector of features cannot accurately express correlations between the data. At the same time, due to the length limit, the model cannot retain all the important information. Therefore, in this literature we add spatial Attention mechanism on the basis of channel Attention, selectively learns the input sequence, retains the intermediate results of the BiLSTM encoder, correlates the pulse wave sequences in the output, and forms a mixed Attention mechanism to learn the feature correlations of pulse wave in the channel and spatial dimensions. The structure diagram is shown in Figure 2.

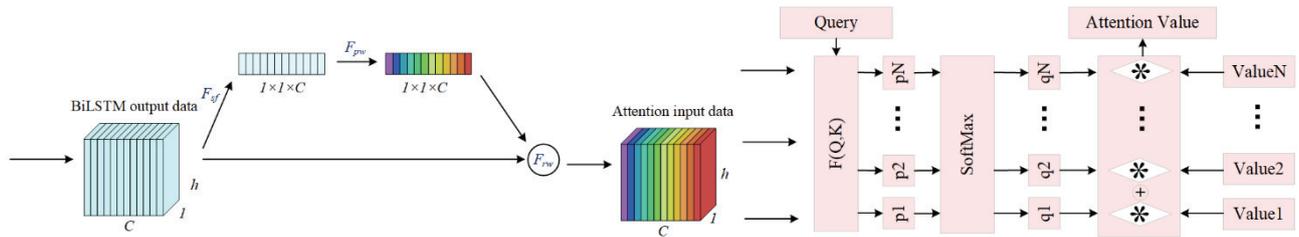


Fig.2 Architecture of mixed attention mechanism

To focus on the relevance of channel dimensions, spatial dimension information [26] needs to be compressed, and the expression is as follows.

$$s = F_{sf}(d_c) = \frac{1}{h} \sum_{i=1}^h d_c(i) \quad (9)$$

Where, d_c is the c th characteristic of the input matrix; h is the input feature height. Channel weights are obtained by hidden layers, including global pooling, ReLU, fully connected layer and Sigmoid. The expression is as follows:

$$P_c = F_{pw}(s, W) = \sigma(g(s, W)) = \sigma(W_2 \delta(W_1 s)) \quad (10)$$

Where, W_1 is the weight of first fully connected layer, activation function δ is ReLU, W_2 is the weight of the second fully connected layer, and activation function σ is sigmoid. The original features are re-calibrated by the weight of channels, and finally the weighted channel dimension information is obtained.

$$r = F_{rw}(d_c, P_c) = P_c \cdot d_c \quad (11)$$

Where, F_{rw} represents the context channel product and d_c represents the feature graph.

To focus on the correlation of spatial dimensions, multi-layer neural network (MLP) is applied for the weight coefficient ^[27-28] of the corresponding *Value* of each *Key* by calculating correlations between *Query* of certain element and its *Keys*. The expression is as follows.

$$p_i = \text{Similarity}(\text{Query}, \text{Key}_i) = \text{MLP}(\text{Query}, \text{Key}_i) \quad (12)$$

The correlation score is converted numerically, and the original calculated score is sorted into a probability distribution with the sum of the weights of all elements equal to 1 through normalization. It can also be expressed as the critical weights of important elements through the internal mechanism of SoftMax. The formula is as follows.

$$q_i = \text{softmax}(p_i) = \frac{\exp(p_i)}{\sum_j(p_j)}, i = 1, 2, \dots, N; j = 1, 2, \dots, i. \quad (13)$$

Finally, we take the weighted sum of the *Values* to get the final *Attention Value*.

$$\text{Attention}(\text{Query}, \langle \text{Key}, \text{Value} \rangle) = \sum_{i=1}^N q_i \cdot \text{Value}_i \quad (14)$$

Because the global pooling layer will ignore some important characteristics, the global attention mechanism is added as auxiliary information. The overall architecture of the proposed model is shown in Figure 3. The model training pseudocode is shown in Tab.1.

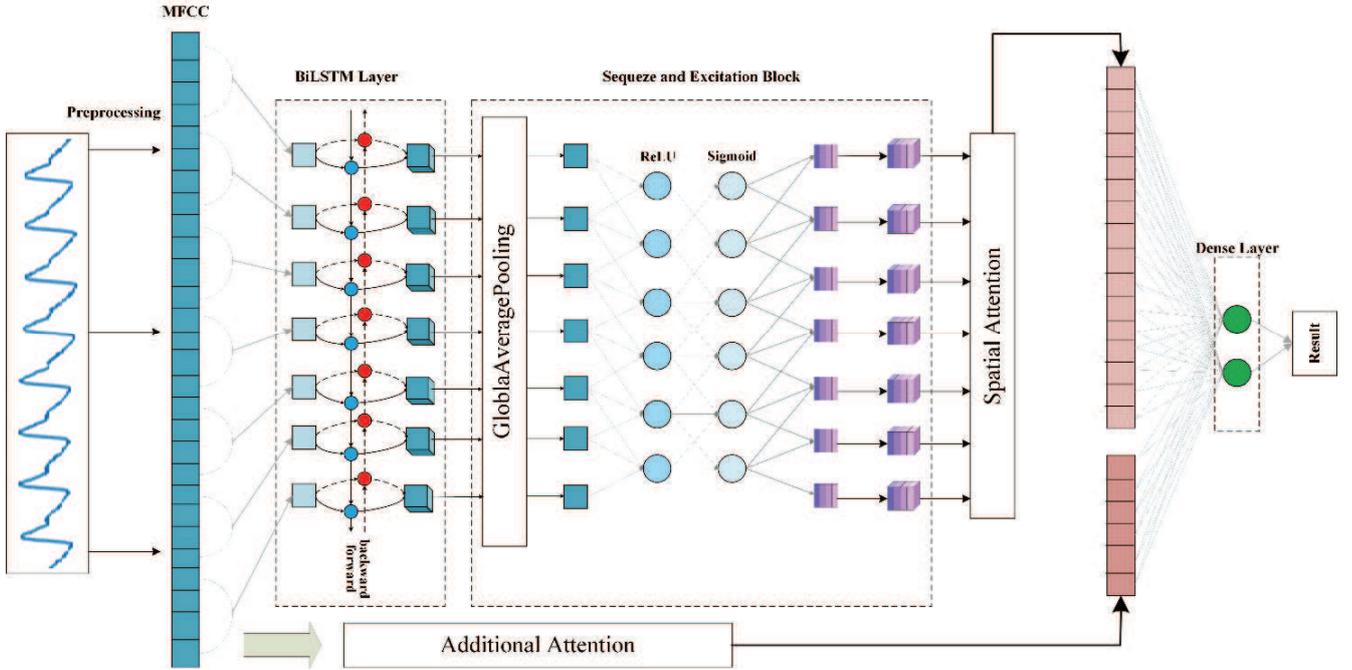


Fig.3 The improved BiLSTM model with mixed attention mechanisms
Tab.1 Prediction of H-type Hypertension based on pulse wave MFCC features

Input: labelled instances $\{(x^{(n)}, y^{(n)})\}$

Output: Predictive label

- 1 iteration ← 0
- 2 while number of epochs > iteration do
- 3 iteration ← iteration + 1
- 4 for each batch do
- 5 BiLSTM layer: $x_B = W_B x + b_B$ Attention layer: $x_{AA} = \text{Attention}(x)$
- 6 Compress $x_C = F_{sf}(x_B)$
- 7 FC layer: $x_{fc1} = \text{ReLU}(W_{fc1} x_C + b_{fc1})$
- 8 FC layer: $x_{fc2} = \text{sigmoid}(W_{fc2} x_{fc1} + b_{fc2})$
- 9 Attention layer: $x_{AB} = \text{Attention}(x_{fc2})$

5. Experimental Results and Analysis

5.1 Pulse Wave Dataset

The experimental dataset in this literature comes from clinical pulse wave instances of suspected patients in H-type hypertension provided by Longhua Hospital affiliated to Shanghai University of TCM and Hospital of Integrated Traditional Chinese and Western Medicine. The cases are collected by a pulse instrument. Before sampling, the patient should relax for more than 5 minutes, and the sampling time is 60 seconds. The sampling contact pressure depends on the patient's pulse depth, and the sampling frequency consists of 200Hz, 700Hz, and 1000Hz. For data consistency, we lower the sampling frequency from 700Hz and 1000Hz to 200Hz. A total of 337 instances were collected, including 140 males (41.54%) with an average age of (66.26 ± 10.37) years and 197 females (58.46%) with an average age of (71.36 ± 8.51) years. There were 129 patients with H-type hypertension, accounting for 38.28% of the total cases, and 208 with non-H-type hypertension, accounting for 61.72% of the total. The case distribution is shown in Fig.4.

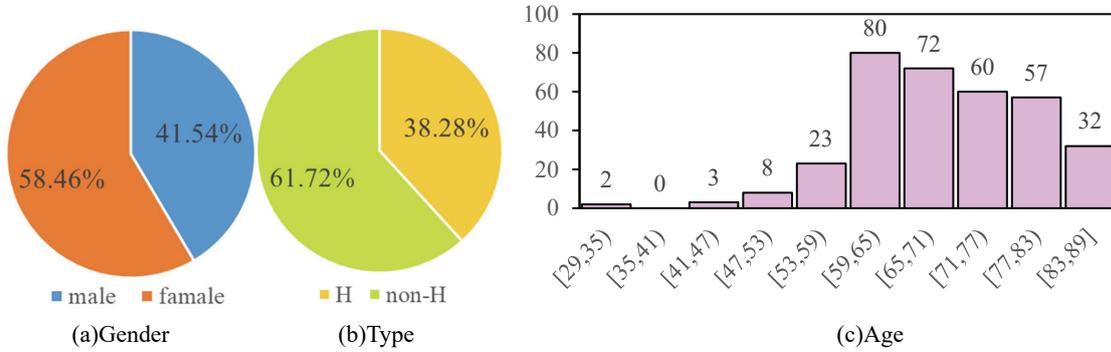


Fig.4 Distributions of pulse wave instances

5.2 Hyperparameters and Evaluation Indexes

Many factors affect clinical pulse wave sampling, even leading to incomplete data. After specific pretreatment, the inconsistency of MFCC feature length still occurs. Therefore, the maximum length of MFCC feature is set to 153 as the standard, and 0 is used to be filled if the length is insufficient. The padding MFCC features are used as network input. The BiLSTM layer has 32 units, dropout is 0.1, and the model uses Adam Optimizer. The initial learning rate is $\alpha=0.001$, and the exponential decay rate is $\beta_1 = 0.9$, $\beta_2 = 0.999$. The batch size is set to 32 and the training epoch is 200. In this literature, four evaluation indexes for binary classification, Receiver Operating Characteristic (ROC) curve and area under curve (AUC) and the Area below Precision-Recall curve called AP(Average Precision) were used to evaluate the classification performance of rhinitis instances, including: Accuracy, Recall, Precision and F1-score. In addition, we consider SNR and RMSE as evaluation indicators of pulse wave denoising as well.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (18)$$

Where, true positive (TP) represents the number of instances with H-type hypertension correctly classified as H-type hypertension; false positive (FP) represents the number of instances with non-H-type hypertension classified as H-type hypertension; true negative(TN) represents the number of instances with non-H-type hypertension

correctly classified as non-H-type hypertension; false negative (FN) represents the number of instances with h-type hypertension classified as non-H-type hypertension.

5.3 Filtering Process

The high-frequency noise and baseline drift can decrease effectively by filtering noises from the original pulse wave. In this literature, wavelet functions with different orders and basic functions are applied to perform noise reduction on original pulse signal. Figure 5 shows the curves of SNR and RMSE after denoising when the wavelet order number is N , and the wavelet basis is *db*, *sym* and *Coif*, respectively. Tab.2 shows the corresponding SNR and RMSE values of the curve in Figure 5. The analysis shows that *sym7* wavelet has a maximum SNR=45.5407 and a minimum RMSE=0.03723. Therefore, *sym7* is selected as the wavelet base. This is because Symlets series *sym* wavelet bases have better symmetry than *db* and *coif*, which can effectively reduce phase distortion and noises in reconstruction of pulse signal. When the support range of *sym7* wavelet base is 13 and the vanishing moment is 7, the pulse wave has good regularity, which can concentrate the pulse wave energy and effectively reduce boundary effect of wavelet transform.

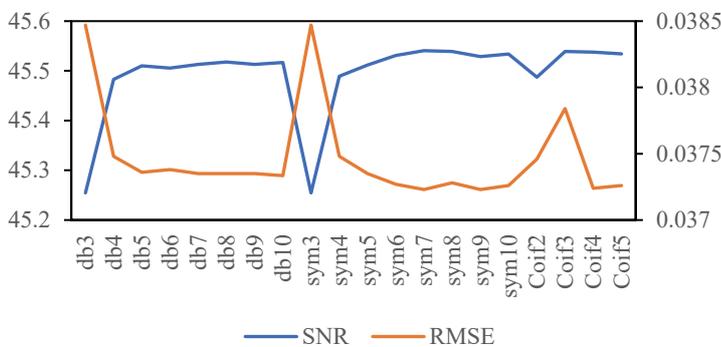


Fig.5 Denoising of different wavelet bases

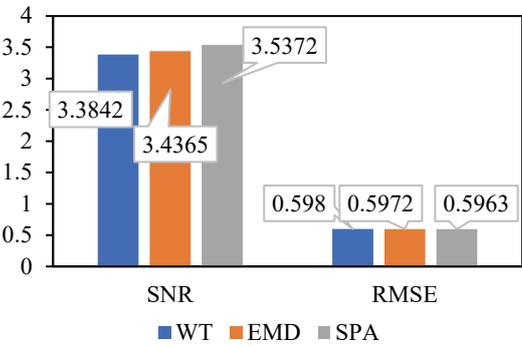


Fig.6 Baseline drift correction of different methods

Tab.2 Comparison of denoising effect of different wavelets

N	<i>db</i>		<i>Sym</i>		<i>Coif</i>	
	SNR	RMSE	SNR	RMSE	SNR	RMSE
1	29.6908	0.23089	29.6908	0.23089	43.5351	0.04679
2	43.3575	0.04787	43.3575	0.04787	45.4871	0.03746
3	45.2548	0.03847	45.2548	0.03847	45.5388	0.03784
4	45.4826	0.03748	45.4891	0.03748	45.5376	0.03724
5	45.5101	0.03736	45.5115	0.03735	45.5341	0.03726
6	45.5056	0.03738	45.5308	0.03727	/	/
7	45.5129	0.03735	45.5407	0.03723	/	/
8	45.5176	0.03735	45.539	0.03728	/	/
9	45.5130	0.03735	45.5287	0.03723	/	/
10	45.5168	0.03733	45.5336	0.03726	/	/

Fig.7 shows the approximation and detail components of each order in the single cycle after wavelet decomposition. Fig.7(a) shows the approximate components of low frequency coefficient of pulse wave, and the different layers of approximation component reflects the variation of low frequency signal. Fig.7 (b) shows the detailed components of high-frequency coefficient. Through the detailed components refactoring we can eliminate the high-frequency signal and noises. As shown in Fig.7(b), high-frequency information exists in the first three layers of detailed components at some sampling points of pulse wave. The detailed components decomposed in layer 1 and layer 3 correspond to the high-frequency interference from 20Hz to 120Hz, and the detailed components decomposed in layer 2 correspond to the high-frequency interference of 50Hz. The fourth and fifth detailed components are the main pulse wave signal. The decomposition coefficients of the first three layers of detailed components are processed by soft threshold and reconstructed to achieve better high effect of high-frequency noise removal. Figure 6 shows the comparative results of various baseline drift removal methods. According to the analysis, for the pulse wave instances of H-type hypertension, SNR of SPA increased by 4.52% and RMSE decreased by 0.28% compared with

WT. Compared with EMD, SNR of SPA was increased by 2.93% and RMSE was decreased by 0.15%. It shows that SPA is more suitable for baseline drift removal of pulse wave. Fig. 8 shows the filtering process. e.g. high-frequency noise and baseline drift removal. It can be seen that the filtering process reduce the average amplitude of pulse wave, eliminate the extreme points, and maintain stable periodicity of pulse wave.

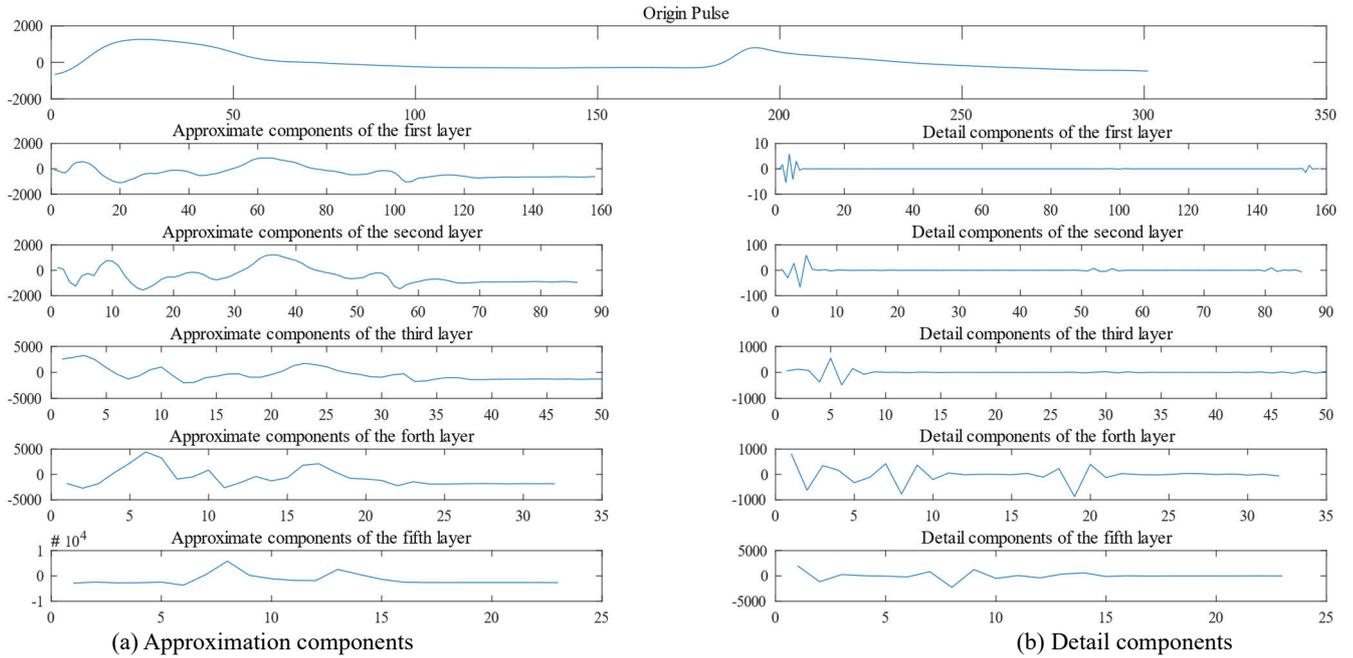


Fig.7 Various components in single cycle after wavelet base, sym7 decomposition

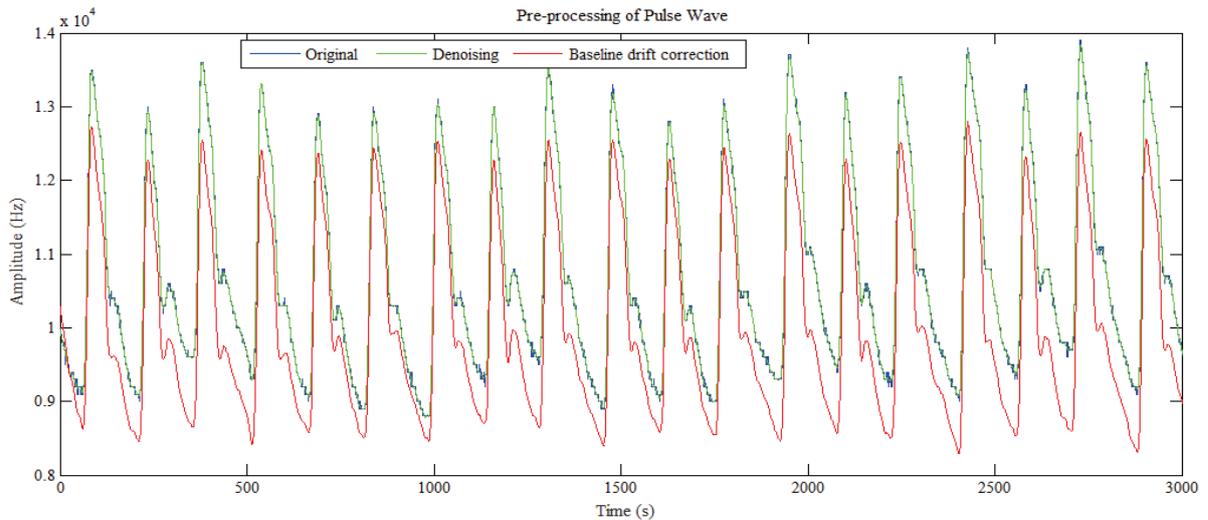


Fig.8 The pulse wave after various filtering

5.4 MFCC-based Feature Extraction

Fig.9 shows the time-domain pulse wave and the IMF components and the residuals of each order in a frame via EMD with adaptive noise. The higher the order, the lower the frequency. Fourier transform can only reflect high-frequency information, but by EMD with adaptive noise, we can obtain different scale frequency information, indicating that pulse wave signal can be expressed with different resolution signals. At the same time, the decomposed pulse wave only retains the same frequency component in the three IMFs and residual components, which reduces mode mixing.

In order to evaluate the effect of filtering and time-frequency domain feature transformation on classification performance, we deal with the original pulse wave instances as the four methods, i.e. the original pulse wave instances (Pulse), the filtered time-domain instance (Handled Pulse), MFCC feature instances (MFCC) and Improved MFCC feature instances (Improved MFCC). The various classification models are applied for comparison of performance, such as Random forest (RF), Decision Tree, support vector machine (SVM), Adaboost, BiLSTM,

BA(BiLSTM with Attention) and BSEAA(BiLSTM with SE-block-Attention and additional Attention). Fig.10 shows evaluation indexes of the means of various models. For time-domain pulse wave instances, different evaluation indexes of each model show different data distributions. The distribution of Accuracy and Precision for Pulse are concentrated, and distribution of Handled Pulse is concentrated in Accuracy and F1-score. At the same time, compared with the time-domain instances, the frequency-domain MFCC feature instances have higher means and smaller variance in each evaluation index, indicating that the features in frequency-domain have better discriminability than in time-domain. Meanwhile, Improved MFCC has achieved higher classification accuracy and generalization performance than MFCC.

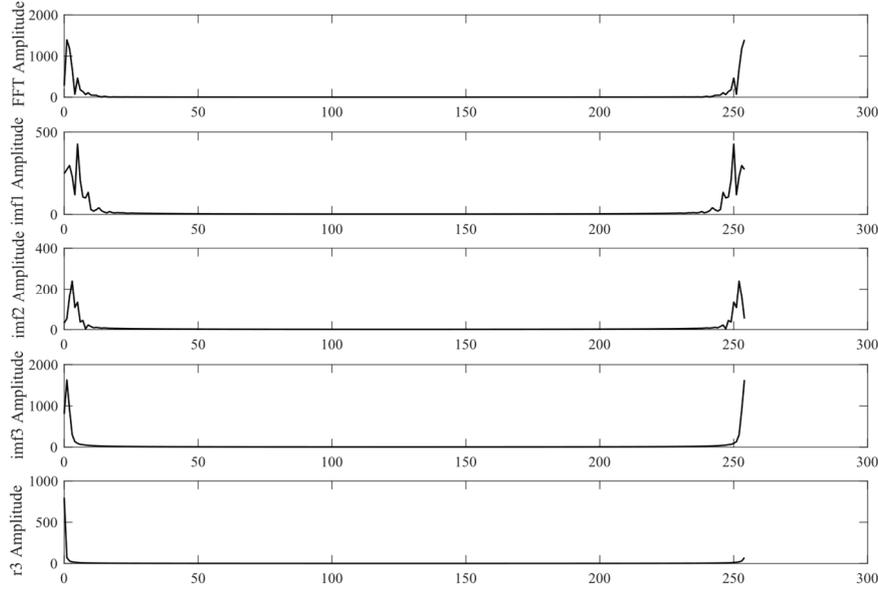


Fig.9 Comparison of pulse wave decomposition via FFT and EMD with adaptive noise

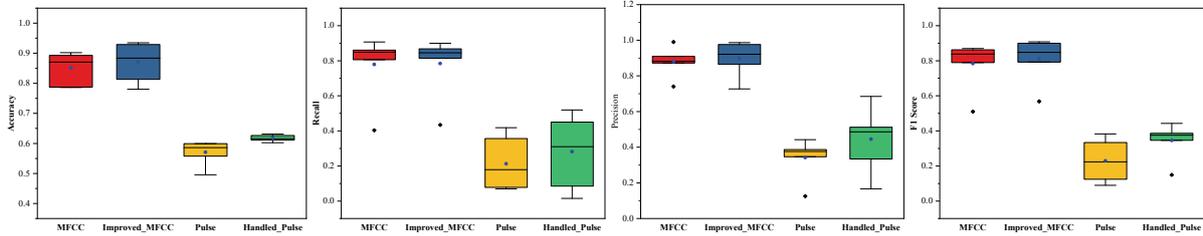


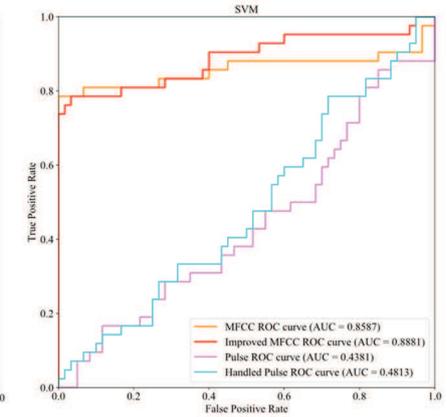
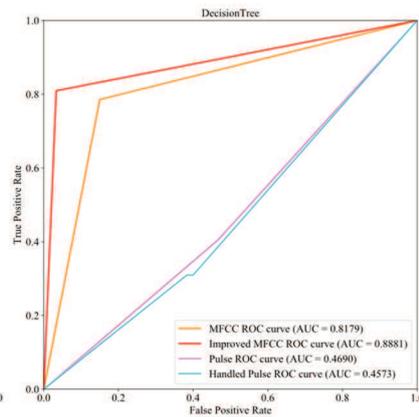
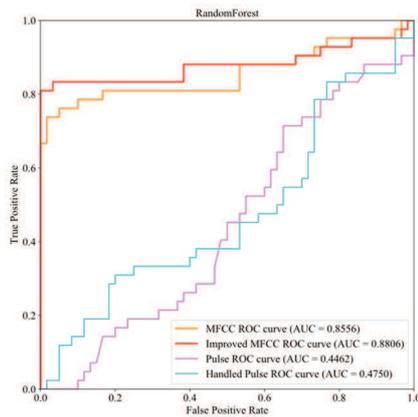
Fig.10 Comparison of performance for various models

Tab.3 shows the classification results of various models. For traditional RF model, compared with MFCC method, the recall of Improved MFCC increases by 0.78%, but the precision falls by 2.45%, respectively. This is because RF model has high sensitivity to class imbalanced instances. The prediction is biased to preference of majority non-H-type class, which leads to high Precision. However, compared with MFCC method, F1 score of Improved MFCC improves by 0.45%. This is because we add the adaptive white noises via HHT and obtain marginal spectrum, which can decrease mode mixing, and reduce the possibility of high-frequency signals with lower amplitude in short time interval. In addition, it accurately reflects the pulse wave as the actual frequency of non-stationary signal, increases the representative characteristics of the MFCC coefficients, so as to improve classification performance.

Fig.11 shows the ROC and PR curves corresponding to the various models and methods. It can be seen that whether for RF and SVM or deep learning model such as BiLSTM, Improved MFCC method has higher various evaluation performance than MFCC, Handled Pulse and Pulse. Therefore, Improved MFCC has better classification accuracy and generalization performance. It means that the filtered frequency-domain MFCC features have higher discriminability than time-domain features, which can effectively reduce mode mixing and enhance classification performance.

Tab.3 Evaluation results of different models and methods

Methods	Models	Accuracy	Recall	Precision	F1 Score
MFCC	RF	0.8662±0.0625	0.8062±0.1293	0.8889±0.1571	0.8244±0.0712
	DecisionTree	0.7862±0.1560	0.9070±0.0570	0.7406±0.2157	0.7896±0.1145
	SVM	0.7689±0.1047	0.4031±0.2875	0.9902±0.0139	0.5100±0.2976
	Adaboost	0.8752±0.0888	0.8527±0.0290	0.8743±0.1777	0.8511±0.0858
	BA	0.8830±0.0457	0.8405±0.0380	0.8795±0.0933	0.8528±0.0505
	BSEAA	0.9020±0.0194	0.8450±0.0610	0.9101±0.0916	0.8690±0.0136
Improved MFCC	RF	0.8723±0.0237	0.8140±0.1005	0.8644±0.0961	0.8289±0.0322
	Decision Tree	0.8129±0.0764	0.8992±0.0791	0.7270±0.1143	0.7928±0.0556
	SVM	0.7803±0.0783	0.4341±0.2083	0.9872±0.0181	0.5683±0.2287
	Adaboost	0.8959±0.0586	0.8682±0.0719	0.8669±0.0979	0.8659±0.0756
	BA	0.8961±0.0225	0.8372±0.0760	0.8959±0.0765	0.8599±0.0295
	BSEAA	0.9348±0.0220	0.8527±0.0667	0.9748±0.0209	0.9077±0.0345
Origin Pulse	RF	0.5993±0.0377	0.1085±0.0667	0.3865±0.1628	0.1650±0.0993
	Decision Tree	0.4955±0.0613	0.3566±0.1523	0.3463±0.0183	0.3339±0.0829
	SVM	0.5994±0.0017	0.0775±0.0439	0.3651±0.0449	0.1246±0.0608
	Adaboost	0.5577±0.0461	0.2481±0.1400	0.3859±0.0908	0.2812±0.1001
	BA	0.6172±0.0016	0.0155±0.0219	0.1667±0.2357	0.0284±0.0401
	BSEAA	0.5727±0.0338	0.4186±0.2424	0.4425±0.0123	0.3828±0.1632
Handled Pulse	RF	0.6142±0.0052	0.4186±0.1245	0.4952±0.0067	0.4439±0.0735
	Decision Tree	0.6024±0.0170	0.4496±0.1145	0.4763±0.0243	0.4565±0.0724
	SVM	0.6113±0.0157	0.2016±0.1825	0.3340±0.2381	0.2341±0.1850
	Adaboost	0.6261±0.0203	0.5194±0.0290	0.5119±0.0252	0.5153±0.0239
	BA	0.5993±0.0261	0.0698±0.0987	0.1250±0.1768	0.0896±0.1266
	BSEAA	0.6321±0.0148	0.0853±0.0290	0.6852±0.2238	0.1492±0.0462



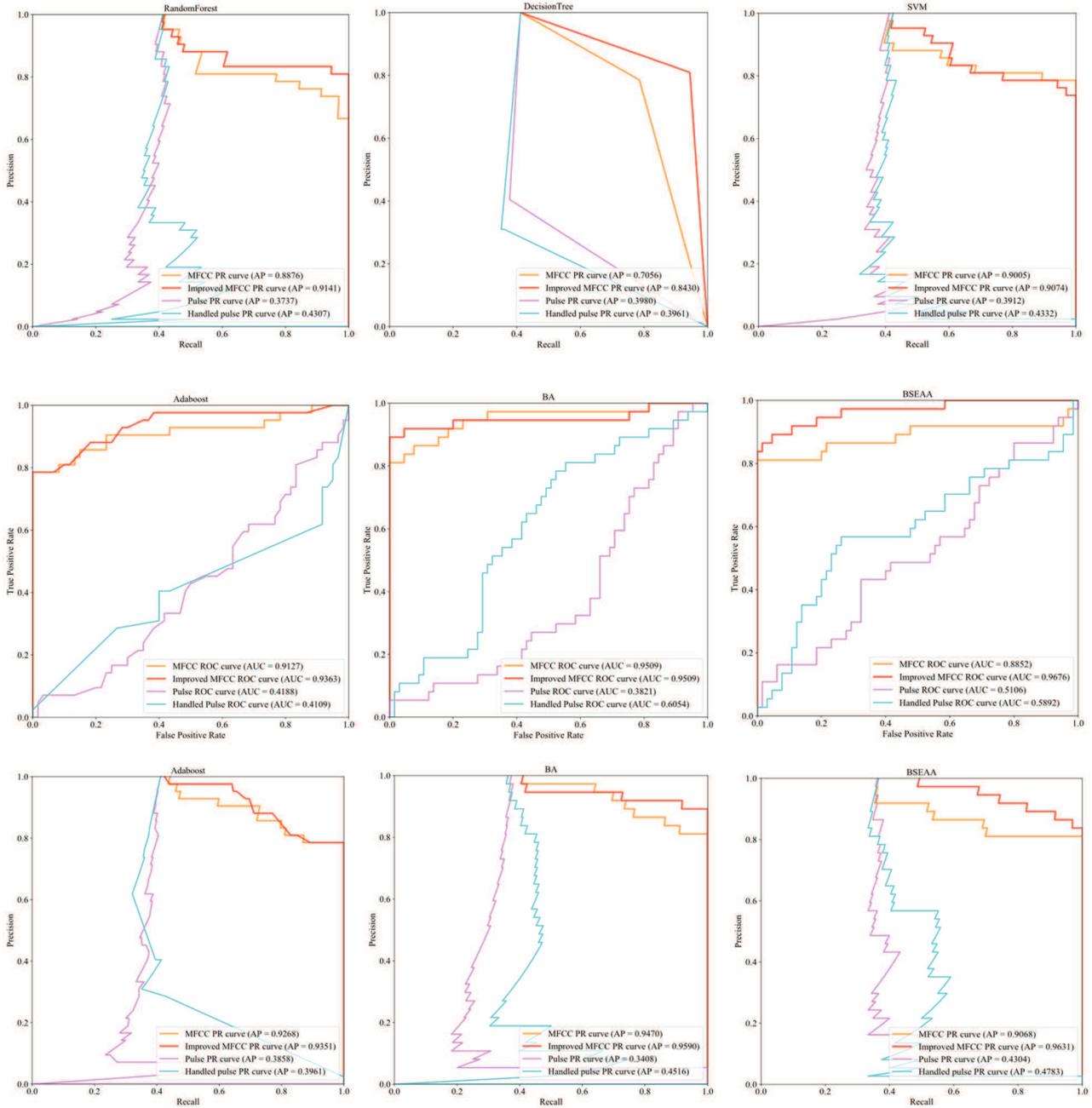


Fig.11 The ROC curves of various methods

5.5 Ablation Study

To evaluate the performance of attention-based mechanism, we conduct ablation study by adding the spatial and channel attention blocks based on BiLSTM model, respectively. The detail is shown in Tab.4 and Figure 12. Fig.13 shows the ROC and PR curves of various models. Compared with BiLSTM, the Accuracy, Recall, Precision, F1-score, AUC and AP of BA model increased by 4.73%, 13.18%, 0.71%, 8.58%, 4.45%, and 2.73%, respectively. It shows that BA has better classification performance than BiLSTM. This is because that the spatial attention mechanism can effectively extract the location important features of pulse wave, and provide more effective information for the fully connected layer, and improve generalization performance of model.

Compared with BiLSTM model, the Accuracy, Recall, Precision, F1 score, AUC and AP of BSE (BSE, BiLSTM with SE block) improve by 4.44%, 11.63%, 1.91%, 8.07%, 4.25% and 3.12% respectively, so BSE has better classification performance on MFCC features of frequency-domain pulse wave. This is because that the channel attention mechanism assigns different weights for different channel dimensions of BSE, which enhances feature extraction ability of each channel, and then improves classification performance.

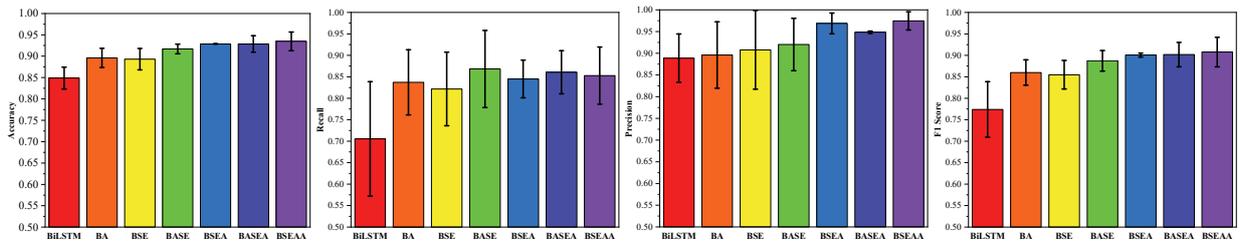


Fig.12 Performance of different models

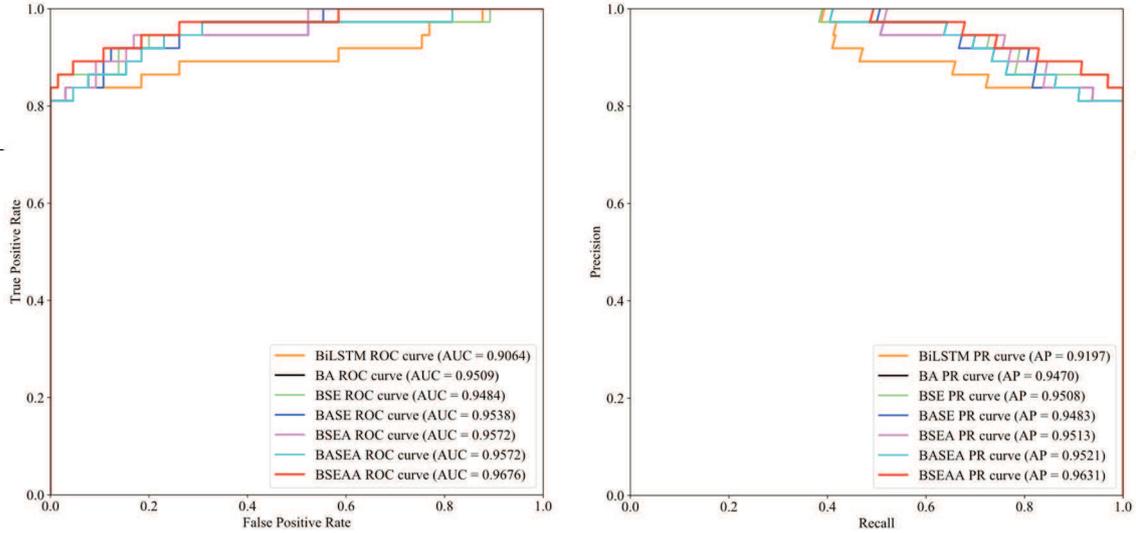


Fig.13 The ROC、PR curve of different models

According to Tab.4, the BSEA (BiLSTM with SE-block-Attention) model uses channel Attention first, and then adds spatial Attention, and its F1-score reaches 0.9006. BASE (BiLSTM with Attention-SE-block) uses spatial Attention first and then adds channel Attention, and its F1-score is 0.8871. It can be seen from the results that BSEA outperform BASE, and the sequence of adding Channel or Spatial Attention mechanism based BiLSTM has some effects on classification performance.

Tab.4 Evaluation indexes of various BiLSTM models with different attention mechanisms

Model	Accuracy	Recall	Precision	F1 Score
BiLSTM	0.8488±0.0255	0.7054±0.1334	0.8888±0.0556	0.7741±0.0647
BA	0.8961±0.0225	0.8372±0.0760	0.8959±0.0765	0.8599±0.0295
BSE	0.8932±0.0250	0.8217±0.0856	0.9079±0.0908	0.8548±0.0334
BASE	0.9169±0.0112	0.8682±0.0897	0.9204±0.0603	0.8871±0.0240
BSEA	0.9288±0.0003	0.8450±0.0439	0.9690±0.0339	0.9006±0.0045
BASEA	0.9288±0.0192	0.8605±0.0502	0.9486±0.0028	0.9017±0.0285
BSEAA	0.9348±0.0220	0.8527±0.0667	0.9748±0.0209	0.9077±0.0345

Before adding attention mechanism, the model needs to compress data via global average pooling, which may lose some detailed features. Therefore, after adding attention mechanism, we add global spatial attention mechanism to supplement the lost features. The BSEAA model is obtained by adding global spatial attention mechanism on the basis of BSEA model. Experimental results show that compared with BSEA, Accuracy of BSEAA increased by 0.6%, reaching to the highest of 0.9348, Recall increased by 0.77%, Precision increased by 0.58%, F1 score increased by 0.71%, AUC increased by 1.08%, and AP increased by 1.2%. BASEA (BiLSTM with Attention-SE-Block and Additional Attention) is obtained by adding global spatial attention mechanism on the basis of BASE model. Compared with BASE model, Accuracy of BASEA increased by 1.19%, Recall decreased by 0.77%, Precision increased by 2.82%, F1 score increased by 1.46%, AUC increased by 0.29%, and AP increased by 0.32%. Therefore, on the basis of BiLSTM, adding global spatial attention mechanism can effectively improve the overall classification performance because the global pooling layer of channel attention mechanism ignores some

important features, and the global attention mechanism with parallel structure can supplement the features lost by global pooling to a certain extent, and enhance classification performance effectively. Compared with BASEA, Accuracy of BSEAA model increased by 0.6%, Recall decreased by 0.78%, Precision increased by 2.62%, F1 score increased by 0.6%, AUC increased by 1.04%, and AP increased by 1.1%. Therefore, BSEAA model achieves the best classification accuracy and generalization performance among various improved models based on BiLSTM.

6. Discussion

6.1 Analysis of Computational Complexity

Fig.14 shows the training time and number of parameters of 3-fold cross-validation for various BiLSTM models with different attention mechanism. It can be seen that after the channel attention mechanism is added to BiLSTM model, the parameters increase by 66, accounting for only 0.55%, and F1 score increases by 8.07%. From the perspective of economics, it shows that the BiLSTM model obtains higher benefits by adding channel attention mechanism. Compared with channel attention mechanism, spatial attention mechanism has higher temporal and spatial complexity. Increasing spatial attention mechanism can improve F1 score by 8.58%, which is similar to that of channel attention mechanism, but the number of parameters nearly doubles. Therefore, when the model is of high complexity, the channel attention mechanism is preferred to improve the classification performance. At the same time, it can be seen that with the increase of number of instances, the training time and the number of model parameters also increase. The number of parameters is proportional to the training time, and the training time difference of each model is about 20s. With the addition of attention mechanism module, the number of parameters increases and the accuracy of classification improves remarkably.

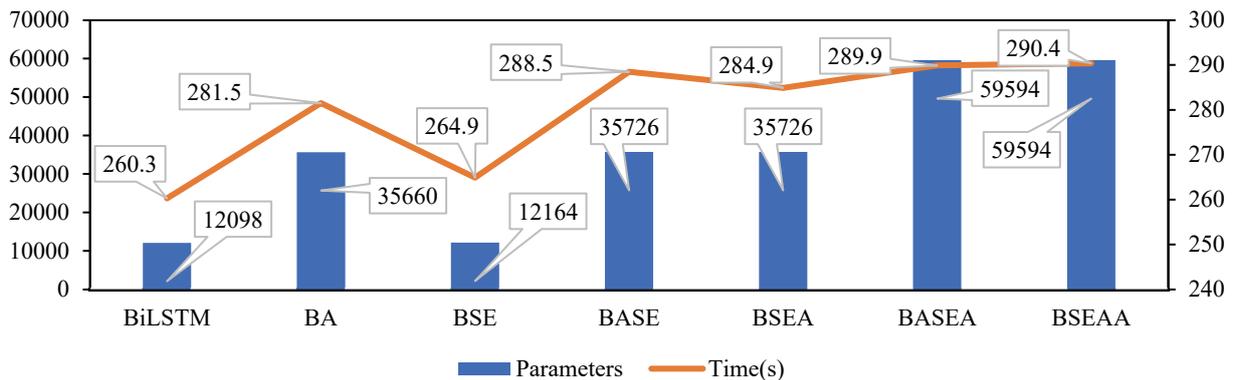


Fig.14 Computational complexity of improved BiLSTM models with different attention mechanisms based on MFCC.

6.2 Analysis of Feature Importance

In this literature, we rank 36 pulse wave MFCC features in frequency-domain by calculating Gini impurity of RF algorithm as shown in Fig.15. The top 5 features are the 8th in Second Order Difference(S), namely S_8., the 12th in First Order Difference (F), F_12, and S_10, S_1, F_8. Most of the features that have major influence on classification are the first or second order difference coefficient. The first-order difference accounts for 42.64%, and the second-order difference accounts for 56.9%, totaling more than 99%. Therefore, the dynamic coefficient characteristics of pulse wave describe the correlation between adjacent frames of pulse wave in details, which has a very important influence on classification performance. At present, many scholars have studied the classification of pulse waves on a single period. The static characteristics of MFCC have a certain impact on classification performance of model. However, the pulse waves between different periods may have more correlations, and more refined classification may be conducted on pulse wave instances of H-type hypertension. Secondly, there are many characteristic components in the first-order and second-order difference coefficient, such as F_8 and S_12, which have great influence on classification. The actual physical meaning of these coefficients in relation to pulse waves has not been directly explained so far. It may play a great role in classification of H-type hypertension. In the future, we would study the characteristics of the first-order and second-order differential coefficients deeply, and explore the correlation between the coefficients of each order and the classification of pulse wave.

In this literature, we calculate the correlation between time-domain pulse graph features and classification of H-type hypertension, and work out the feature importance ranking. As shown in Figure 13, the top 4 importance features, i.e. w_1/T , h_4/h_1 , h_4 and t_1/t_4 , are all higher than 5%. w_1/t refers to the duration of aortic pressure rise, which is related to the appearance time of wave before repeat wave and peripheral resistance. h_4/h_1 mainly reflects the level of peripheral resistance. When the peripheral blood vessels contract, the resistance increases and h_4/h_1 increases (>0.45). On the contrary, when the peripheral resistance decreases, h_4/h_1 decreases (<0.30). h_4 is the amplitude of the dicotic notch, representing the height from the bottom of dicotic notch to the baseline of pulse wave. The height of dicotic notch corresponds to the diastolic blood pressure, which is related to the peripheral resistance of arteries and the function of aortic valve closure. The top 3 importance features all reflect the impact of peripheral vascular resistance of H-type hypertension patients on classification. It also shows that H-type hypertension is an independent risk factor for atherosclerosis and arteriosclerosis. t_1/t_4 is related to the function of cardiac ejection. The larger t_1/t_4 , the slower the acute ejection period of the heart, the weaker systolic function of left ventricular.

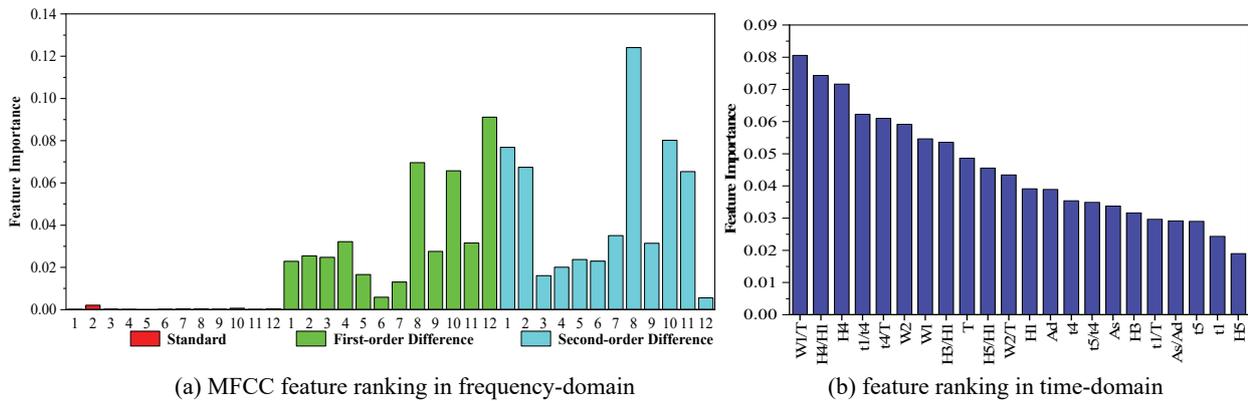


Fig.15 Feature importance ranking of Pulse wave in frequency-domain and time-domain

Therefore, we can research on the correlations between the pulse wave features both in time-domain and frequency-domain and classification of H-type hypertension, assist clinicians to furtherly study the influence of patient's vascular stress and blood flow on prediction of H-type hypertension, further to seek the occurrence mechanisms of H-type hypertension to provide reference for timely prevention and treatment.

7. Conclusion

In this literature, according to the characteristics of pulse wave in TCM, we transform the pulse wave signal into MFCC features in frequency-domain, and build an improved BiLSTM model with mixed channel attention mechanisms to predict suspected patients in H-type hypertension via pulse wave MFCC features. The experimental results show that the MFCC features in frequency-domain are better distinguishable than that in time-domain. Compared with traditional machine learning, the proposed model has higher classification accuracy and generalization performance, and has a good reference value for clinical diagnosis of H-type hypertension.

Funding

This research is funded by the National Natural Science Foundation of China (81973749).

Conflicts of Interest

None declared

Ethical Approval

Not required

REFERENCE

- [1] Liu Qiang Huang, Chong Xin Wu, et al. Clinical characteristics of H-type hypertension and its relationship with the MTHFR C677T polymorphism in a Zhuang population from Guangxi, China[J]. *Journal of Clinical Laboratory Analysis*,2020,34: e23499. <https://doi.org/10.1002/jcla.23499>.
- [2] Zhang, J., Liu, Y., Wang, A. et al. Association between H-type Hypertension and Asymptomatic Extracranial Artery Stenosis. *Sci Rep* 8, 1328 (2018). <https://doi.org/10.1038/s41598-018-19740-0>.
- [3] Hu, S., Ren, L., Wang, Y. et al. Homocysteine-lowering therapy and early functional outcomes of ischemic patients with H-type hypertension: a retrospective analysis of CNSR. *Australas Phys Eng Sci Med* 38, 785–791 (2015).<https://doi.org/10.1007/s13246-015-0406-x>.
- [4] ZHOU Chaofei, GUO Yi, DENG Juclin. H-type Hypertension and Atrial Fibrillation [J]. *Advances in Cardiovascular Diseases*,2019,40(9):1205-1207.
- [5] WANG Yiqin, LI Fufeng, YAN Haixia, et al. A Review of Digitalized Traditional Diagnostic Techniques [J]. *Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*,2007,9(3):96-101.
- [6] WEI Hong, LIU Minglin, ZHENG Hongxin. The modern research route of pulse-taking [J]. *Journal of Basic Chinese Medicine*,2004,10(2):69-71.
- [7] ZHANG X, XU L, CHEN K, et al. A new method for locating feature points in pulse wave using wavelet transform [C]. *W RI World Congress on Computer Science and Information Engineering*,2009,5: 367-371.
- [8] Lu Hongsheng, Ye Weiqiong, Hu Yue. Research on Feature Extraction of Pulse Wave Signal[J]. *Information and Communications*,2019.9(7):3-5. (In Chinese)
- [9] Bao-cun FAN, Yan WANG, Chen-chen HUANG, et al. Pulse Wave Signal Feature Recognition Based on Time-domain Differential Period Ratio[J]. *Acta Photonica Sinica*, 2020, 49(12): 1206003-1206003. (In Chinese)
- [10] Zhi-yu Luo, Ji Cui, Xiao-juan Hu, et al. A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave.[J] *BioMed Research International*,2018(7):1-12
- [11] FENG Bing, LI Shaozi. Unsupervised clustering analysis of human-pulse signal in traditional Chinese medicine[J]. *CAAI Transactions on Intelligent Systems*, 2018, 13(4): 564-570. (In Chinese)
- [12] Zhang Shiyu, Yang ke, Xia Chunming et al. Research on dimension reduction and classification of pulse signal features based on random forest [J].*Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology*. 2020,22(07): 2418-2426. (In Chinese)
- [13] Zhang N . Pulse Condition Recognition Based on Convolutional Neural Network with Dimension Enlarging[J]. *Computer Science*, 2018,6(45):506-507,535. (In Chinese)
- [14] LIU Guohua, ZHOU Wenbin. Pulse wave signal classification algorithm based on time-frequency domain feature aliasing using convolutional neural network [J]. *Journal of Jilin University (Engineering and Technology Edition)*. 2020,50(05): 1818-1825. (In Chinese).
- [15] YAN Jianjun, CHEN Songye, YAN Haixia, et al. Wrist pulse analysis and recognition based on recurrence plot and convolution neural network[J]. *Computer Engineering and Applications*. 2020,56(7):170-175. (In Chinese)
- [16] Kang G, Xiao Y, Liu J, et al. Tatt-BiLSTM: Web service classification with topical attention-based BiLSTM. *Concurrency Computat Pract Exper*. 2021. <https://doi.org/10.1002/cpe.6287>
- [17] HUANG Wei, YU Jiangwei. Exploration and analysis research of medical consultation content and objectification in traditional Chinese medicine [J]. *China Journal of Traditional Chinese Medicine and Pharmacy*.2019,34(8):3666-3668.
- [18] JI Zhong, LIU Xu. Study on feature points recognition of pulse wave based on waveform feature and wavelet [J]. *Chinese Journal of Scientific Instrument*.2016,37(2):379-386.
- [19] T.L. Thomas, C.S. Antony. Detection of transients in noise with the undecimated discrete wavelet transform[J]. *IEEE Transactions on Signal Processing*.,2000, 48(5): 1458~1462.
- [20] Bibin Sam Paul S, Antony Xavier Glittas, Lakshminarayanan Gopalakrishnan. A low latency modular-level deeply integrated MFCC feature extraction architecture for speech recognition[J]. *INTEGRATION-THE VLSI JOURNAL*. Volume 76, January 2021,

Pages 69-75.

- [21] Muqing Deng, Tingting Meng, Jiuwen Cao, et al. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks[J]. *Neural Networks*. Volume 130, October 2020, Pages 22-32.
- [22] Shuaipu Jin, Xiufeng Wang, Leilei Du, et al. Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN[J]. *Applied Acoustics* 172 (2021) 107562. Pages 1-11.
- [23] LI Yang, DONG Hongbin. Text sentiment analysis based on future fusion of convolution neural network and bidirectional long short-term memory network[J]. *Journal of Computer Applications*,2018,38(11):3075-3080.
- [24] Pham, T.D. Time–frequency time–space LSTM for robust classification of physiological signals. *Sci Rep* 11, 6936 (2021). <https://doi.org/10.1038/s41598-021-86432-7>
- [25] GREFF K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: a search space odyssey[J]. *IEEE Transactions on Neural Network and Learning Systems*,2015,28(10):2222-2232.
- [26] Jie Hu, Li Shen, Gang Sun, et al. Squeeze-and-Excitation Networks. *IEEE Conference on Computer Vision and Pattern Recognition*. hu2018senet.
- [27] Deepak Jain, Akshi Kumar, Geetanjali Garg. Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN[J]. *Applied Soft Computing*. Volume 91, June 2020, 106198.
- [28] Tharindu Fernando, Simon Denman, Sridha Sridharan, et al. Soft-Hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection[J]. *Neural Networks*. Volume 108, December 2018, Pages 466-478.