Wrist pulse signal diagnosis using modified Gaussian models and Fuzzy C-Means classification

Yinghui Chen, Lei Zhang[*,1], David Zhang, Dongyu Zhang

Biometrics Research Center, Dept. of Computing, The Hong Kong Polytechnic University, Hung Kom, Kowloon, Hong Kong, China

ARTICLE INFO

ABSTRACT

Wrist pulse signal contains important information about the health status of a person and pulse signal diagnosis has been employed in oriental medicine for thousands of years. In this research, a systematic approach is proposed to analyze the computerized wrist pulse signals, with the focus placed on the feature extraction and pattern classification. The wrist pulse signals are first collected and pre-processed. Considering that a typical pulse signal is composed of periodically systolic and diastolic waves, a modified Gaussian model is adopted to fit the pulse signal and the modeling parameters are then taken as features. Consequently, a feature selection scheme is proposed to eliminate the tightly correlated features and select the disease-sensitive ones. Finally, the selected features are fed to a Fuzzy C-Means (FCM) classifier for pattern classification. The proposed approach is tested on a dataset which includes pulse signals from 100 healthy persons and 88 patients. The results demonstrate the effectiveness of the proposed approach in computerized wrist pulse diagnosis.

1. Introduction

Pulse diagnosis has been successfully used for thousands of years in oriental medicine [1–6]. In traditional pulse diagnosis, practitioners use fingertips to feel the pulse beating at the measuring position of the radial artery. Since the wrist pulse signals contain vital information and can reflect the pathological changes of a person's body condition, the practitioners can then determine the person's health conditions. However, the accuracy of pulse diagnosis depends heavily on the practitioner's skills and experience. Different practitioners may not give identical results for the same patient [2,3]. Therefore, it is necessary to develop computerized pulse signal analysis techniques to standardize and objectify the pulse diagnosis method.

The computerized pulse signal analysis has shown promises to the modernization of traditional pulse diagnosis, such as the pulse pattern reorganization, the arterial wave analysis and so on [7,8]. Generally speaking, computerized pulse signal diagnosis can be divided into three stages: data collection, feature extraction and pattern classification. In the first stage of our work, the pulse signals are collected using a Doppler ultrasound device and some pre-processing of the collected pulse signals has been performed. At the second stage, some diagnostic features that can reflect the characteristics of the measured pulse signals are extracted. These features can be time-domain features (like Doppler parameters [9,10]), frequency-domain features extracted by the Fourier transform [7] or time-frequency features extracted by the wavelet transform [11–15]. By taking the extracted features as inputs, pattern classification can be carried out at the third stage to classify the signals into different groups, i.e. the healthy subjects or patients with particular types of diseases. The pattern classification methods adopted can be statistical methods, such as Support Vector Machine (SVM) classifier [13,16] and Bayesian classifier [5], or artificial neural network (ANN) methods [17]. Although some of the existing pulse signal diagnosis approaches have shown good results, the effectiveness of these methods needs further assessment due to the limited number of testing samples and the types of diseases.

This paper aims to establish a systematic approach to the computerized pulse signal diagnosis, with the focus placed on feature extraction and pattern classification. The collected wrist pulse signals are first denoised by the wavelet transform. To effectively and reliably extract the features of the pulse signals, a two-term Gaussian model is then adopted to fit the pulse signals. The reason of using this model is because each period of a typical pulse signal is composed of a systolic wave and a diastolic wave, both of which are bell-shaped. The obtained Gaussian models can provide reliable and distinctive features of the wrist pulse signal, such as the relative differences of the two waves with respect to amplitude, phase and shape. Instead of directly using these features for pattern classification, a two-step feature selection scheme is performed. Firstly, the tightly correlated features are eliminated so that the pattern dimension is reduced to ensure the efficiency of

---

* Corresponding author. Tel.: +852 27667355.
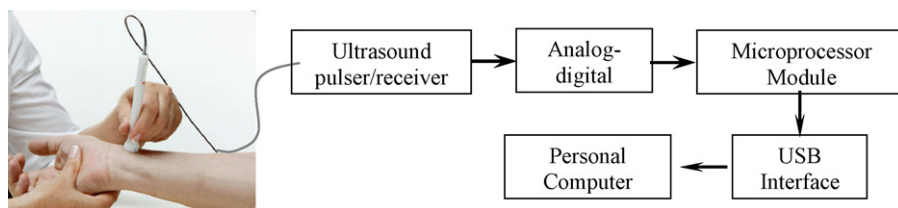*E-mail address:* cslzhang@comp.polyu.edu.hk (L. Zhang).

Fig. 1. Pulse signal collection using ultrasonic blood analyzer.

computation. Secondly, the disease-sensitive features, which can best describe the symptoms and sings of disease, are selected by using the training datasets to improve the classification performance. These selected features are taken as the inputs to the Fuzzy C-Means (FCM) classifier for pattern classification. In this paper, a pulse signal dataset, which contains pulse signals from 100 healthy persons and 88 patients, was established to validate the effectiveness of the proposed approach.

The reminder of this paper is organized as follows. Section 2 describes the wrist pulse signal collection and pre-processing. A modified Gaussian model is proposed in Section 3 to model pulse signals and extract features. The feature selection scheme is also presented in this section. Section 4 presents the FCM clustering classification method to classify the pulse signals. Section 5 performs extensive experiments to validate the proposed method. Finally, the paper is concluded in Section 6.

2. Wrist pulse signal collection and pre-processing

In our work, a USB-based Doppler ultrasonic blood analyzer (Edan Instruments, Inc.) is used to collect the wrist pulse signals (see Fig. 1). Through an USB interface, the collected signals are transmitted and stored in a PC for further processing and analysis. The signal collection process includes three steps. First is to find a rough location in the wrist. In traditional pulse diagnosis, the practitioners usually use three fingertips (Index, Middle and Ring fingers) to feel the pulse fluctuation on three positions, named 'Cun', 'Guan' and 'Chi', in a patient's wrist [1,3]. Since there is only one probe of the Doppler ultrasound device, we can only detect the pulse fluctuation at one position. Hence the pulse signal at the 'Guan' position is detected because the fluctuation of pulse at this position is bigger than other positions. The second step is to get the most significant signal by moving the probe around the rough location whilst changing the angle of the probe against the skin; and finally, the wrist pulse signal can be recorded and saved in the form of Doppler spectrograms.

These three steps are repeated several times to collect several measurements of a subject so that the measurement error can be reduced. Compared with detecting pulse signal by using the pressure sensor, which is heavily interfered by the artery blood flowing in the wrist, capturing pulse signal through ultrasound scanning is more accurate by locating the probe directly on the styloid processes. In addition, ultrasound scanning can provide new information, which is not available by using the pressure sensor, because it can reflect the deep radial artery changes beneath the skin [18,19].

Fig. 2(a) shows the collected Doppler ultrasonic spectrogram of a typical wrist pulse signal. Before extracting features, the collected wrist pulse samples are pre-processed. First, the maximum velocity envelope of each spectrogram is extracted in order to reduce dimension of the signal (see Fig. 2(b)). Afterward, the low-frequency drift and high-frequency noise contained in the maximum velocity envelopes should be removed without the phase shift distortion. In this paper, both the low-frequency drift and the high-frequency noise can be reduced simply by using a

7-level 'db6' wavelet transform [20]. By subtracting the 7th level wavelet approximation coefficients, the low-frequency drift of the waveform is eliminated. Similarly, the high-frequency noise can be removed by subtracting the 1st level wavelet detail coefficients from the waveform. The result of drift and noise removal is shown in Fig. 2(c).

It can be seen from Fig. 2 that the wrist pulse signal is not a random process but a cyclic wave with regularly occurring systolic and diastolic waves, which is confirmed in [21]. In this study, a sampled pulse signal is segmented into single-period waveforms for further analysis. The procedures to extract each period are described as follows (illustrated in Fig. 3):

1. Perform the Fourier transform to find out the base frequency, denoted as f, of the signal. Then the base period T is calculated as $T = 1/f$.
2. Detect the peak point of the pulse signal within the time interval $[0,T]$. The obtained peak point, denoted as P_1, is the maximum point of the first period, and its corresponding time instant is t_1.
3. After t_1 is determined, we can find out the second peak point P_2 within the time interval $[t_1 + (T/2), t_1 + (3T/2)]$. Its corresponding time instant is denoted as t_2. The time interval between the two peak points P_1 and P_2 is calculated as $T' = t_2 - t_1$.
4. Similarly, the next peak point P_3 can be detected within the time interval $[t_2 + (T'/2), t_2 + (3T'/2)]$, and its corresponding time
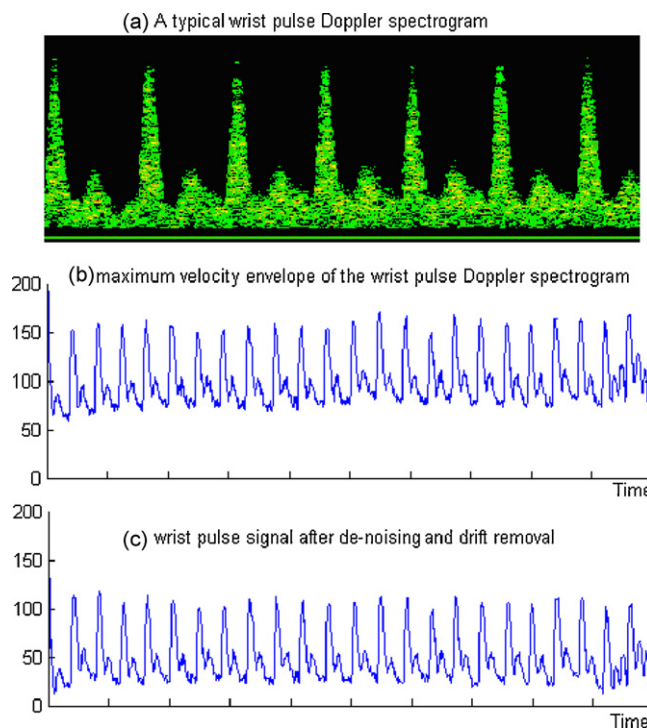

Fig. 2. (a) A typical wrist pulse Doppler spectrogram, (b) the maximum velocity envelope of this Doppler spectrogram, and (c) the wrist pulse signal after de-noising and drift removal.
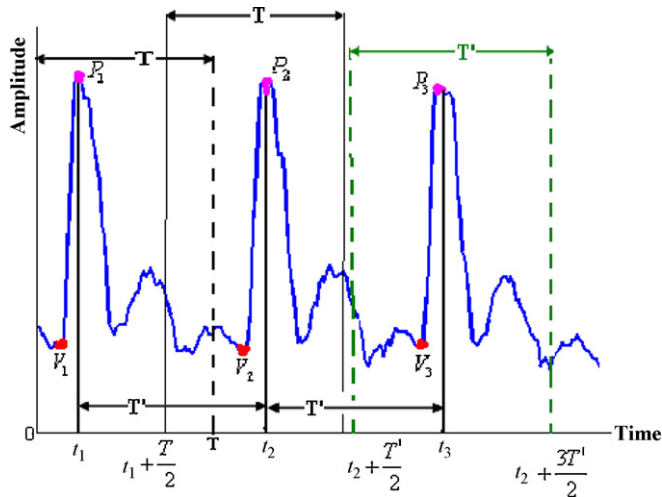
Fig. 3. Illustration of the wrist pulse signal segmentation process.

Fig. 5. Illustration of the decomposition of a single-period wrist pulse waveform.

instant is t_3. It should be noted that T' is used here instead of T. The time interval between P_2 and P_3 is denoted as $T' = t_3 - t_2$.

5. Repeat step 4 until all the peak points for the pulse signal are detected. These peak points are denoted as P_i, $i = 1, \ldots, n$.

6. Once the peak points are detected, we can search for the start points on the left side of each peak point. On the left side of each P_i ($i = 2, \ldots, n$), find out the local minimum point (denoted as V_i) which is nearest to the P_i. The corresponding time instant of V_i is denoted as t_i'.

As a result, each start point V_i of the pulse signal can be detected. The pulse signal at each time interval $[t_i', t_{i+1}']$ ($i = 2, \ldots, n$) consists of two complete waves: a systolic wave and a diastolic wave. Thus the pulse signal can be partitioned into multiple cycles according to the start points (see Fig. 4 for examples).

3. Feature extraction and feature selection

3.1. A two-term Gaussian model

With the method described in Section 2, the wrist pulse signal can be partitioned into several single-period waveforms for further feature extraction. Fig. 5(a) shows one typical period of a wrist
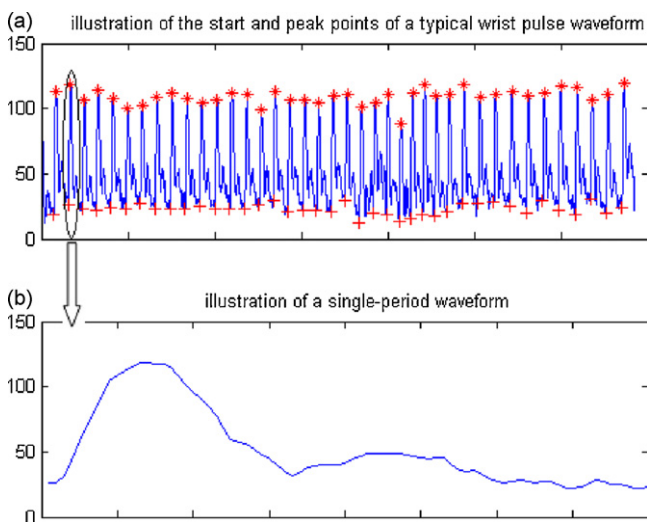

Fig. 4. (a) Illustration of the start and peak points for a typical wrist pulse signal; (b) a single-period waveform of the wrist pulse signal divided using the start point.
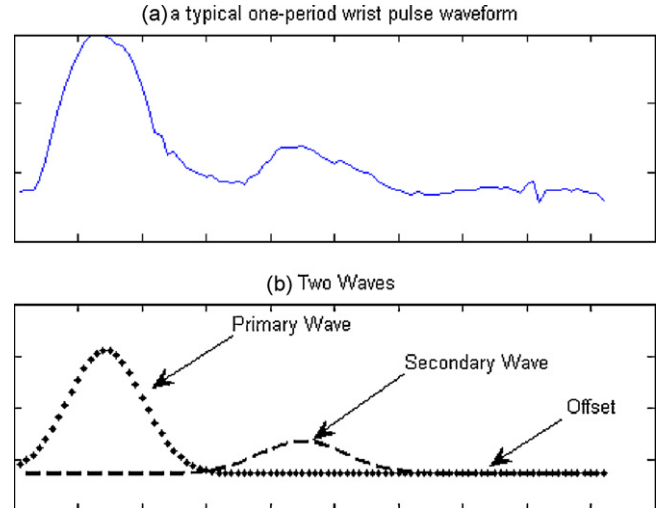
pulse signal. A further examination of the waveform in Fig. 5(a) reveals that this single-period wrist pulse signal can be seen as the superimposition of two waves: a primary wave with higher amplitude and a secondary wave with lower amplitude and a phase shift. This distinctive characteristic is caused by the rhythmic contraction and relaxation of the heart [22]. The primary wave, also called as the systolic component, is generated when the left ventricle of the heart is in contraction forcing blood into the aorta. The secondary wave is due to the phenomenon of wave reflection, which is an echo of the primary wave and usually occurring when the left ventricle of the heart is in relaxation following systole. The primary wave mainly contains information of the heart itself while the secondary wave contains information of the reflection sites and the periphery of the arterial system [22]. Moreover, the secondary wave tends to increase the load to the heart and plays a major role in determining the wrist pulse waveform patterns [23]. Therefore, how to extract these two waves, particularly the secondary wave from the wrist pulse signal is crucial for diagnosis.

Since both of the two waves are 'bell-shaped' curves with relative phase shift to each other, the pulse signal in Fig. 5(a) can be expressed by a two-term Gaussian function with an offset:

$$f(x|A_1, \tau_1, \sigma_1, A_2, \tau_2, \sigma_2, d) = A_1 * e^{-((x-\tau_1)/\sigma_1)^2}$$
$$+ A_2 * e^{-((x-\tau_2)/\sigma_2)^2} + d \tag{1}$$

where the primary wave and the secondary wave are extracted as $A_1 * e^{-((x-\tau_1)/\sigma_1)^2} + d$ and $A_2 * e^{-((x-\tau_2)/\sigma_2)^2} + d$, respectively (refer to Fig. 5(b)). It can been seen from Eq. (1) that there exist seven coefficients in the Gaussian model: A_1, A_2, τ_1, τ_2, σ_1, σ_2 and d. Among them, A_1 and A_2 determine the amplitudes of the two waves, d is the offset, τ_1 and τ_2 are the phases of the two waves, while σ_1 and σ_2 determine the width of two bell-shaped waves.

These coefficients are obtained by using the nonlinear least squares formulation to fit the Gaussian model to the wrist pulse signal. For simplicity, we assume that the Gaussian model for data fitting can be expressed as:

$$y = f(X, \beta) + \varepsilon \tag{2}$$

where y is an n-by-1 wrist pulse signal, f is a function of β and X. β is a parameter vector including the seven coefficients in the Gaussian model. X is the n-by-m design matrix for the model. ε is an n-by-1 vector of errors.

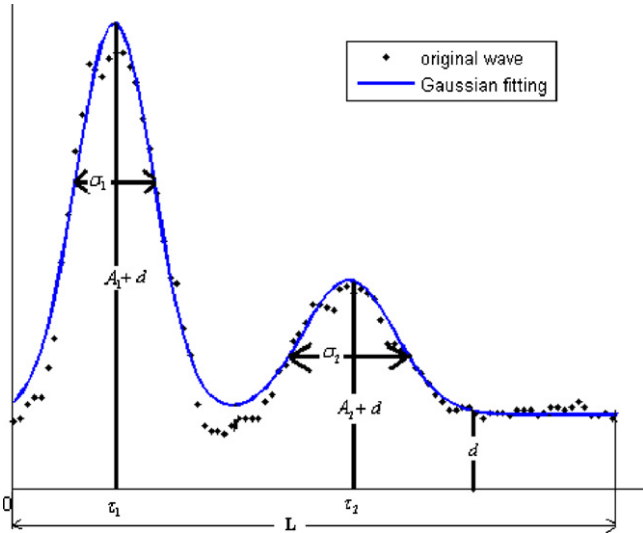The fitting process can then be determined as follows:

Fig. 6. Gaussian model fitting parameters for a typical single-period waveform.

1. Initial estimate of each parameter. Based on our experimental experience, some reasonable starting values of these parameters are made.
2. Produce the fitted curve for the current set of coefficients. The fitted response value $\hat{y}$ is given by $\hat{y} = f(X, \beta)$ and involves the calculation of the Jacobian of $f(X, \beta)$, which is defined as a matrix of partial derivatives taken with respect to the coefficients.
3. Adjust the coefficients and determine whether the fit improves. The direction and magnitude of the adjustment depend on the fitting algorithm. Some algorithms, such as trust-region [24], Levenberg–Marquardt [25] and Gauss–Newton [26] can be the options. In this study, trust-region algorithm is selected because it can solve difficult nonlinear problems more efficiently than the other algorithms.
4. Iterate the process by returning to step 2 until the fit reaches the specified convergence criteria.

As an example, Fig. 6 shows an original wrist pulse signal in a single-period and its Gaussian fitting result. It can be seen that the fitted curve using the two-term Gaussian model is in good agreement with the original signal.

Except for the above Gaussian parameters, the length of the single-period waveform L is also calculated. After we separate a wrist pulse signal into single periods, the length of each single period can be determined (i.e. L is the number of points between two consecutive start points V_i and V_{i+1}). To summarize, the obtained parameters can be divided into the ones associated with magnitude, like A_1, A_2 and d, and the ones associated with time, like τ_1, τ_2, σ_1, σ_2 and L. These parameters are illustrated in Fig. 6 as well.

By using the curve fitting technology, the models of the primary wave and secondary wave in the pulse signal can be obtained. Obtaining parameters by Gaussian curve fitting has two advantages. First, the noise contained in the original pulse signal can be reduced. Second, the information contained in the primary wave and the secondary wave can be obtained in a straightforward way. Particularly, even when the primary wave and secondary wave contained in a pulse signal cannot be easily distinguished either because of noise or because of the intrinsic characteristic of the pulse signal, this curve fitting using Gaussian model can still reliably extract related parameters.

After the Gaussian model has been identified, the parameters $\{A_i, \tau_i, \sigma_i, L\}$ ($i = 1, 2$), which represent the amplitude, phase and shape information of the two waves as well as the length infor-

Table 1
Feature candidates for wrist pulse signals.

Relative parameters	Parameter value
Ratio of the amplitude of the primary wave to the amplitude of the secondary waves	A_2/A_1
Ratio of the phase of the primary wave to the phase of the secondary waves	τ_2/τ_1
Ratio of the shape of the primary wave to the shape of the secondary waves	σ_2/σ_1
Ratio of the phase of the primary wave to the length of a single-period waveform	τ_1/L
Ratio of the phase of the secondary wave to the length of a single-period waveform	τ_2/L
Ratio of the shape of the primary wave to the length of a single-period waveform	σ_1/L
Ratio of the phase of the secondary wave to the length of a single-period waveform	σ_2/L

mation, can be obtained. Generally, the relative values between these two waves can provide more reliable information and therefore are taken as feature candidates. The seven relative values used in this research are illustrated in Table 1. A feature vector, which represents the characteristics of a single-period waveform, is then constructed using these relative values.

3.2. Feature selection

In the previous section, we have used a Gaussian model to extract a feature vector for a single-period of a wrist pulse signal. However, there may be some closely correlated parameters in the feature vector and these parameters need to be eliminated. Since a typical wrist pulse signal contains many periods, a 'pool' of feature vectors, each corresponds to one period of the pulse signal, is obtained. The feature vectors for a typical pulse signal are illustrated in Fig. 7. It can be seen that the feature will vary with the period. The correlation coefficient matrix of these features is shown in Table 2. It can be seen that there exist three tightly correlated feature pairs: $(\tau_2/\tau_1, \tau_1/L)$, $(\sigma_2/\sigma_1, \sigma_2/L)$ and $(\tau_2/L, \tau_1/L)$. Since these pairs of features provide similar information, using only one
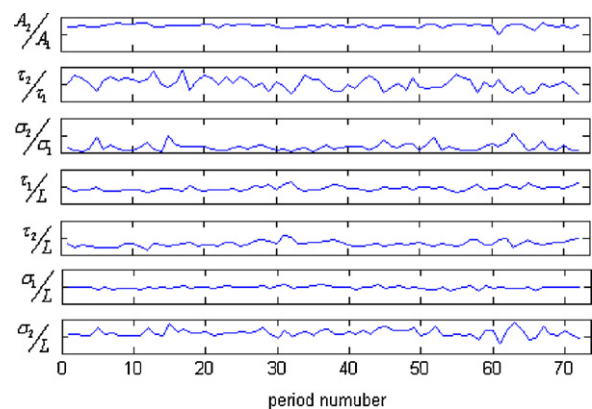
Fig. 7. Variability of the Gaussian fitting parameters of a wrist pulse waveform.

Table 2
Cross-correlation coefficients for the features.

	A_2/A_1	τ_2/τ_1	σ_2/σ_1	τ_1/L	τ_2/L	σ_1/L	σ_2/L
A_2/A_1	1.00	0.16	0.19	−0.28	−0.39	−0.17	0.35
τ_2/τ_1	0.16	1.00	−0.25	**−0.90**	−0.33	0.36	−0.04
σ_2/σ_1	0.19	−0.25	1.00	0.02	−0.35	−0.30	**0.86**
τ_1/L	−0.28	**−0.90**	0.02	1.00	**0.79**	−0.16	−0.13
τ_2/L	−0.39	−0.33	−0.35	**0.79**	1.00	0.31	−0.30
σ_1/L	−0.17	0.36	−0.30	−0.16	0.31	1.00	0.15
σ_2/L	0.35	−0.04	**0.86**	−0.13	−0.30	0.15	1.00

feature from each pair is enough for classification. The redundant features, such as τ_1/L, τ_2/L and σ_2/L, can then be eliminated from the feature vector.

So far a feature vector has been obtained which contains no tightly correlated elements. However, the remaining elements in this feature vector are still subject to further selection. Since the purpose of the study is for disease diagnosis, only the disease-sensitive features are required. A statistical difference based approach is used here to select the disease-sensitive features. Assuming a training dataset is available which contains N_1 wrist pulse signals from the healthy persons and N_2 pulse signals from the patients. For a given feature α as an example, a group of this feature for the healthy person, denoted as $\{\alpha\}_H = \{\{\alpha\}_{H,1}, \{\alpha\}_{H,2}, \ldots, \{\alpha\}_{H,N_1}\}$, is established, where $\{\alpha\}_{H,i}$ ($i = 1, \ldots, N_1$) is the features extracted from the ith healthy person. Similarly, a group of this feature for patients with a certain decease is established and is denoted as $\{\alpha\}_P = \{\{\alpha\}_{P,1}, \{\alpha\}_{P,2}, \ldots, \{\alpha\}_{P,N_2}\}$, where $\{\alpha\}_{P,i}$ ($i = 1, \ldots, N_2$) is the features extracted from the ith patients. The statistical difference between these two groups can be calculated as:

$$\text{statistical difference of } \alpha = \frac{|\overline{\{\alpha\}}_H - \overline{\{\alpha\}}_P|}{S_{\overline{\{\alpha\}}_H, \overline{\{\alpha\}}_P}} \tag{3}$$

where $\overline{\{\alpha\}}_H$ and $\overline{\{\alpha\}}_P$ are the means of $\{\alpha\}_H$ and $\{\alpha\}_P$, respectively. $S_{\overline{\{\alpha\}}_H, \overline{\{\alpha\}}_P}$ is defined as:

$$S_{\overline{\{\alpha\}}_H, \overline{\{\alpha\}}_P} = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}} \tag{4}$$

where S_1^2 and S_2^2 are the variances of $\{\alpha\}_H$ and $\{\alpha\}_P$, respectively.

For each feature, its statistical difference between the healthy persons and patients with a certain disease reflects the sensitivity of this feature to the disease; therefore, the statistical difference determines whether this feature should be selected. If the statistical difference of a feature is relatively large, then this feature is a good indictor for this disease and should be selected for classification. Otherwise, the feature is not good enough and should not be used.

4. FCM clustering

The selected features are then used as inputs to the classifier for further pattern classification, which is to determine from these features that whether the pulse signals are from healthy persons or from patients with certain disease. In this study, a FCM classifier is adopted.

Clustering is a common technique for statistical data analysis. It aims to cluster data points into clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible [27]. There are many algorithms for fuzzy clustering, and the FCM is one of the most widely used ones [28]. In this study, after selecting the disease-sensitive features, we use the FCM to make the pattern classification. The FCM is used in this study due to it ability to classify data belonging to two or more groups. Moreover, another aspect of the FCM is the use of membership function, which means an object can belong to several clusters at the same time but with different degrees. Such characteristic is important for the disease diagnosis.

5. Experimental result

By collaborating with the Harbin 211 hospital (Harbin, Heilongjiang Province, China), we collected the wrist pulse signals using a Doppler ultrasonic blood analyzer from both healthy persons (100 samples) and patients with different diseases (88 samples). Half of these data, which are randomly selected, are used

Table 3
Gaussian fitting parameters for a healthy person (Group H), a pancreatitis patient (Group P) and a DBU patient (Group DBU).

Gaussian fitting parameters	A_1	A_2	τ_1	τ_2	σ_1	σ_2	L
Group H	130.4	44.6	15.8	50.3	8.0	9.6	90
Group P	85.7	34.7	17.0	41.5	6.7	26.6	120
Group DBU	110.2	27.8	13.8	45.4	8.7	18.7	63

Table 4
Mean cross-correlation coefficients for Group H.

	A_2/A_1	τ_2/τ_1	σ_2/σ_1	τ_1/L	τ_2/L	σ_1/L	σ_2/L
A_2/A_1	1.00	0.24	0.08	−0.38	−0.27	−0.34	−0.03
τ_2/τ_1	0.24	1.00	−0.41	**−0.88**	0.13	0.42	−0.23
σ_2/σ_1	0.08	−0.41	1.00	0.24	−0.40	−0.43	**0.91**
τ_1/L	−0.38	**−0.88**	0.24	1.00	0.34	−0.19	0.15
τ_2/L	−0.27	0.13	−0.40	0.34	1.00	0.46	−0.24
σ_1/L	−0.34	0.42	−0.43	−0.19	0.46	1.00	−0.02
σ_2/L	−0.03	−0.23	**0.91**	0.15	−0.24	−0.02	1.00

for the training purpose, and the remaining data are used for testing. The testing dataset includes 50 signals from healthy persons (Group H), 23 signals from patients with pancreatitis (Group P) and 21 signals from patients with Duodenal Bulb Ulcer (Group DBU).

As was described previously, these signals are first partitioned into single-period waveforms. Then the modified Gaussian model is used to fit each single-period waveform. As an example, Fig. 8 illustrates three pulse signals (single-period), which are from Group H, Group P and Group DBU, respectively, as well as the corresponding fitting results using Gaussian models. The values of the Gaussian fitting parameters A_1, A_2, τ_1, τ_2, σ_1, σ_2 and the length of each waveform L are calculated (see Table 3).

As was discussed in Section 3.2, the cross-correlation analysis is used to find out the tightly correlated features. Tables 4–6 show the mean value of the cross-correlation coefficients for Group H (Table 4), Group P (Table 5) and Group DBU (Table 6). All these calculations are based on the training dataset. Two observations can be reached from these tables: first, the magnitude related feature, for example A_2/A_1, is not correlated with the time-relevant features, such as τ_2/τ_1 and σ_2/σ_1; second, for all the three groups, two tightly correlated pairs can be detected: (τ_2/τ_1, τ_1/L) and (σ_2/σ_1, σ_2/L).

As a result, two features can be eliminated because of the two tightly correlated pairs. The resulted feature vector contains 5 features: A_2/A_1, τ_2/τ_1, σ_2/σ_1, τ_2/L and σ_1/L. The statistical differences

Table 5
Mean cross-correlation coefficients for Group P.

	A_2/A_1	τ_2/τ_1	σ_2/σ_1	τ_1/L	τ_2/L	σ_1/L	σ_2/L
A_2/A_1	1.00	0.43	0.62	0.35	0.29	0.07	0.68
τ_2/τ_1	0.43	1.00	−0.29	**0.89**	0.22	0.37	**0.85**
σ_2/σ_1	0.62	−0.29	1.00	−0.43	−0.10	0.08	0.10
τ_1/L	0.35	**0.89**	−0.43	1.00	**0.77**	0.25	**0.79**
τ_2/L	0.29	0.22	−0.10	**0.77**	1.00	0.01	0.38
σ_1/L	0.07	0.37	0.08	0.25	0.01	1.00	0.33
σ_2/L	0.68	**0.85**	0.10	**0.79**	0.38	0.33	1.00

Table 6
Mean cross-correlation coefficients for Group DBU.

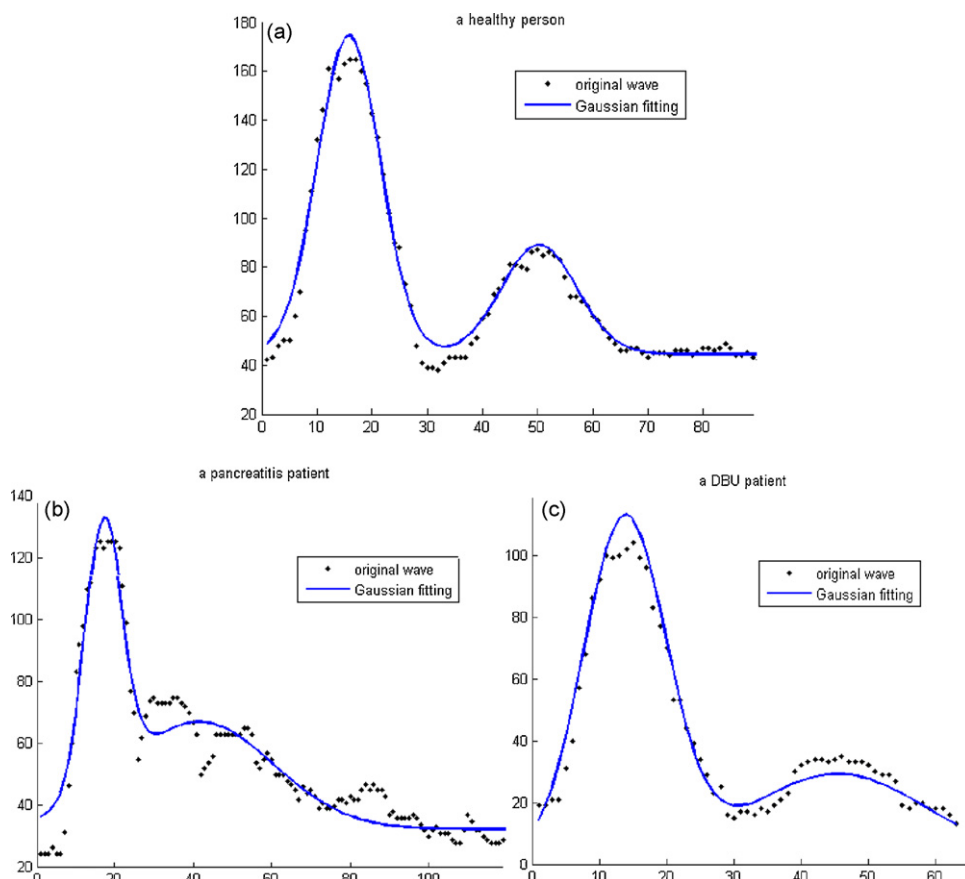|	A_2/A_1	τ_2/τ_1	σ_2/σ_1	τ_1/L	τ_2/L	σ_1/L	σ_2/L
A_2/A_1	1.00	−0.09	−0.10	0.18	0.24	−0.12	−0.06
τ_2/τ_1	−0.09	1.00	−0.14	**−0.84**	−0.39	0.47	−0.02
σ_2/σ_1	−0.10	−0.14	1.00	−0.30	−0.70	−0.85	**0.94**
τ_1/L	0.18	**−0.84**	−0.30	1.00	**0.82**	−0.08	−0.43
τ_2/L	0.24	−0.39	−0.70	**0.82**	1.00	0.41	**−0.79**
σ_1/L	−0.12	0.47	−0.85	−0.08	0.41	1.00	**−0.78**
σ_2/L	−0.06	−0.02	**0.94**	−0.43	**−0.79**	**−0.78**	1.00

Fig. 8. Gaussian curve fitting results for (a) a healthy person, (b) a pancreatitis patient, and (c) a DBU patient, respectively, where the dots represent the single-period wrist pulse waveform and the solid line is its Gaussian fitting result.

Table 7
The statistical differences of the selected features for the three groups.

	Group H vs. Group P	Group H vs. Group DBU	Group P vs. Group DBU
A_2/A_1	7.01	5.65	8.98
τ_2/τ_1	3.71	2.15	4.62
σ_2/σ_1	14.29	7.79	9.47
τ_2/L	8.64	8.61	10.36
σ_1/L	1.53	2.96	4.01

of the selected features are calculated and the results are show in Table 7. Again, the calculation is based on the training dataset. It can be seen that the feature σ_2/σ_1, which has the largest statistical difference, is the best parameter to be used in order to distinguish between Group H and Group P. Similarly, the other two features τ_2/L and A_2/A_1 can also be selected to due to their relatively large

statistical difference for all these three groups. In conclusion, three features (σ_2/σ_1, τ_2/L and A_2/A_1) with relatively large statistical difference for all the three groups are selected as the features for classification.

The classification results on the testing dataset using the FCM classifier are shown in Table 8. It can be seen that an accuracy of 94.5% is obtained in distinguishing the Group H and Group P using feature σ_2/σ_1 alone, which means only 2 of the 50 healthy persons and 2 of the 23 pancreatitis patients are misclassified. The classification result for the healthy person and DBU patients is relatively low, which confirms the previous results of statistical difference in Table 7. By mixing all these three groups together, the accuracy of the classification can still reach 85%, which is quite encouraging because no previous work has been done in distinguishing more than two groups. In Table 8, the classification results of the proposed method are compared with the previous wavelet transform

Table 8
Classification result using FCM.

Sample class	Testing samples		Classification results	Accuracy (%) (Gaussian model)		Accuracy (%) (AR model [18])		Accuracy (%) (WT method [13])	
Group H	50	73	48(2)	96.0	94.5	88.9	86.3	86.7	84.8
Group P	23		21(2)	91.3		80.6		80.5	
Group H	50	71	42(8)	84.0	85.9	85.7	82.3	88.9	85.4
Group DBU	21		19(2)	90.4		74.3		77.1	
Group P	23	44	21(2)	91.3	90.9	85.7	87.3	80.6	82.2
Group DBU	21		19(2)	90.4		88.9		83.7	
All three groups									
Group H	50	94	48(2)	84.0	85.1	86.0	78.9	82.4	71.7
Group DBU	21		21(2)	76.2		77.1		72.7	
Group P	23		42(8)	95.7		73.5		60.0	

method [13] and the auto-regressive method [18]. It can be found that the proposed method can provide a better classification accuracy (i.e. about 8% higher than the AR model for distinguishing healthy person and patients with pancreatitis). This result indicates that the proposed method has a great potential in pulse diagnosis for the current situation.

6. Conclusion

The wrist pulse signal of a person contains important information about the pathologic changes of the person's body condition. Extracting this information from the wrist pulse waveforms is important for computerized pulse diagnosis. A modified Gaussian model was proposed in this paper to extract useful features from each single-period waveform of a wrist pulse signal. The features were then selected by using the cross-correlation analysis and the statistical difference calculation. The performance of the selected parameters was evaluated using the established testing dataset, including both healthy persons and patients. The experimental results demonstrate that the proposed method performs well for the current research: an accuracy of over 90% can be reached in distinguishing the healthy person from the patients with some specific types of diseases. Moreover, an accuracy of over 85% can be reached in distinguishing healthy persons from the mixed kinds of diseases.

Conflict of interest

All the authors declare that we have no conflict of interest.

References

[1] Lukman S, He Y, Hui S. Computational methods for traditional Chinese medicine: a survey. Computer Methods and Programs in Biomedicine 2007;88:283–94.
[2] Hammer L. Chinese pulse diagnosis—contemporary approach. Eastland Press; 2001.
[3] Zhu L, Yan J, Tang Q, Li Q. Recent progress in computerization of TCM. Journal of Communication and Computer 2006;3(7).
[4] Wang K, Xu L, Zhang D, Shi C. TCPD based pulse monitoring and analyzing. In: Proceedings of the 1st ICMLC conference. 2002.
[5] Wang H, Cheng Y. A quantitative system for pulse diagnosis in Traditional Chinese Medicine. In: Proceedings of the 27th IEEE EMB conference. 2005.
[6] Lau E, Chwang A. Relationship between wrist-pulse characteristics and body conditions. In: Proceedings of the EM2000 conference. 2000.
[7] Lu W, Wang Y, Wang W. Pulse analysis of patients with severe liver problems. IEEE Engineering in Medicine and Biology Magazine 1999;18(January/February (1)):73–5.
[8] Wang B, Luo J, Xiang J, Yang Y. Power spectral analysis of human pulse and study of traditional Chinese medicine pulse-diagnosis mechanism. Journal of Northwestern University (Natural Science Edition) 2001;31(1):22–5.
[9] Wang Y, Wu X, Liu B, Yi Y. Definition and application of indices in Doppler ultrasound sonogram. Journal of Biomedical Engineering of Shanghai 1997;18:26–9.
[10] Ruano M, Fish P. Cost/benefit criterion for selection of pulsed Doppler ultrasound spectral mean frequency and bandwidth estimators. IEEE Transactions on BME 1993;40:1338–41.
[11] Leonard P, Beattie TF, Addison PS, Watson JN. Wavelet analysis of pulse oximeter waveform permits identification of unwell children. Journal of Emergency Medicine 2004;21:59–60.
[12] Zhang Y, Wang Y, Wang W, Yu J. Wavelet feature extraction and classification of Doppler ultrasound blood flow signals. Journal of Biomedical Engineering 2002;19(2):244–6.
[13] Zhang D, Zhang L, Zhang D, Zheng Y. Wavelet based analysis of Doppler ultrasonic wrist-pulse signals. In: Proceedings of the ICBBE 2008 conference, vol. 2. 2008. p. 539–43.
[14] Chen B, Wang X, Yang S, McGreavy C. Application of wavelets and neural networks to diagnostic system development, 1, feature extraction. Computers and Chemical Engineering 1999;23:899–906.
[15] Hera1 A, Hou Z. Application of wavelet approach for ASCE structural health monitoring benchmark studies. Journal of Engineering Mechanics 2004;1:96–104.
[16] Burges C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 1998;2:121–67.
[17] Chiu C, Yeh S, Yu Y. Classification of the pulse signals based on self-organizing neural network for the analysis of the autonomic nervous system. Chinese Journal of Medical and Biological Engineering 1996;16:461–76.
[18] Chen Y, Zhang L, Zhang D, Zhang D. Pattern classification for Doppler ultrasonic wrist pulse signals. In: 5th ICBBE conference. 2009.
[19] Yoon Y, Lee M, Soh K. Pulse type classification by varying contact pressure. IEEE Engineering in Medicine and Biology Magazine 2000;19:106–10.
[20] Xu L, Zhang D, Wang K. Wavelet-based cascaded adaptive filter for removing baseline drift in pulse waveforms. IEEE Transactions on Biomedical Engineering 2005;52(11):1973–5.
[21] Xia C, Li Y, Yan J, Wang Y, Yan H, Guo R, et al. A practical approach to wrist pulse segmentation and single-period average waveform estimation. In: The ICBEI conference. 2008. p. 334–8.
[22] Walsh S, King E. Pulse diagnosis: a clinical guide. Elsevier; 2007.
[23] Shu J, Sun Y. Developing classification indices for Chinese pulse diagnosis. Complementary Therapies in Medicine 2007;15:190–8.
[24] More JJ. Recent developments in algorithms and software for trust region methods. In: Mathematical programming. NY: Springer-Verlag; 1983. p. 258–87.
[25] Mor JJ. The Levenberg–Marquardt algorithm: implementation and theory. Berlin/Heidelberg: Springer; 2006.
[26] Jorge N, Stephen W. Numerical optimization. New York: Springer; 1999.
[27] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press; 1981.
[28] Wang X, Wang Y, Wang L. Improving fuzzy c-means clustering based on feature-weight learning. Pattern Recognition Letters 2004;25:1123–32.