ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.165

Analysis and Identification of Pulse Signals for Patients with Duodenal Bulbar Ulcer Based on Wavelet Packet Transform and Sample Entropy

Lin Fan [1], Xiaokang Zhang [2], Zhongmin Wang [1], Rong Zhang [1], Yan He [1]

[1] School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing

Xi 'an Key Laboratory of Big Data and Intelligent Computing, Xi'an, China

[2] School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, China

Abstract. Objective. The aim of this study is to classify and identify the pulse signals of healthy individuals and patients with duodenal bulb ulcer (DBU) using machine learning algorithm to provide objective parameters for the clinical application of the Traditional Chinese Medicine (TCM) pulse diagnosis. Methods. In order to effectively identify the chronic disease DBU by pulse signals, the fusion of energy features with time-frequency characteristics and sample entropy feature are used to address the shortcomings of traditional pulse feature extraction methods, and the fused feature vectors are used as the input data of the K-nearest neighbor (KNN) classifier for model training and classification identification. Results. There are significant differences in the energy feature and sample entropy feature between pulse signals of normal subjects and patients, respectively. After the two features are fused and fed into the classifier, the results show that the differences are more obvious, which verifies the effectiveness of the proposed feature fusion method. Compared with the traditional methods, the feature fusion method proposed in this paper can effectively represent more accurate pulse information with the accuracy of 89.3\%. Conclusion. Therefore, it can be determined that the linear and nonlinear fused features used in this study can be used as a favorable disease predictor based on pulse signals, and the method also provides a new perspective for pulse diagnosis in TCM.

Keywords: pulse signal, energy feature, sample entropy

1. Introduction

The wrist pulse contains information about the physiology and pathology of the human body, so physicians can use pulse information to determine the health status of the human body. Traditional pulse diagnosis has important clinical value in the diagnosis and prevention of diseases, especially in cardiovascular and visceral diseases. However, traditional Chinese medicine pulse diagnosis relies on the physician's experience, subjective judgment and description, and accurate pulse diagnosis can only be performed by experienced practitioners, which has greatly limited its clinical application and development. Therefore, there is a need to promote objective and quantitative pulse diagnosis using computerized signal analysis techniques, which can help establish objectivity and standardization in the diagnosis of TCM diseases[1] [2].

Currently, many researchers at home and abroad have proposed different methods for pulse analysis. Gong et al. [3] designed a prototype system for wrist pulse detection and analysis, including a pulse detection device and a cirrhosis diagnosis scheme based on the pulse signals, which achieved an accuracy of 87.09% for the identification of cirrhosis. Wang et al. [4] [5] integrated a pressure sensor with optical sensors together, fused different pulse acquisition methods, designed a multi-channel sensor array structure to obtain more pulse information, and proposed a robust signal preprocessing framework for wrist pulse analysis, which is superior to previous pulse signal preprocessing methods, applied period segmentation and pulse signal normalization to feature extraction, and obtained a classification accuracy of 91.6% for diabetes diagnosis. Guo et al. [6] treated TCM pulse signals as time series and further used Hilbert-Huang transform to extract energy features and sample entropy features of pulse signals and achieved an average recognition

accuracy of 90.21%. Liao et al. [7] calculated 10 relative energy values of spectral harmonics by pulse signal spectrum analysis to determine the three stages of pregnancy pulse signal. Huang et al. [8] analyzed the characteristics of frequency domain of pulse signal and determined the proportion of energy distribution in different harmonic bands using the spectral harmonic energy ratio (SHER), and verify that the total energy of the fourth to sixth harmonics was significantly lower in patients with palpitations than in normal subjects using this method. Luo et al. [9] used questionnaire information and pulse signal's time domain features as input variables to construct an unused machine learning model to classify and predict the pulse signals of the hypertensive and healthy groups with an accuracy of 86.41%. In addition, many experiments have shown that spectral analysis can reveal more physiological and pathological information in pulse diagnosis analysis compared to time domain analysis [10-12]. Through the above methods, a large number of quantifiable feature parameters can be obtained, and then the pulse signals can be classified and identified by machine learning algorithms such as Bayesian classification, support vector machines, and artificial neural networks [13].

Although the above methods have achieved breakthrough results, most of them remain linear analysis methods in time-frequency domain, which have great limitations. The nonlinear dynamics time series methods can obtain nonlinear information about the pulse signal that cannot be obtained by conventional analysis methods and have wide applications in the study of physiological signals [14] [15]. A pulse diagnosis method with the fusion of linear and nonlinear features is proposed, using wavelet packet analysis and sample entropy algorithm to classify and identify normal and patients with DBU and normal subjects in this study.

2. Methods

In this paper, a pulse signal disease recognition model based on wavelet packet and sample entropy feature fusion is proposed. It mainly includes the extraction and fusion of linear and nonlinear features, and machine learning algorithm are used for classification and identification. The framework of the model is shown in Fig. 1. The pulse signal is first decomposed and reconstructed by wavelet packet, then the sample entropy of the pulse signal is extracted, and the eigenvalues representing the energy weight of the sub-band signal in the reconstructed nodes and the sample entropy values are fused to construct a feature vector set. Finally, the KNN algorithm is used as a classifier for classification and recognition.

The energy features extracted by the wavelet packet transform combine the time-frequency characteristics of the signal, while compensating for the lack of high-frequency signal characteristics of the commonly used wavelet transform. Since the higher energy of the frequency band signal after wavelet packet decomposition, the more complex it physiological information represents. Therefore, the pulse signal is first decomposed by wavelet packet, and then the sample entropy of the signal after decomposing the reconstructed nodes is calculated separately as follows: (1) Select the *dmey* as the wavelet basis function, performing 3-layer wavelet packet decomposition on the pulse signal, and extracting the wavelet packet energy distribution. (2) Calculate the energy magnitude of the 8 node signals after wavelet packet decomposition, and normalize the calculated energy values to form the energy characteristic parameters in a certain order. (3) The sample entropy of each cycle of the pulse signal is extracted by the sample entropy algorithm, and fused with the feature from wavelet packet transform to form the feature vector for disease identification.

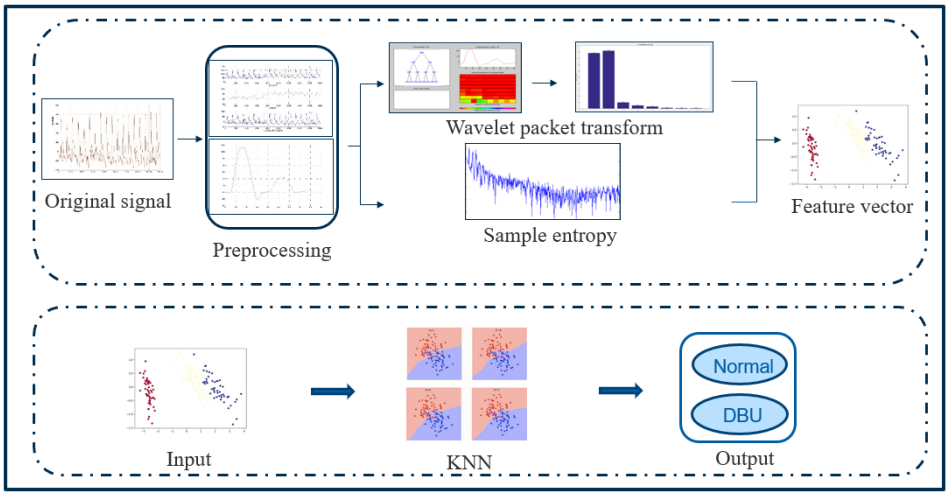

Fig. 1: Framework of pulse signal disease recognition system based on wavelet packet and sample entropy feature fusion. The pulse data are processed sequentially as follows: data preprocessing, extraction of linear and nonlinear features, feature fusion, and KNN classifier for classification and recognition.

2.1. Wavelet packet decomposition

Wavelet Packet Decomposition, is a further optimization of the wavelet transform. The main idea of the algorithm is that, at each level of signal decomposition, except to the low frequency sub-bands, the high frequency sub-bands are also further decomposed based on the wavelet transform. Finally, the optimal signal decomposition path is calculated by minimizing a cost function, and the original signal is decomposed by this decomposition path. Considering the fact that wrist pulse signals are naturally a time series, the wavelet packet can be used to analyze the time series and then extract the disease-sensitive features. The algorithm for for wavelet packet decomposition is shown below:

$$\begin{cases} \mathrm{d}_k^{\mathrm{j+1},2n} = \sum_l h_{0(k-2l)} d_l^{j,n} \\ \mathrm{d}_k^{\mathrm{j+1},2n+1} = \sum_l h_{1(k-2l)} d_l^{j,n} \end{cases} \tag{1}$$

In (1): j is the signal decomposition scale; k is the signal decomposition position; n is the signal frequency; l is the variable of the signal; and are the filter coefficients used for wavelet packet decomposition. The wavelet packet reconstruction algorithm is shown below:

$$d_l^{j,n} = 2[\sum_k h_{0(l-2k)} d_{\mathrm{k}}^{j+1,2n} + \sum_k h_{1(l-2k)} d_{\mathrm{k}}^{j+1,2n+1}] \tag{2}$$

In (2): $j = 1, 2, \cdots\cdots$; $l = 0, 1, \cdots\cdots, N - 1$; $n = 0,1, \cdots\cdots, 2^{j-1}$. Suppose $x(t)$ is the pulse signal, and the 3-layer wavelet packet decomposition is performed on the pulse signal, then the reconstructed signal is defined as follows:

$$S = S_{30} + S_{31} + S_{32} + S_{33} + \ldots\ldots + S_{3j-1} + S_{3j}$$

in the formula: $S_{3j}(j = 0,1,\ldots\ldots,7)$ denotes the (3, j)th node coefficient reconstruction signal in the 3-layer wavelet packet decomposition process. Taking the 3-layer wavelet packet transform as an example, first find the energy of each frequency band, i.e. $S_{3j} = E_{3j}$, then:

$$E_{3j} \int \left| s_{3j}(t) \right|^2 dt = \sum_{k=1}^{n} \left| x_{jk} \right|^2 \tag{3}$$

In (3): $x_{jk}(j = 0,1,\ldots\ldots,7; k = 1,2,\ldots\ldots,n)$ indicates the amplitude of the reconstructed pulse signal S_{3j} .And the total signal energy is $E = (\sum_{j=0}^{7} \left| E_{3j} \right|^2)^{1/2}$, Using energy as an element, the vector T can be constructed as follows:

$$T = \left[E_{30}, E_{31}, \ldots\ldots, E_{36}, E_{37} \right] \tag{4}$$

In (4): T' is the feature vector obtained by normalization that can be used as a feature parameter of the pulse signal.

A. Sample entropy algorithm

Assume that the original data sequence is: $x(1), x(2), \cdots\cdots, x(N)$,there are a total of N sampling points, and the specific procedure for calculating the sample entropy of a segment of the signal is as follows:
(1) Given the number of fault signal dimensions m， Composing an m-dimensional vector from the data of the original signal as follows:

$$x(i) = \left[x(i), x(i+1), ..., x(i+m+1)\right], \quad i = 1, \ 2, \ ....., \ N-m+1 \tag{5}$$

(2) Calculate the distance between $x(i)$ and $x(j)$ as follows:

$$d(i,j) = \max_{k=1\to m-1} \left|x(i+k) - x(j+k)\right|, \quad k = 0, \ 1, \ ....., \ m-1 \tag{6}$$

(3) Set Threshold r， for each value of i count the number of $d(i,j) < r$, the ratio of these numbers to the total number of distances $N-m+1$ is denoted as $B_i^m(r)$, $B_i^m(r) = \frac{\left[d(i,j) < r\right]}{N-m+1}$,$1 \le j \le N-m$, $j \ne i$. Referring to the following equation (7), the average value of all i can be solved using the following equation:

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \tag{7}$$

(4) And then given the dimension $m+1$, repeat steps (1)-(3), it can get $B^{m-1}(r)$.
(5) If N is a finite data value, the sample entropy of the pulse signal is calculated as follows:

$$SampEn(m,r,N) = -\ln \frac{B^{m+1}(r)}{B^m(r)} \tag{8}$$

In this paper, the sample entropy of the pulse signal is calculated by following parameters: $N = 1024$, $m = 2$， $r = 0.2$。

2.2. K nearest neighbor classifier

3. Results

The wrist pulse signal dataset used in this study was obtained from a database established by Chen [16] in collaboration with Harbin 211 Hospital (Harbin, Heilongjiang Province, China), who collected pulse signals from subjects of different ages and diseases, including 100 healthy individuals and 42 cases of patients with DBU.

3.1. Data Processing and Analysis

Since the pulse signal is a weak physiological signal, it has a high background noise and is susceptible to external interference (poor sensor acquisition point contact, industrial frequency interference, environmental noise, its own biologic signals, and other factors). Meanwhile, the pulse signal is subject to baseline drift due to interference from the respiration and body movement of the person being acquired. The preprocessing including noise reduction of the original signal and elimination of baseline drift, providing a better basis for the subsequent use of the data. It starts with a filter to remove high frequency noise from the signal, followed by a smoothing process, and then the obtained signal is divided into single cycles. The flow chart of the pulse signal preprocessing is given in Fig. 2.

The baseline drift of the pulse signal needs to be eliminated. The original pulse signal contains baseline interference signals (low frequency noise) that can adversely affect the signal analysis. Using the zero-phase low-pass filtering method, the pulse signal is first low-pass filtered, then the trend signal is subtracted from the original signal to obtain the detrended pulse signal, as shown in Fig. 3. In addition, a complete sampled

pulse signal will also be segmented into single-cycle waveforms for further analysis. A simple peak detection algorithm is used to first find the peak of the pulse signal that crosses the threshold by manually selecting the amplitude threshold.

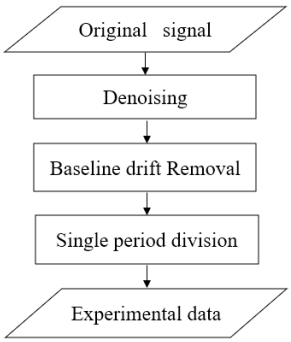

Fig. 2: Flow chart of pulse signal preprocessing.

If the difference between the adjacent local maxima and local minima is greater than the threshold, then mark this local minimum as the start of the cycle and this local maximum as the main wave amplitude of the corresponding cycle. The signal from the start of the current pulse to the start of the next pulse is taken as a cycle. It is concluded from the experiment that if the threshold value is set too small, then some non-starting points will be marked as the start of the period. If the threshold value is set too large, then some start points that need to be identified will be missed. By the above method, the pulse signal after period division and superposition is obtained, as shown in Fig. 4.

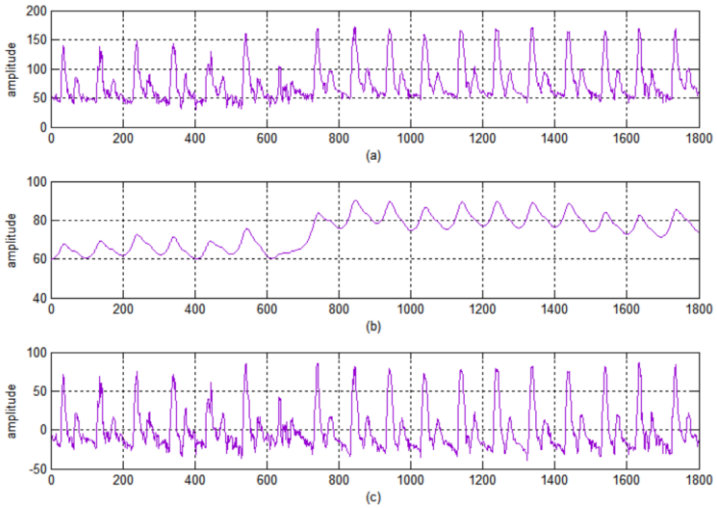

Fig. 3 Removal of baseline drift of pulse signal. (a) Original signal. (b) Trend line. (c) The signal of removing baseline drift.

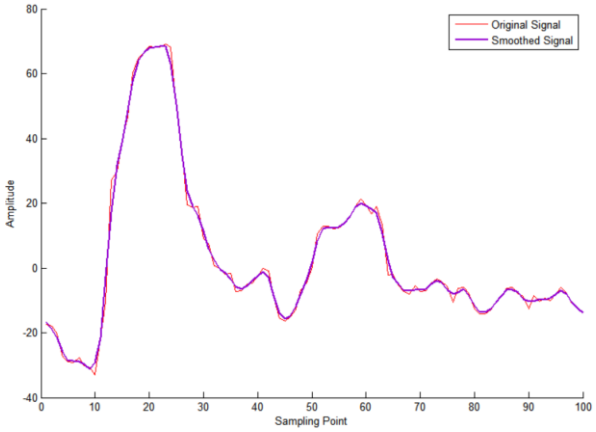

Fig. 4: A single period of pulse signal.

3.2. Feature Extraction

Wavelet packet transform and sample entropy algorithm are used for feature extraction of the pulse signal. The *dmey* is chosen as the wavelet basis, and the preprocessed signal is decomposed into 3 layers of wavelet packets. The 8 sub-bands of the 3rd layer obtained from the decomposition are reconstructed to form a feature vector of dimension 8. The two main parameters of the sample entropy are the embedding dimension m and the threshold r. In this paper, we take the parameters as follows: m = 2 and r = 0.2. Fig. 5(a) shows the reconstructed signal of each node of the normal subjects pulse signal after the 3-layer wavelet packet decomposition. Fig. 5(b) shows the reconstructed signal of each node of the pulse signal of the patient with DBU after 3-layer wavelet packet decomposition.

After the 3-layer wavelet packet decomposition and the normalization of the frequency band energy, the pulse signal is transformed from a one-dimensional signal to 8 reconstructed signals, and it can be seen that the pulse signal energy of DBU patients is reflected in the high frequency part. As can be seen from TABLE I, after reconstructing the eight sub-bands of the pulse signal obtained from the 3-layer wavelet packet decomposition, the percentage of high-frequency components of the reconstructed signal of DBU patients is significantly higher compared to normal subjects, reflecting the irregularity of the pulse signal of DBU patients.

To further verify the above information, the sample entropy of each cycle of the pulse signal was extracted in this study. Sample entropy is a new measure of time series complexity by measuring the magnitude of the probability of generating new patterns in the signal. As shown in Fig. 6, the sample entropy of the pulse signal of DBU patients is higher than that of normal samples overall. The values of sample entropy are then fused with the energy features extracted from the wavelet packet transform to obtain a new feature vector.

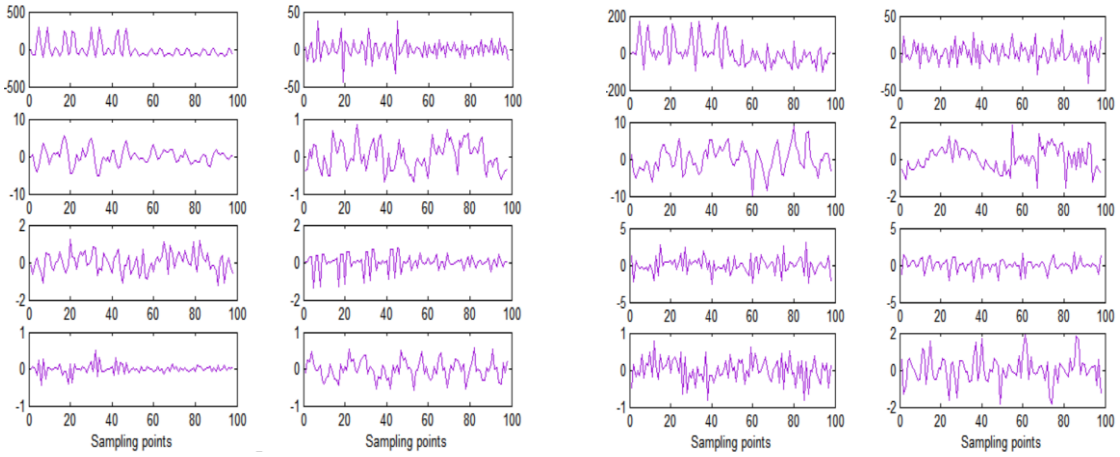

Fig. 5: The reconstruction signal of 3-layer wavelet packet decomposition of the pulse signal. Left: The reconstructed signal of normal subject. Right:The reconstructed signal of patient with DBU.

3.3. Pulse signal recognition based on KNN classifier

In this study, KNN will be used as a classifier to train the model and implement recognition classification for the band energy features and sample entropy features of the two sets of pulses in the sample. In order to select the best k value for the KNN model, multiple k values are provided to be applied to the KNN algorithm and the k value that results in the optimal classification is selected. The classification accuracy decreases initially and reaches a maximum of 89.3% when the test size is set to 40%, k=2 and the two features of band energy and sample entropy are fused, The recognition results are shown in Fig. 7.

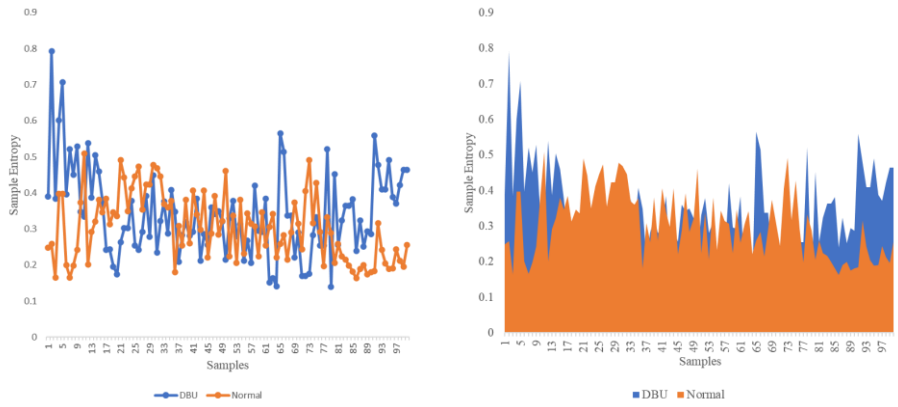

Fig. 6: The sample entropy plot of the samples. (a) Broken line plot of sample entropy. (b) Histogram of sample entropy.

Table 1: The proportion of energy in each frequency band

Frequency Band	Normal subjects		Patients with DBU	
	Energy	Proportion (%)	Energy	Proportion (%)
0-8HZ	1.0975e+06	98.8632	4.4580e+05	96.2382
8-16HZ	1.4380e+04	1.0876	1.5995e+04	3.4529
16-24HZ	435.4250	0.0341	1.1399e+03	0.2461
24-32HZ	13.2852	0.0027	39.5280	0.0085
32-40HZ	26.8891	0.0080	121.3918	0.0262
40-48HZ	19.2825	0.0022	66.7318	0.0144
48-56HZ	1.7123	2.147e-04	9.6199	0.0021
56-64HZ	6.5185	0.0021	53.8614	0.0116
Total	1.1124e+06		4.6323e+05	

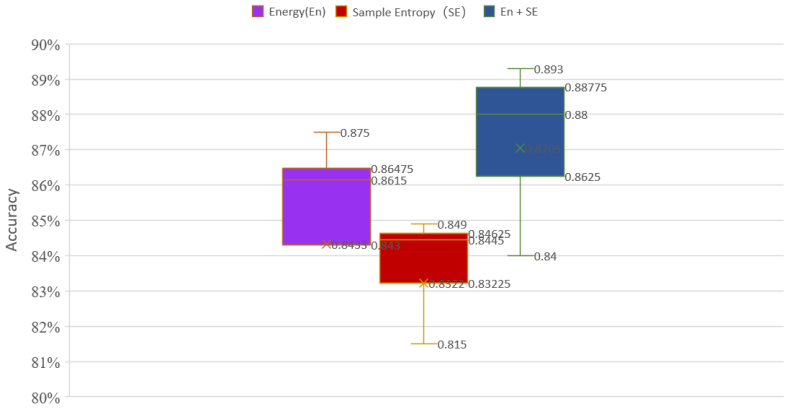

Fig. 7: Classification accuracy with different feature combinations.

4. ClassificatioDiscussion

Wavelet packet transform has high time-frequency resolution and local orthogonal adaptive characteristics, while the method is also superior to wavelet transform and other signal analysis methods, and it retains the high frequency information in the pulse signal within the range that can be studied. The sample entropy has two advantages in the field of physiological signal processing: (1) the sample entropy does not contain a comparison of its own data segments and is not sensitive to missing data, and the technique does not have to depend on the length of the data by; (2) the sample entropy has better consistency and shorter computation time than the approximate entropy. Therefore, the sample entropy algorithm is chosen in this paper to extract the features of the pulse signal. According to the sample entropy principle, the greater the probability of new pattern generation and the greater the complexity of the sequence, the greater the sample entropy value of the time series. As shown in Fig. 6, the values of sample entropy of the pulse signal of DBU patients, in general, are significantly higher than that of normal subjects, indicating that the complexity of the pulse signal of patients is higher than that of normal subjects.

In this study, the KNN classifier is used to classify and identify the energy characteristics and sample entropy of the impulse signal wavelet packet reconstruction signal. The values of energy, the values of sample entropy, and the feature vectors corresponding to the fused features of both, were used as the input data of the classifier, and the corresponding average recognition rates were 87.9%, 84.9%, and 89.3%, respectively. Compared with using energy or sample entropy alone as feature vectors, fusing energy and sample entropy as feature vectors improved the recognition accuracy of the patients with DBU group and the normal group.

5. Conclusions

In this paper, a new method of classifying duodenal bulb ulcers based on pulse signals is proposed. The energy and sample entropy of different frequency bands of the pulse signal are extracted as pulse signal

features and K-nearest neighbors are used as classifiers. The results show that the proposed pulse signal processing and classification method is effective. Despite the lack of sufficient proof of histological lesions or other pathological indicators, this prediction model can help physicians to initially identify disease categories with fewer features and give clinical treatment recommendations. Although the method proposed in this paper has achieved encouraging results, it still needs to be improved due to the limited sample size and types of diseases.

6. Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 61373116, Grant 62002287, and the general project of Shaanxi Province in the field of industry (2018GY-013), and the project of Xianyang Science and Technology Bureau (2017k01-25-2).

7. References

[1] O'Brien, Kylie, et al., "Traditional East Asian Medical Pulse Diagnosis: A Preliminary Physiologic Investigation," Journal of Alternative and Complementary Medicine, vol. 19, no. 10, pp. 793 – 798, 2013.

[2] Velik, Rosemarie, "An Objective Review of the Technological Developments for Radial Pulse Diagnosis in Traditional Chinese Medicine," European Journal of Integrative Medicine, vol. 7, no. 4, pp. 321 – 331, 2015.

[3] Gong, Shujie, et al., "Accurate Cirrhosis Identification with Wrist-Pulse Data for Mobile Healthcare," Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare, p. 6, 2012.

[4] Wang, Dimin, et al., "A Novel Multichannel Wrist Pulse System With Different Sensor Arrays," IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 7, pp. 2020 – 2034, 2015.

[5] Wang, Dimin, et al., "A Robust Signal Preprocessing Framework for Wrist Pulse Analysis," Biomedical Signal Processing and Control, vol. 23, pp. 62 – 75, 2016.

[6] Guo, Rui, et al., "Analysis and Recognition of Traditional Chinese Medicine Pulse Based on the Hilbert-Huang Transform and Random Forest in Patients with Coronary Heart Disease," Evidence-BasedComplementary and Alternative Medicine, vol. 2015, no. 2015, pp. 895749 – 895749, 2015.

[7] Liao, Yin-Tzu, et al., "The Pulse Spectrum Analysis at Three Stages of Pregnancy," Journal of Alternative and Complementary Medicine, vol. 18, no. 4, pp. 382 – 386, 2012.

[8] Huang, Chin-Ming, et al., "Developing the Effective Method of Spectral Harmonic Energy Ratio to Analyze the Arterial Pulse Spectrum," Evidence-Based Complementary and Alternative Medicine, vol. 2011, 2011, pp. 342462 – 342462, 2011.

[9] Luo, Zhi-Yu, et al., "A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave," BioMed Research International, vol. 2018, p. 2964816, 2018.

[10] O' Rourke, M. F, "Time Domain Analysis of the Arterial Pulse in Clinical Medicine," Artery Research, vol. 5, no. 4, 2011.

[11] Chen, Fun Jou, et al., "Pulse Analysis in Bipolar Disordered and Nonpsychotic Human Subjects," The American Journal of Chinese Medicine, vol. 40, no. 3, pp. 455 – 465, 2012.

[12] Nanyue, Wang, et al., "Pulse Diagnosis Signals Analysis of Fatty Liver Disease and Cirrhosis Patients by Using Machine Learning," The Scientific World Journal, vol. 2015, 2015.

[13] H.-Y. Wang and S. Xu, "Automatic pulse recognition method based on bayesian classifier," Chinese Journal of Biomedical Engineering, vol. 28, no. 5, pp. 735 – 742, 2009.

[14] Hong-wei, LU, "A Fine Graining Complexity Measure Algorithm and Its Application in Cardiac RR Intervals Analysis," Space Medicine \& Medical Engineering, 2004.

[15] Songtao, Tong, "Nonlinear Dynamic Fast Analysis and Analysis of Epileptic's EEG," Journal of Biomedical Engineering Research, 2008.

[16] L.Chen and L. J. Yang, "Identifying pulse signals based on fractal theories and wavelet," Aeronautical Computing Technique, vol. 38, no. 4, pp. 321 – 328, 2008.