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

Jingdong Yang1 · Lei Chen1 · Shuchen Cai1 · Tianxiao Xie2 · Haixia Yan2

Received: 5 November 2021 / Accepted: 2 February 2022 / Published online: 8 July 2022
© Taiwanese Society of Biomedical Engineering 2022

Abstract

Purpose 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 article, we propose an effective prediction on frequency-domain pulse wave features.

Methods 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. Finally, an improved BiLSTM model, combined with mixed attention mechanism is built to be applied for prediction of H-type hypertension.

Results With 337 clinical cases from the Longhua Hospital affiliated to Shanghai University of TCM and Hospital of Integrated Traditional Chinese and Western Medicine, the threefold 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. 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 correlation between features and classification.

Conclusion The proposed model achieves better generalization performance than the classical traditional models, and 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 μmol/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
pathological pulse has become an important basis for the diagnosis of diseases. To improve the accuracy of pathological diagnosis, lately multiple artificial intelligence approaches have also been applied in the field of pulse diagnosis of TCM.

2 Related Work

In recent years, different pulse wave acquisition instruments and varied methods of pulse wave characteristic analysis have been applied to TCM. 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 et al. [10] applied AdaBoost on hypertension prediction based on time-domain pulse wave, and reached classification accuracy of 86.41%. Feng and Li [11] used the fuzzy C-means clustering to classify the characteristics of frequency-domain pulse wave. Zhang et al. [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 foregoing methods did not deal with the sequential characteristic of pulse wave. With the advancement 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 and Zhou [14] extracted single-period and multiperiod features of pulse wave based on CNN and combined with frequency-domain features for classification with an accuracy of 93%. Yan et al. [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 perspective for feature extraction of pulse wave. However, the previous stated approaches only classify the time-domain waveform without considering frequency-domain characteristics of pulse wave.

In this article, 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, it is difficult to set the filtering parameters. SPA only adjusts the frequency response by changing the smoothing parameters, which can effectively improve the filtering speed. The formula is as follows.

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

(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 $\lambda$. The baseline drift frequency is 0.2–0.3 Hz, and the sampling frequency is 200 Hz. 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 0.5 Hz 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

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

(2)

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} [p(n) - p(n)]^2}{N}}$$

(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 frequency 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 excels 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 article. The detailed process is shown in Fig. 1.

In the pre-emphasis process, a Gaussian filter was applied to framing (sampling) operation, and 256 sampling points were taken as an observation unit to ensure that an observation unit contains at least one pulse wave period. In the process of windowing, we selected a 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 leap 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 article, we added the adaptive white noise to pulse wave decomposition stage to superimpose and eliminate pulse wave components in different time scales, and also eliminated 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.

\[ p'(t) = p(t) + g'(t) \] (4)

where, \(g'(t)(i = 1, 2, \ldots, I)\) represents Gaussian noise, and \(i\) is the total number of pulse wave instances.

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

\[
\begin{align*}
  r_k(t) &= p(t) - IMF_1^i(t), k = 1 \\
  r_k(t) &= r_{k-1}(t) - IMF_k^i(t), k > 1
\end{align*}
\] (5)

\[
\begin{align*}
  IMF_{k+1}^i(t) &= \frac{1}{2} \sum_{m=1}^{I} IMF_{k+1}^i, k = 0 \\
  IMF_{k+1}^i(t) &= \frac{1}{2} \sum_{m=1}^{I} \{r_k(t) + \epsilon_k E_k[g'(t)]\}, k > 0
\end{align*}
\] (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 article, 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.

\[
\begin{align*}
  T_m(k) &= \begin{cases} 
    0, k < f(m-1) \text{ or } k > f(m+1) \\
    \frac{2(k-f(m-1))}{(f(m)+f(m-1))(f(m)-f(m-1))}, f(m-1) \leq k \leq f(m) \\
    \frac{2(k-f(m+1))}{(f(m)+f(m+1))(f(m)-f(m+1))}, f(m) \leq k \leq f(m+1)
  \end{cases}
\end{align*}
\] (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.

\[
C(n) = \sum_{m=0}^{N-1} \ln \left( \sum_{k=0}^{N-1} |H_n(k)|^2 T_n(k) \cos \left( \frac{\pi n(m - 0.5)}{M} \right) \right), 0 \leq m \leq M, n = 1, 2, \ldots, L
\]

where \(H_n(k)\) 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 article, 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 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 correlation of features 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 correlation generated after the channel expansion is often ignored. Therefore, this paper adds the channel attention mechanism on the basis of BiLSTM model to learn the correlation of pulse wave feature in the channel dimension.

However, when the input sequence is overly long and the redundancy of input data from two directions increases, the vector of features cannot accurately express correlation between the data. Meanwhile, due to the length limit, the model cannot retain all the important information. Therefore, in this article we add spatial Attention mechanism on the basis of channel Attention, selectively learn the input sequence, retain the intermediate results of the BiLSTM encoder, correlate the pulse wave sequences in the output, and form a mixed Attention mechanism to learn the correlation of pulse wave feature in the channel and spatial dimensions. The structure diagram is shown in Fig. 2.

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

\[
s = F_g(c) = \frac{1}{h} \sum_{i=1}^{h} d_c(i)
\]

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

\[
P_C = F_{pw}(s, W) = \sigma(g(s, W)) = \sigma(W_2 \delta(W_1 s))
\]

where \(W_1\) is the weight of the first fully connected layer, activation function \(\delta\) is ReLU, \(W_2\) is the weight of the second fully connected layer, and activation function \(\sigma\) is sigmoid. The original features are re-calibrated by the weight of channels, and finally the weighted channel dimension is obtained.

\[
r = F_{rw}(d_c, P_C) = P_C \cdot d_c
\]

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 correlation between Query of certain element and its Keys. The expression is as follows.

\[
p_i = \text{Similarity}(Query, Key_i) = \text{MLP}(Query, Key_i)
\]
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.

\[ q_i = \frac{\exp(p_i)}{\sum_j \exp(p_j)}, \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots, i. \]  (13)

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

\[ \text{Attention}(\text{Query}, < \text{Key}, \text{Value}>) = \sum_{i=1}^{N} q_i \cdot \text{Value}_i \]  (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 Fig. 3. The model training pseudocode is shown in Table 1.

**Table 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 | 11 return all prediction |
| 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: \(x_{AB} = \text{Attention}(x)\) | |
| 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: \(x_{AB} = \text{Attention}(x_{fc2})\) | |
| 10 FC layer: \(pre = \text{sigmoid}(W_{fc3} (x_{AB} + x_{AA}) + b_{fc3})\) | |
5 Experimental Results and Analysis

5.1 Pulse Wave Dataset

The experimental dataset in this article comes from clinical pulse wave instances of suspected H-type hypertension patients provided by the Longhua Hospital affiliated to Shanghai University of TCM and the Hospital of Integrated Traditional Chinese and Western Medicine. The inclusion criteria for hypertension in this study are in accordance with the Guidelines for Prevention and Treatment of Hypertension in China (2018 revised edition). The diagnosis of H-type hypertension is based on the level of homocysteine, with \( H_{cy} < 15 \mu mol/L \) in the non-H-type hypertension group and \( H_{cy} \geq 15 \mu mol/L \) in the H-type hypertension group. The exclusion criteria in this study include: ① Secondary hypertension; ② Other systemic serious diseases, such as severe anemia, hyperthyroidism, severe heart failure, aortic valve disease, congenital heart disease, cardiomyopathy, constrictive pericarditis, arteriovenous fistula, malignant tumor, autoimmune diseases and; ③ Stress state (including pregnancy, surgery, trauma, etc.); ④ Unable to cooperate with collection or incomplete data. The cases are collected by a pulse instrument. Before sampling, the patient should relax for more than 5 min, and the sampling time is 60 s. The sampling contact pressure depends on the patient’s pulse depth, and the sampling frequency consists of 200 Hz, 700 Hz, and 1000 Hz. In the interest of data consistency, we lowered the sampling frequency from 700 and 1000 Hz to 200 Hz. A total of 337 instances were collected, including 140 males (41.54%) with an average age of \((66.26 \pm 10.37)\) years and 197 females (58.46%) with an average age of \((71.36 \pm 8.51)\) 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.

5.2 Hyperparameters and Evaluation Indexes

Clinical pulse wave sampling is subject to multiple factors, and sometimes there will be incomplete data. After specific pretreatment, 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 article, four evaluation indexes of confusion metrics, including Accuracy, Recall, Precision and F1-score for binary classification, Receiver Operating Characteristic (ROC) curve and area under curve (AUC) and the Area below Precision-Recall curve called AP(Average Precision) are used to evaluate the classification performance of rhinitis instances. In addition, we consider SNR and RMSE as evaluation indicators of pulse wave denoising as well.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{15}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{16}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{17}
\]

\[
\text{F1score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}
\]

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

![Fig. 4 Distributions of pulse wave instances](image-url)
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 paper, 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. Table 2 shows the corresponding SNR and RMSE values of the curve in Fig. 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. 6).

Figure 7 shows the approximation and detailed components of each order in the single cycle after wavelet decomposition. Figure 7a 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. Figure 7b 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. 7b, 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 20 to 120 Hz, and the detailed components decomposed in layer 2 correspond to the high-frequency interference of 50 Hz. 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

Table 2 Comparison of denoising effect of different wavelets

| N | \( db \) | | \( Sym \) | | \( Coif \) |
|---|---|---|---|---|---|
| | 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 | – | – |

The significance of bold indicate the highest value of what to be stated.
by 0.28% compared with WT. Compared with EMD, SNR of SPA increased by 2.93% and RMSE decreased by 0.15%. It shows that SPA is more suitable for baseline drift removal of pulse wave. Figure 8 shows the filtering process, e.g. high-frequency noise and baseline drift removal. It can be seen that the filtering process reduces the average amplitude of pulse wave, eliminates the extreme points, and maintains stable periodicity of pulse wave.

### 5.4 MFCC-Based Feature Extraction

Figure 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 process the original pulse wave instances with 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 spatial Attention) and BSEAA(BiLSTM with SE-block-Attention and additional Attention). Figure 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 Original 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.

Table 3 shows the classification results of various models. For the 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.

Figure 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 evaluation performance than MFCC, Handled Pulse and Original Pulse. Therefore, Improved MFCC has better classification accuracy and generalization performance. It follows that the filtered frequency-domain MFCC features have higher discriminability than time-domain features, which can effectively reduce mode mixing and enhance classification performance.

5.5 Ablation Study

To evaluate the performance of attention-based mechanism, we conducted ablation study by adding the spatial and channel attention blocks based on BiLSTM model, respectively. Figure 12 shows the performance of different models, and Table 4 lists the evaluation indexes of various BiLSTM models with different attention mechanisms. Figure 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 increase 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 the spatial attention mechanism can effectively extract the important location features of pulse wave, 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 (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 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.

According to Table 4, the BSEA (BiLSTM with SE-block-Attention and spatial Attention) model uses channel Attention first, then adds spatial Attention, and its F1 score reaches 0.9006. BASE (BiLSTM with Attention-SE-block)
Fig. 11 The ROC curves of various methods
uses spatial Attention first, then adds channel Attention, and its F1-score is 0.8871. It can be seen from the results that BSEA outperforms BASE, and the sequence of adding Channel or Spatial Attention mechanism based BiLSTM has some effects on classification performance. Data compression via global average pooling may cause loss of some detailed features. Therefore, after adding channel attention mechanism, we also 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 the highest of 0.9348, Recall increased by 0.77%, Precision increased by 0.58%, F1 score increased by 0.71%, AUC 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

Figure 14 shows the training time and number of parameters of threefold 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 number of parameters increases 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. In the meantime, 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 20 s. With the addition of attention mechanism module, the number of parameters increases and the accuracy of classification improves remarkably.

6.2 Analysis of Feature Importance

In addition, we added 29 healthy controls (male 10.3%, female 89.7%, age 20.34 ± 0.61), and calculated the P-values and 95% confidence intervals of the characteristics of sex, age and pulse wave in the time–frequency domain, including healthy control group and H-type, H-type and non-H-type. Table 5 shows significance test of top 5 feature importance ranking in time and frequency domain including healthy controls vs hypertensive patients and H-type vs non-H-type Hypertension.

In this article, we ranked 36 pulse wave MFCC features [29] in frequency-domain by calculating Gini impurity of RF algorithm [30, 31] 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 crucial influence on classification performance. Presently, 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 correlation, and more refined classification may be conducted on pulse wave instances of H-type hypertension. In addition, 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 is yet to be explained so far. It may play a great role in classification of H-type hypertension. Our future directions will include the study of the characteristics of the first-order and second-order differential coefficients, and the correlation.

![Fig. 14 Computational complexity of improved BiLSTM models with different attention mechanisms based on MFCC](image-url)
between the coefficients of each order and the classification of pulse wave.

In this article, we also calculated the correlation between time-domain pulse graph features and classification of H-type hypertension, and worked out the feature importance ranking. As shown in Fig. 13, the top 4 importance features, i.e. $w_1/T$, $h_4/h_1$, $W_1/T$, $H_4/H_1$, $h_4$ and $t_1/t_4$ [32, 33] 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 dicrotic notch, representing the height from the bottom of dicrotic notch to the baseline of pulse wave. The height of dicrotic 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. Studies [34] have proved that H-type hypertension is an independent risk factor for atherosclerosis and atherosclerosis, which is also correlated with the results of our study, suggesting that H-type hypertension has a certain correlation with atherosclerosis and atherosclerosis.

$t_1/t_4$ 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.

Therefore, we can research on the correlation between the pulse wave features both in time-domain and frequency-domain and classification of H-type hypertension, aid clinicians to furtherly study the influence of patient’s vascular

### Table 5 Significance test of top 5 feature importance ranking in time and frequency domain

| Clinical indicators | Healthy controls vs hypertensive patients | H-type vs non-H type hypertension |
|---------------------|------------------------------------------|----------------------------------|
|                     | 95% confidence interval | P-value | 95% confidence interval | P-value |
| Age                 | 48.457–50.603 | < 0.001 | 0.359–4.777 | 0.023 |
| Height              | 0.174–5.785 | 0.038 | 1.446–5.133 | 0.001 |
| Weight              | 6.341–12.986 | < 0.001 | 0.757–6.394 | 0.013 |
| Gender              | 0.109–0.367 | 0.001 | 0.155–0.369 | < 0.001 |
| Inactivity          | 0.009–0.457 | 0.003 | 0.040–0.258 | 0.008 |
| $W_1/T$             | 0.012–0.058 | 0.004 | 0.013–0.047 | 0.001 |
| $H_4/H_1$           | 0.066–0.159 | < 0.001 | 0.009–0.523 | 0.042 |
| $H_4$               | 807.376–918.05 | < 0.001 | 0.052–0.145 | < 0.001 |
| $t_1/t_4$           | 0.098–0.148 | < 0.001 | 0.007–0.076 | 0.018 |
| $t_4/T$             | 0.130–0.182 | < 0.001 | 0.014–0.092 | 0.007 |
| $S_8$               | 0.881–1.756 | < 0.001 | 0.155–0.212 | 0.017 |
| F$_{12}$            | 0.011–0.043 | 0.006 | 0.022–0.027 | 0.005 |
| S$_{10}$            | 0.049–0.175 | 0.012 | 0.082–0.122 | 0.018 |
| S$_1$               | 1.433–1.652 | 0.006 | 0.529–0.812 | 0.036 |
| F$_8$               | 0.181–0.790 | 0.003 | 0.012–0.223 | 0.049 |

![Fig. 15 Feature importance ranking of pulse wave in frequency-domain and time-domain](image-url)
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 article, 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 was funded by the National Natural Science Foundation of China (No. 81973749) and the Science and Technology Plan of Shanghai Municipal Science and Technology Commission (No. 21DZ2271000).

Declarations

Conflict of interest None declared.

Ethical Approval Not required.

References

1. Huang, L. Q., Wu, C. X., et al. (2020). Clinical characteristics of H-type hypertension and its relationship with the MTHFR C677T polymorphism in a Zhuang population from Guangxi, China. Journal of Clinical Laboratory Analysis, 34, e23499. https://doi.org/10.1002/jcla.23499
2. Zhang, J., Liu, Y., Wang, A., et al. (2018). Association between H-type hypertension and asymptomatic extracranial artery stenosis. Science and Reports, 8, 1328. https://doi.org/10.1038/s41598-018-19740-0
3. Hu, S., Ren, L., Wang, Y., et al. (2015). Homocysteine-lowering therapy and early functional outcomes of ischemic patients with H-type hypertension: a retrospective analysis of CNSR. Australasian Physical Engineering Science & Medicine, 38, 785–791. https://doi.org/10.1007/s13246-015-0406-x
4. Zhou, C., Guo, Y., & Deng, J. (2019). H-type hypertension and atrial fibrillation. Advances in Cardiovascular Diseases, 40(9), 1205–1207.
5. Wang, Y., Li, F., Yan, H., et al. (2007). A review of digitized traditional diagnostic techniques. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 9(3), 96–101.
6. Wei, H., Liu, M., & Zheng, H. (2004). The modern research route of pulse-taking. Journal of Basic Chinese Medicine, 10(2), 69–71.
7. Zhang, X., Xu, L., Chen, K., et al. (2009). A new method for locating feature points in pulse wave using wavelet transform. WRI World Congress on Computer Science and Information Engineering, 5, 367–371.
8. Lu, H., Ye, W., & Hu, Y. (2019). Research on feature extraction of pulse wave signal. Information and Communications, 9(7), 3–5. (In Chinese).
9. Fan, B., Wang, Y., Huang, C., et al. (2020). Pulse wave signal feature recognition based on time-domain differential period ratio. Acta Photonica Sinica, 49(12), 1206003–1206003. (In Chinese).
10. Luo, Z., Hu, J. X., et al. (2018). A study of machine-learning classifiers for hypertension based on radial pulse wave. BioMed Research International, 7, 1–12.
11. Feng, B., & Li, S. (2018). Unsupervised clustering analysis of human-pulse signal in traditional Chinese medicine. CAAI Transactions on Intelligent Systems, 13(4), 564–570. (In Chinese).
12. Shiyu, Z., Ke, Y., Chunming, X., et al. (2020). Research on dimension reduction and classification of pulse signal features based on random forest. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 22(07), 2418–2426. (In Chinese).
13. Zhang, N. (2018). Pulse condition recognition based on convolutional neural network with dimension enlarging. Computer Science, 6(45), 506–507. (In Chinese).
14. Liu, G., & Zhou, W. (2020). Pulse wave signal classification algorithm based on time-frequency domain feature aliasing using convolutional neural network. Journal of Iljin University (Engineering and Technology Edition), 50(05), 1818–1825. (In Chinese).
15. Yan, J., Chen, S., Yan, H., et al. (2020). Wrist pulse analysis and recognition based on recurrence plot and convolution neural network. Computer Engineering and Applications, 56(7), 170–175. (In Chinese).
16. Kang, G., Xiao, Y., Liu, J., et al. (2021). Tatt-BiLSTM: Web service classification with topical attention-based BiLSTM. Concurrency and Computation. https://doi.org/10.1002/cpe.6287
17. Huang, W., & Yu, J. (2019). Exploration and analysis research of medical consultation content and objectification in traditional Chinese medicine. China Journal of Traditional Chinese Medicine and Pharmacy, 34(8), 3666–3668.
18. Ji, Z., & Liu, X. (2016). Study on feature points recognition of pulse wave based on waveform feature and wavelet. Chinese Journal of Scientific Instrument, 37(2), 379–386.
19. Thomas, T. L., & Antony, C. S. (2000). Detection of transients in noise with the undecimated discrete wavelet transform. IEEE Transactions on Signal Processing, 48(5), 1458–1462.
20. Paul, S. B. S., Glittas, A. X., & Gopalakrishnan, L. (2021). A low latency modular-level deeply integrated MFCC feature extraction architecture for speech recognition. Integration-The VLSI Journal, 76, 69–75.
21. Deng, M., Meng, T., Cao, J., et al. (2020). Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Networks., 130, 22–32.
22. Jin, S., Wang, X., Lu, L., et al. (2021). Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN. Applied Acoustics, 172, 1–11.
23. Li, Y., & Dong, H. (2018). Text sentiment analysis based on future fusion of convolution neural network and bidirectional long short-term memory network. Journal of Computer Applications, 38(11), 3075–3080.
24. Pham, T. D. (2021). Time–frequency time–space LSTM for robust classification of physiological signals. Science and Reports, 11, 6936. https://doi.org/10.1038/s41598-021-86432-7
25. Greff, K., Srivastava, R. K., Koutnik, J., et al. (2015). LSTM: A search space odyssey. IEEE Transactions on Neural Network and Learning Systems, 28(10), 2222–2232.

26. Hu, J., Shen, L., Sun, G., et al. Squeeze-and-excitation networks. In IEEE Conference on Computer Vision and Pattern Recognition. hu2018senet

27. Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Applied Soft Computing, 91, 106198.

28. Fernando, T., Denman, S., Sridharan, S., et al. (2018). Soft-Hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection. Neural Networks, 108, 466–478.

29. Abdouni, A., Vargiolu, R., & Zahouani, H. (2018). Impact of finger biophysical properties on touch gestures and tactile perception: Aging and gender effects. Scientific Reports, 8, 12605.

30. Aler, R., Valls, J. M., & Boström, H. (2020). Study of Hellinger Distance as a splitting metric for Random Forests in balanced and imbalanced classification datasets. Expert Systems with Applications, 149, 113264.

31. Seifert, S., Gundlach, S., & Szymczak, S. (2019). Surrogate minimal depth as an importance measure for variables in random forests. Bioinformatics, 35(19), 3663–3671.

32. Zhao, Y., Cheng, F., Pham, M., et al. (2019). Relationships between homocysteine and arterial stiffness in patients with h-type arterial hypertension. Journal of Hypertension, 39, e314–e315.