

# Analyzing Wrist Pulse Signals Measured with Polyvinylidene Fluoride Film for Hypertension Identification

Ruqiang Yan,<sup>\*</sup> Mengjie Zhou, Wenjun Sun, and Jingjing Meng

School of Instrument Science and Engineering, Southeast University  
No. 2, Sipailou, Nanjing 210096, China

(Received March 21, 2017; accepted July 18, 2017)

**Keywords:** wrist pulse signal, wavelet packet transform, hidden Markov model, local discriminant bases

The human wrist pulse signal, which often exhibits nonstationarity, can reflect changes in mechanisms and pathophysiology in the blood and viscera. In this paper, we present an integrated approach for identifying hypertension using the wavelet packet transform (WPT) and continuous hidden Markov models (HMM) to analyze wrist pulse signals. The approach starts with decomposing the wrist pulse signals into a number of frequency sub-bands through the WPT. Then the local discriminant bases (LDB) algorithm is used to obtain the best representation of the wrist pulse signal in the optimal frequency sub-bands. After that, energy features are extracted from those sub-bands and the optimal features associated with corresponding sub-bands are selected using the Fisher linear discriminant criterion. The optimal features are subsequently used as input to a continuous HMM classifier for hypertension identification. Experimental results indicate that the presented approach can differentiate the hypertensive wrist pulses from healthy wrist pulses effectively. In addition, when compared with other classifiers, the results demonstrate that the continuous HMM classifier yields a better classification result.

## 1. Introduction

Hypertension, also known as high blood pressure, is a chronic disease with a rise in arterial blood pressure as the main symptom. Most of the more than one million deaths due to cerebrovascular disease every year are caused by high blood pressure. Since wrist pulse signals carry a great deal of physiological information regarding the health of a human body, research on wrist pulse signal analysis has been attracting more and more attention over the past few years. However, the wrist pulse signal is very complex; it is related to the vascular structure, blood characteristics, blood circulation system, and many other factors<sup>(1)</sup> that give rise to challenges in effective pulse signal analysis.

For computerized pulse signal diagnosis, pulse signals are firstly acquired by sensors,<sup>(2,3)</sup> and then processed to extract effective features for different classification tasks, such as diagnosis of disease, pulse waveform classification, and analysis of health conditions. Since the pulse signal is weak, nonstationary, and predominantly in the low-frequency region, appropriate signal processing methods must be adopted for the signal analysis. With the rapid development of signal processing technology, time-frequency analysis methods, such as short-time Fourier transform,<sup>(4)</sup> Hilbert–

---

<sup>\*</sup>Corresponding author: e-mail: ruqiang@seu.edu.cn  
<http://dx.doi.org/10.18494/SAM.2017.1606>

Huang transform (HHT),<sup>(5)</sup> empirical mode decomposition (EMD),<sup>(6)</sup> and wavelet transform,<sup>(7)</sup> have provided effective methods of extracting features from transient, time-varying signals. For example, the pulse signal has previously been processed by HHT to determine arterial stiffness, which can be used as an indicator for differentiating subjects with and without diabetes.<sup>(8)</sup> The ensemble EMD has been used to extract statistical features for pulse feature representation.<sup>(9)</sup> Previous works<sup>(10)</sup> confirmed that wavelet transform is superior to frequency-domain analysis for the feature extraction of the pulse waveform and can decompose signals into elementary time-frequency blocks that are well localized in both time and frequency domains.<sup>(11,12)</sup> Automated pulse recognition by using the wavelet transform to compute the main time-domain characteristic parameters of pulse signals could also be realized.<sup>(13,14)</sup> Furthermore, wavelet packet transform (WPT),<sup>(15,16)</sup> as an extension of wavelet transform, has attracted increasing attention owing to its ability to provide more flexible time-frequency decomposition.<sup>(17)</sup> It was used for frequency sub-band energy or other statistical feature extractions for pulse feature representations.<sup>(18)</sup> Although many studies have shown that wavelet transform is effective in pulse signal analysis, research has rarely focused on hypertensive features learned through the pulse signal, and features extracted by wavelet transform are chosen mostly by subjective judgment so they may not be representative.

In this paper, an integrated approach, in which WPT is combined with continuous hidden Markov models (HMM), is designed for more effective pulse signal analysis with specific application to the health status assessment of hypertensive patients. WPT is used to decompose the pulse signals into various frequency sub-bands, with energy features being extracted from those sub-bands to characterize the pulse signal. The local discriminant bases (LDB) algorithm and Fisher linear discriminant (FLD) measure are applied when choosing the optimal features to characterize pulse signals. The HMM, a probability distribution model, is then used to diagnose hypertensive subjects. The paper is organized as follows. Section 2 gives the theoretical knowledge of WPT, LDB, FLD, and HMM. In addition, some methods, such as K-means clustering and the genetic algorithm (GA), are also studied for HMM parameter initialization. Then, human wrist pulse signals measured from healthy and hypertensive subjects are analyzed and discussed in Sect. 3. Finally, conclusions are drawn in Sect. 4.

## 2. Theoretical Framework

### 2.1 WPT-based feature extraction

WPT,<sup>(15,16)</sup> different from discrete wavelet transform, can decompose a signal into equally spaced frequency sub-bands. Mathematically, a wavelet packet consists of a set of linearly combined wavelet functions that can be defined by the following recursive relationships:

$$\begin{cases} \psi^{2k}(t) = \sqrt{2} \sum_n h(n) \psi^k(2t - n), \\ \psi^{2k+1}(t) = \sqrt{2} \sum_n g(n) \psi^k(2t - n), \end{cases} \quad (1)$$

where  $\psi^0(t) = \phi(t)$  is the scaling function and  $\psi^1(t) = \psi(t)$  is the wavelet function. The symbols  $h(n)$  and  $g(n)$  represent the low-pass and high-pass filter coefficients, respectively, determined by the scaling function and wavelet function.<sup>(19)</sup> Furthermore,  $h(n)$  and  $g(n)$  are related to each other by

$g(n) = (-1)^n h(1-n)$ . Utilizing the low-pass and high-pass filter coefficients, a time-domain signal  $x(t)$  can be decomposed recursively as

$$\begin{cases} x_{j+1}^{2k}(t) = \sum_m h(m-2k)x_j^k(t), \\ x_{j+1}^{2k+1}(t) = \sum_m g(m-2k)x_j^k(t), \end{cases} \quad (2)$$

where the symbols  $j$  and  $k$  denote the decomposition level and frequency sub-band, respectively. Furthermore,  $x_j^k(t)$  denotes the wavelet coefficients at the  $j$ -th level,  $k$ -th frequency sub-band, represented by node  $(j, k)$ , and  $x_0^0(t) = x(t)$ . If each frequency sub-band at  $(j, k)$  can be considered as a subspace  $\mathbf{B}_{j,k}$ , which is spanned by a series of base vector  $\{\omega_{i,j,k}\}_{m=0}^{m=2^{N-j}-1}$ , where  $m$  is equal to the dimension of  $\mathbf{B}_{j,k}$ ,  $2^N$  corresponds to the length of the signal, and then the signal  $x_i$  can be expressed as

$$x_i = \sum_{j,k,m} [x_{j,k,m}]_i \cdot \omega_{j,k,m}, \quad (3)$$

where  $x_{j,k,m}$  is equal to the  $m$ -th coefficient of  $\mathbf{B}_{j,k}$ .

Unlike traditional signal processing methods, WPT can provide many important signal characteristics that are useful in classification. Because of the great diversity of the existing wavelet packet bases, WPT has various decomposition styles for a signal. Different wavelet packet bases possess different natures and reflect the different properties of a signal. To solve this problem, the LDB algorithm, introduced in the following subsection, has been employed to select the optimal wavelet packet bases for signal decomposition based on the data gathered.

## 2.2 LDB

The LDB algorithm is a pruning algorithm that identifies the subspaces that exhibit high discrimination attributes between signal classes, using a given dissimilarity measure.<sup>(20,21)</sup> The LDB algorithm has been employed to select the optimal wavelet packet bases to address real-world classification problems in audio,<sup>(22)</sup> radar,<sup>(23)</sup> and biomedical engineering,<sup>(24)</sup> etc. The optimal choice of LDB subspaces for a given dataset is driven by the nature of the dataset and the dissimilarity measures used to distinguish between classes. In this study, relative entropy is selected as the discriminant measure in searching for the optimal wavelet packet subspaces because of its prominent ability in characterizing the relationship between two signals.<sup>(25)</sup> In the case of two categories, the relative entropy can be defined as

$$D(p^{(1)}, p^{(2)}) = \sum_{i=1}^n p_i^{(1)} \log \frac{p_i^{(1)}}{p_i^{(2)}}, \quad (4)$$

where  $p_i^{(1)}$  and  $p_i^{(2)}$  are two non-negative sequences satisfying the condition that the sum of every sequence is one. We assume that  $\log 0 = -\infty$  and  $\log(p_i/0) = +\infty$  for  $p_i > 0$ ,  $0(\pm\infty) = 0$ . The discriminant information  $D(p^{(1)}, p^{(2)})$  between the two sequences measures how different the

distributions of the two classes are. For multiple class problems, the discriminant measure based on relative entropy is expressed as

$$D(\{p^l\}_{l=1}^L) = \sum_{i=1}^{L-1} \sum_{j=i+1}^L D(p^{(i)}, p^{(j)}), \quad (5)$$

where  $L$  is the number of classes.

Suppose that  $A_{j,k}$  represents the desired local discriminate basis restricted to the subspace of  $B_{j,k}$ , which is a set of basis vectors at node  $(j, k)$ . Then, for a given training dataset consisting of  $L$  classes of signals  $\left\{ \left\{ x_i^{(l)} \right\}_{i=1}^{N_l} \right\}_{l=1}^L$ , where  $N_l$  is the total number of training signals in class  $l$ , the steps for implementing the LDB algorithm are shown as follows.

**Step 1:** Choose a time-frequency decomposition method, such as WPT used in this study, to decompose the signals contained in the training dataset.

**Step 2:** Construct time-frequency energy maps  $C_l$  for  $l = 1, 2, \dots, L$  on the wavelet packet coefficients using following equation.

$$C_l(j, k, m) = \sum_{i=1}^{N_l} (\omega_{j,k,m}^T x_i^{(l)})^2 / \sum_{i=1}^{N_l} \|x_i^{(l)}\|^2 \quad (6)$$

**Step 3:** Suppose that  $A_{J,K} = B_{J,K}$ . Then, set  $\Delta_{j,k} = D(\{C_l(j, k, \cdot)\}_{l=1}^L)$  as in Eq. (7), which contains the dissimilarity measure of the node  $(j, k)$  for  $k = 0, 1, \dots, 2^J - 1$ .

$$D(\{C_l(j, k)\}_l^L) = \sum_{m=0}^{2^{n_0-j}-1} D(C_1(j, k, m), \dots, C_l(j, k, m)) \quad (7)$$

**Step 4:** Search for the best subspace  $A_{j,k}$ , defined as follows. For  $j = J-1, J-2, \dots, 0$  and  $k = 0, 1, \dots, 2^J - 1$ , if  $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$ , that is, the dissimilarity measure of the parent node is greater than those of cumulative children nodes, then  $A_{J,K} = B_{J,K}$ . Otherwise,  $A_{j,k} = A_{j+1,2k} \oplus A_{j+1,2k+1}$  and set  $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$ .

**Step 5:** Rank the complete basis functions found in step 3 from high to low according to their discrimination power.

Through LDB, a set of optimal wavelet packet subspaces have been selected. In this study, the sum of the squares of the coefficients in each subspace (i.e., energy feature  $E_j^k = \sum_m x_{j,k,m}^2$ ) is calculated to construct the feature vector as

$$F = \{f_1, f_2, \dots, f_t\}, \quad (8)$$

where  $f_t$  is the energy of each subspace and  $t$  is the number of subspaces chosen using LDB.

However, preliminary analysis has revealed that not all the feature components in the feature vector contribute directly to distinguishing the hypertensive subjects from the healthy subjects. This is because the human wrist pulse frequency range is between 0 and 20 Hz,<sup>(26)</sup> whereas about 99% of the energy is distributed between 0 and 10 Hz. Furthermore, an excessively high vector dimension involved in the input to a classifier may not necessarily improve the performance of classification and also increase computational time. Therefore, the FLD measure is introduced in the following subsection to reduce the feature dimension.

### 2.3 FLD for feature selection

FLD, a distance measure, has been applied to components differentiation within a class pair. Its basic idea is to project a high-dimensional model of the initial samples onto the best vector space for identification, so as to reduce the feature space dimension and extract significant information for classification. Generally, the greater the distance between two feature components within a class pair, the easier it will be to separate them. The FLD power can be expressed as

$$J_{f_l}(i, j) = \frac{|\mu_{i,f_l} - \mu_{j,f_l}|^2}{\sigma_{i,f_l}^2 + \sigma_{j,f_l}^2}, \quad (9)$$

where the symbols  $\mu_{i,f_l}$  and  $\mu_{j,f_l}$  and  $\sigma_{i,f_l}^2$  and  $\sigma_{j,f_l}^2$  are the mean values and the variances of the  $l$ -th feature  $f_l$ , for class  $i$  and class  $j$ , respectively. Features with a high discriminant power are more representative than those with low discriminant power, and will be selected as input for the subject's health status assessment.

### 2.4 HMM classifier design

HMM is a statistical model that has been successfully applied to speech recognition since the 1980s.<sup>(27)</sup> In recent years, it has been widely used, for example in text recognition and fault diagnosis.<sup>(28,29)</sup> HMM has been developed from the Markov chain. A Markov chain is a sequence of events, usually called states. They are not directly observed but are associated with a probability function. The generation of a random sequence is the result of a random walk in the chain and of an observation (also called an emission) at each visit to a state. An HMM has the following elements.<sup>(30)</sup>

- (1)  $M$  represents the number of distinct observations for each state. The observation symbols correspond to the physical output of the system being established. The individual observation symbols are denoted as  $V = \{v_1, v_2, \dots, v_M\}$ , and the state at time  $t$  is denoted as  $s_t$ .
- (2)  $N$  stands for the number of states in the model. Although the states are hidden, there are often some physical signals attached to the states of the model.
- (3)  $A = \{a_{ij}\}$  stands for the state transition probability distribution, where  $a_{ij} = P[s_{t+1} = j | s_t = i]$ ,  $1 \leq i, j \leq N$ .
- (4)  $B = \{b_{jk}\}$  stands for the observation probability distribution in state  $j$ , where  $b_{jk} = P[v_k | s_t = j]$ ,  $1 \leq j \leq N, 1 \leq k \leq M$ .
- (5)  $\pi = \{\pi_i\}$  represents the initial state distribution, where  $\pi_i = P[s_1 = i]$ ,  $1 \leq i \leq N$ .

It can be seen that a complete HMM requires the specifications of  $N$ ,  $M$ ,  $A$ ,  $B$ , and  $\pi$ . Since there is a definite relation between these parameters (e.g.,  $N$  and  $M$  can be determined if  $A$  and  $B$  are known),<sup>(31)</sup> a compact notation is often used in the literature to indicate the complete parameter set of the model for convenience.

$$\lambda = (\pi, A, B) \quad (10)$$

A typical HMM is shown in Fig. 1, where  $h_1$  from one to  $h_N$  symbolize the hidden states. Arrows indicate the relationship between variables, and the symbols above arrows indicate the state transfer probability.

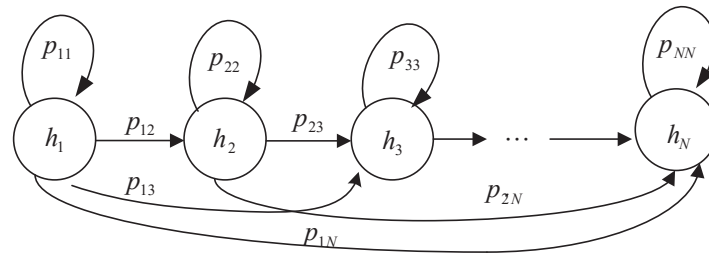


Fig. 1. A typical structure of the HMM.

In general, HMM can be divided into two categories, the discrete HMM and the continuous HMM. The latter is applied to analyzing the wrist pulse signals of healthy and hypertensive patients in this study. Compared with the typical discrete HMM, the observation probability distribution  $B$  in the continuous HMM is assumed to be generated by the Gaussian probability density function. In practical applications, one Gaussian probability density function cannot fit the probability distributions of the observation sequence, so the linear combination of several Gaussian probability density functions is used for generating the observation sequence. The observation probability distribution is expressed as<sup>(32)</sup>

$$b_j(O) = \sum_{l=1}^M C_{jl} \frac{\prod_{d=1}^D \exp[-(O - \mu_{jl})^2 / 2U_{jl}^2]}{(2\pi)^{D/2} (\prod_{d=1}^D U_{jl}^2)^{1/2}}, \quad (11)$$

where  $D$  is the number of components in  $O$  and  $M$  is the number of Gaussian probability density functions. Furthermore,  $C_{jl}$  is the  $l$ -th mixed coefficient belonging to the  $j$ -th state,  $\mu_{jl}$  is the  $l$ -th average of the mixed Gaussian belonging to the  $j$ -th state, and  $U_{jl}$  is the  $l$ -th covariance of the mixed Gaussian belonging to the  $j$ -th state. Therefore, a continuous HMM could be denoted as<sup>(33)</sup>

$$\lambda = \{\pi, A, \mu_{jl}, U_{jl}, C_{jl}\}. \quad (12)$$

When building an HMM, three basic problems must be solved: evaluation, decoding and training. Firstly, when observation sequence  $O$  and HMM  $\lambda$  are given, the evaluation problem is how to calculate the probability of observation sequence  $O$  under model  $\lambda$ . The forward-backward algorithm is used to solve this problem.<sup>(34)</sup> Secondly, the decoding problem is how to determine a best state sequence that can generate the best observation sequence when one observation sequence and a model are given. The Viterbi algorithm is used to solve this problem.<sup>(35)</sup> Thirdly, the training problem is how to determine the optimum HMM to maximize the probability of an observation sequence when the observation sequence is given, and the Baum-Welch algorithm is introduced to deal with this problem.<sup>(36)</sup> The Baum-Welch algorithm is a type of expectation maximization (EM) method based on the maximum likelihood criterion that could be defined as

$$Best = \arg \max_{i \in k} P(O|\lambda_i), \quad (13)$$

where  $O$  indicates the observation sequence and  $P(O|\lambda_i)$  indicates the probability of the output observations  $O$  under the HMM. Then, the best parameter set  $\lambda$  of the model is determined.

## 2.5 Parameter initialization

To use a continuous HMM as a classifier, the HMM parameters should be estimated first. Since it is generally recognized that the choice of the initial values of  $\pi$  and  $A$  has little effect on the results, the influence of the parameters  $\mu_{jl}$ ,  $U_{jl}$ , and  $C_{jl}$  on the classification performance is investigated in this paper. In addition to random initialization, two other methods are studied to initialize the continuous HMM parameters, as briefly introduced below.

### 2.5.1 K-means initialization

K-means clustering is a method of cluster analysis aimed at distributing  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean.<sup>(37)</sup> K-means clustering often uses Euclidean distance as the evaluation index of comparability, and is one of the popular cluster algorithms today. For HMM parameter initialization in this study,  $k$  feature vectors are randomly selected as initial cluster center points. Then, depending on the similarity criterion (Euclidean distance), each sample belongs to the cluster with the shortest distance to the clustering center, and the new central point of each cluster is calculated. If the mean square covariance of two iteration results tends to be convergent, the calculation is over; otherwise, this process is repeated to adjust the cluster center. Finally, the mean value and standard deviation of each cluster are used to estimate the Gaussian probability density function for initialization.

### 2.5.2 GA initialization

GA is an optimization algorithm that simulates Darwin's genetic choice and biological natural selection process.<sup>(38,39)</sup> It uses a coding scheme to realize parameter optimization through iterative genetic operations, including selection, crossover, and mutation, where a fitness function is usually defined to find the optimal solution. For HMM parameter initialization in this study, floating point number coding is used as the coding scheme due to the high precision requirement when the mixed Gaussian probability density function is employed for generating the observation sequence. The matrices of  $\mu_{jl}$ ,  $U_{jl}$ , and  $C_{jl}$  are then built and connected together as individual populations. The fitness function is defined as<sup>(40)</sup>

$$f(\lambda) = \sum_{k=1}^K \ln(P(O^{(k)}|\lambda)), \quad (14)$$

where  $O^{(k)}$  indicates the  $k$ -th observation sequence of the training model and  $P(O^{(k)}|\lambda)$  is the likelihood probability. After the result of the fitness function is obtained, individuals and the population are varied by genetic operations. This study employs a one-point crossover operator. It also takes a real value variation as a mutation operator and adopts a roulette wheel selection method. Supposed there are  $n$  genes and  $M$  sequences  $O^k$  ( $k = 1, 2, \dots, M$ ) of length  $T$ . In accordance with the roulette wheel selection method, one gene from every sequence is selected and the probability for gene  $g$  being selected is shown as<sup>(32)</sup>

$$P_c = \frac{f(g_c, O^k)}{\sum_k f(g_k, O^k)}. \quad (15)$$

After that the fitness function is recalculated. If the fitness matches the expectation error ( $<10^{-3}$ ) or reaches the maximum number of iterations, the results are considered to have been obtained. Otherwise, the above process is repeated.

## 2.6 Wrist pulse signal diagnosis system

With the information described above, an integrated algorithm for wrist pulse signal diagnosis can be designed. Suppose there are  $k$  kinds of targets, and each target is represented by an HMM and recorded as  $\lambda_1, \lambda_2, \dots, \lambda_k$ . The flow chart of this wrist pulse signal diagnosis system is shown in Fig. 2. Firstly,  $k$  different categories of wrist pulse signals are collected and they are each decomposed through WPT. Secondly, the LDB algorithm is applied to the selection of the optimal bases and FLD is used to select the best energy characteristic vectors. Thirdly, each class of signals is divided into two parts. One is used to train the model, and the other is used to test the model. Fourthly, the energy characteristic vectors from the training dataset are used to

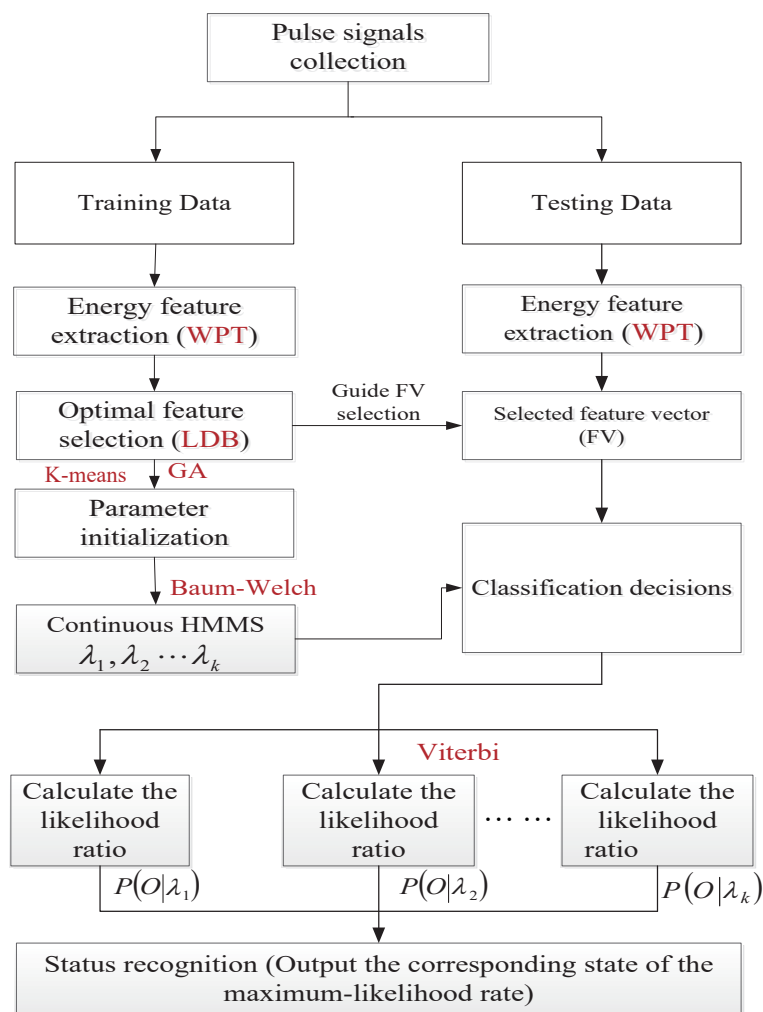


Fig. 2. (Color online) Flow chart of the pulse signal diagnosis system.



obtain the  $k$  classes of continuous HMMs using the Baum–Welch algorithm. Finally, the energy characteristic vectors from the testing dataset are put into the trained models, and the likelihood ratio is calculated using the Viterbi algorithm. A wrist pulse signal series in the testing will belong to the continuous HMM that corresponds to the largest likelihood ratio.

### 3. Experimental Evaluation

#### 3.1 Wrist signal measurement

Experimental study was conducted at Nanjing ZhongDa Hospital (affiliated to Southeast University) with informed consent that the authors have obtained. A pulse sensor, which converts the pulse to an electric signal through a polyvinylidene fluoride (PVDF) film, was placed on the surface of the wrist of the subjects. A total of 94 volunteers, including 38 healthy and 56 hypertensive subjects with an average age of 42 years old, were involved in the experiment. The ratio of males to females was close to 1. Typical healthy and hypertensive pulse signals are represented in Fig. 3. It can be seen that the descent stage of the hypertensive pulse wave has less fluctuation than that of a healthy pulse wave because the blood pressure of hypertensive patients increases sustainability.

#### 3.2 Energy feature extraction

In this study, for hypertensive and healthy subjects, each pulse signal is decomposed into four levels using the WPT algorithm with the Db4 wavelet, whose shape is more similar than those of other wavelets to the shape of the pulse wave. Since energy can be used to describe the characteristics of a given signal, the corresponding energy values at all frequency sub-bands are calculated and chosen as the feature to characterize the wrist pulse signal. As an example, Fig. 4 illustrates the energy distribution of wrist pulse signals from both the healthy and hypertensive subjects at the fourth decomposition level. It can be seen that the energy of the healthy and hypertensive signals is mainly concentrated in the low frequency sub-band, which also matches the characteristics of the low frequency of the pulse signal. The energy amplitude of the healthy signal is significantly higher than that of the hypertensive signal, and thus the energy features of the package can be used to classify health and hypertension.

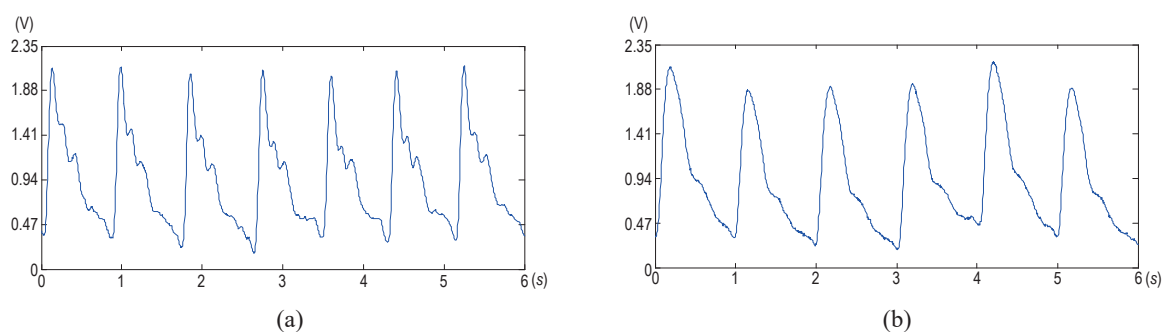


Fig. 3. (Color online) Wrist pulse signals measured using PVDF film. (a) Typical healthy pulse signal. (b) Hypertensive pulse signal.

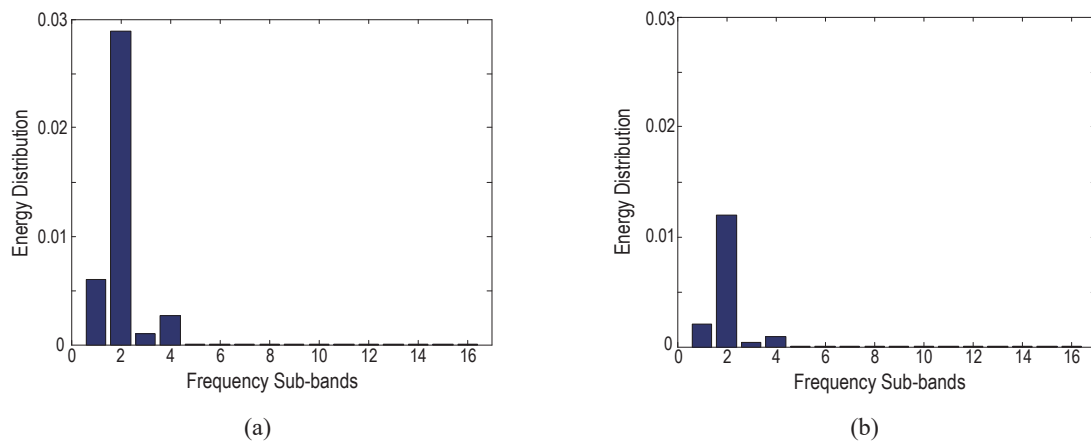


Fig. 4. (Color online) Energy features at the fourth decomposition level of the WPT. (a) Energy distribution of a healthy pulse signal. (b) Energy distribution of a hypertensive pulse signal.

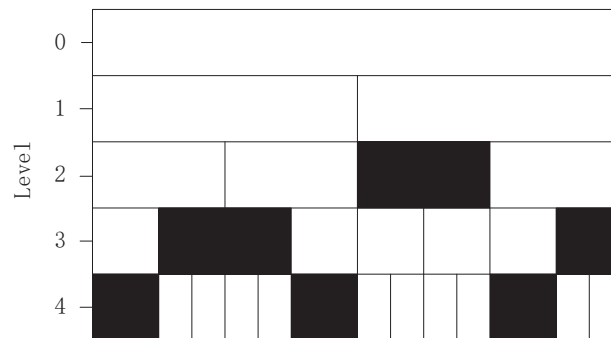


Fig. 5. Selected wavelet packet nodes.

LDB and FLD are then used to select the appropriate optimal features from all the four decomposition levels. Figure 5 shows the selected wavelet packet nodes that contain the best discriminant information for classifying health conditions after applying the LDB algorithm. It can be seen from Fig. 5 that the following subspaces are chosen as optimal orthogonal subspaces:  $\{B_{4,0}; B_{4,1}; B_{4,6}; B_{4,7}; B_{4,12}; B_{4,13}; B_{3,1}; B_{3,2}; B_{3,7}; B_{2,2}\}$ .

The corresponding FLD powers of all chosen frequency sub-bands are obtained and listed in Table 1. The first four large discriminant powers at frequency sub-bands  $B_{3,3}$ ,  $B_{4,6}$ ,  $B_{4,13}$ , and  $B_{3,2}$  are chosen. Therefore, the energy features  $E_3^3$ ,  $E_4^6$ ,  $E_4^{13}$ , and  $E_3^2$  are selected as the optimal features for continuous HMM classification.

### 3.3 HMM classification

In accordance with the human wrist pulse detection process, a continuous HMM structure (shown in Fig. 6.) that has three hidden states is chosen to simulate the changing processes of the wrist pulse. The states simulate the main, tide, and repulse waves of the wrist pulse.

The experiments were divided into three groups on the basis of different parameter initialization methods: random generation, K-means, and GA. The initial state distribution is set as  $\pi = [1, 0, 0]$ ,

Table 1  
Discriminant power of sub-bands.

Sub-band	$B_{4,0}$	$B_{4,1}$	$B_{4,6}$	$B_{4,7}$	$B_{4,12}$
Discriminant power	0.0592	0.0748	0.2186	0.1009	0.0385
Sub-band	$B_{4,13}$	$B_{3,2}$	$B_{3,3}$	$B_{3,7}$	$B_{2,2}$
Discriminant power	0.1449	0.1033	0.3283	0.0683	0.0012

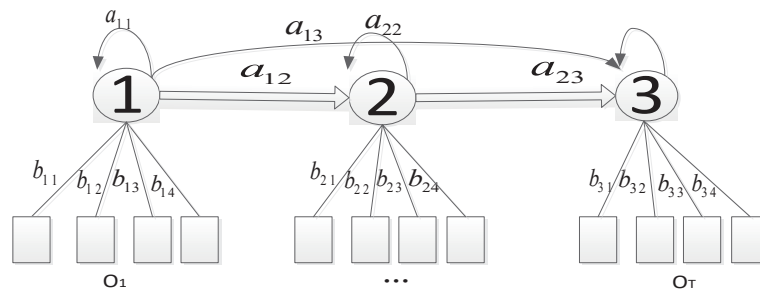


Fig. 6. Continuous HMM structure of the wrist pulse signal.

and the initial state transition probability distribution matrix  $A$  is expressed as

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (16)$$

Tables 2 and 3 are the initial parameters generated by K-means and GA, respectively. Features from a group of 20 healthy subjects and 20 hypertensive subjects are used as training data. The remaining 18 healthy subjects and 36 hypertensive subjects are used as test data and are input to the trained continuous HMMs.

The classification results of these three methods are listed in Table 4. It can be seen that continuous HMM, as a classifier, is effective in differentiating hypertensive subjects from healthy subjects. K-means and the GA algorithm, which are used to initialize the continuous HMM parameters, can improve the accuracy of the continuous HMM for classification, and GA is more effective than K-means because of its strong local search ability.

### 3.4 Comparison with other classifiers

From the discussion in Sect. 3.3, it can be seen that the continuous HMM classifier can effectively differentiate the hypertensive wrist pulses from healthy wrist pulses. To compare the classification accuracy of different classifiers, the artificial neural network (ANN) classifier, support vector machine (SVM), and k-nearest neighbor (KNN) classifier are to class the wrist pulse signals instead of the HMM classifier.<sup>(41–43)</sup> Classification results are shown in Table 5.

It can be seen that the classification accuracy rates of these three classifiers are less than 80%. The comparison of results shows that the continuous HMM classifier can effectively improve the classification results.

Table 2  
Initial parameters generated by K-means.

HMM parameters	$U_{jl}$	$\mu_{jl}$	$C_{jl}$
Hypertensive	300.4	173.9	0.2320
	19.40	27.89	0.7765
	375.6	4660	0.0038
Healthy	22.90	18.98	0.4376
	43.60	0.469	0.1458
	977.1	146.0	0.4266

Table 3  
Initial parameters generated by GA.

HMM parameters	$U_{jl}$	$\mu_{jl}$	$C_{jl}$
Hypertensive	36.30	13.20	0.1051
	39.01	0.1695	0.0923
	9485	194.7	0.8949
Healthy	4.641	2.452	0.3190
	39.41	0.7319	0.1776
	7787	146.38	0.5034

Table 4  
Classification results of three HMM parameter initialization methods.

Initialization method	Health condition	Number of subjects	Classification results	Classification accuracy (%)	Total accuracy (%)
Random values	Healthy	18	16	88.9	83.3
	Hypertensive	36	29	80.6	
K-means	Healthy	18	17	94.4	87.0
	Hypertensive	36	30	83.3	
GA	Healthy	18	16	88.9	90.7
	Hypertensive	36	33	91.7	

Table 5  
Classification results of three other classifiers.

Classifier	Total accuracy (%)
SVM	70.4
KNN	75.9
ANN	77.8

#### 4. Conclusions

An integrated signal processing approach for wrist pulse signal analysis was investigated. This approach utilizes the capability of the WPT in analyzing non-stationary signals to construct the sub-band energy features of pulse signals, and then uses them as input to a continuous HMM classifier. Three methods were used to initialize the HMM parameters and the classification results were compared. The experimental results indicated that the integrated approach is effective

in differentiating hypertensive subjects from healthy subjects, and provides a viable tool for pulse signal analysis. Furthermore, the parameter initialization approach can effectively improve the classification performance.

## References

- 1 D. Wang and F. Chen: *J. Emerg. Tradit. Chin. Med.* **13** (2004) 41.
- 2 C. Tyan, S. Liu, J. Chen, J. Chen, and W. Liang: *IEEE Trans. Biomed. Eng.* **55** (2008) 288.
- 3 P. Wang, W. Zuo, and D. Zhang: *IEEE Trans. Instrum. Meas.* **63** (2014) 1556.
- 4 E. Sejdić, I. Djurović, and J. Jiang: *Digital Signal Process.* **19** (2009) 153.
- 5 N. E. Huang, S. Zheng, and R. L. Steven: *Proc. R. Soc. London, Ser. A* **454** (1998) p. 903.
- 6 P. Flandrin, G. Rilling, and P. Goncalves: *IEEE Signal Process Lett.* **11** (2004) 112.
- 7 I. Daubechies: *IEEE Trans. Inf. Theory* **36** (1990) 961.
- 8 H. Wu, C. Lee, and A. Liu: *IEEE Trans. Biomed. Eng.* **58** (2011) 243.
- 9 H. Wu, C. Lee, C. Sun, J. Hsu, R. Huang, and C. Tang: *IEEE Trans. Instrum. Meas.* **61** (2012) 162.
- 10 L. J. Yang, Z. C. Song, S. P. Wang, and L. Chen: *Tech. Autom. Appl.* **25** (2006) 9.
- 11 I. Daubechies: *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics, Philadelphia, 1992) Chap. 1.
- 12 S. Mallat and W. L. Hwang: *IEEE Trans. Inf. Theory* **38** (1992) 617.
- 13 Zhang and H. Wang: *EURASIP J. Adv. Signal Process.* **2008** (2007) 1.
- 14 H. Wang and P. Zhang: *J. Zhejiang University SCIENCE A* **9** (2008) 1382.
- 15 M. V. Wickerhauser: *INRIA Lectures on Wavelet Packet Algorithms*, No. 31 (1991).
- 16 R. R. Coifman and M. V. Wickerhauser: *IEEE Trans. Inf. Theory* **38** (1992) 713.
- 17 L. Q. Zhang, Y. Li, and X. J. Zhang: *Beijing Biomed. Eng.* **24** (2005) 214.
- 18 L. Liu and W. Zuo: *IEEE Trans. Inf. Technol. Biomed.* **16** (2012) 598.
- 19 R. Yan and R. X. Gao: *Struct. Health Monit.* **10** (2011) 631.
- 20 Y. H. Wu, M. X. Shan, Y. N. Qian, and R. Yan: *Mech. Mater.* **226** (2012) 740.
- 21 N. Saito and R. R. Coifman: *J. Math. Imaging Vision* **5** (1995) 337.
- 22 K. Umamathy, S. Krishnan, and R. K. Rao: *IEEE Trans. Audio Speech Lang. Process.* **15** (2007) 1236.
- 23 S. Wang, G. Zhu, and Y. Y. Tang: *Int. J. Pattern Recognit. Artif. Intell.* **17** (2003) 127.
- 24 K. Englehart, B. Hudgin, and P. A. Parker: *IEEE Trans. Biomed. Eng.* **48** (2001) 302.
- 25 V. Vedral: *Rev. Mod. Phys.* **74** (2002) 197.
- 26 R. Xu and X. Yan: *J. Clin. Rehabil. Tissue Eng. Res.* **12** (2008) 671.
- 27 P. S. Aleksic and A. K. Katsaggelos: *IEEE Trans. Circuits Syst. Video Technol.* **14** (2004) 682.
- 28 S. Espana-Boquera, M. J. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez: *IEEE Trans. Pattern Anal. Mach. Intell.* **33** (2011) 767.
- 29 J. Ying, T. Kirubarajan, and K. R. Pattipati: *IEEE Trans. Syst. Man Cybern.* **30** (2000) 463.
- 30 M. Dong and D. He: *Mech. Syst. Signal Process.* **21** (2007) 2248.
- 31 G. Panahandeh, N. Mohammadiha, and A. Leijon: *Proc. IEEE I2MTC (Graz, 2012)* p. 991.
- 32 Q. Shao and J. Feng: *Mech. Sci. Technol. Aerosp. Eng.* **28** (2009) 1440.
- 33 J. Yu: *IEEE Trans. Instrum. Meas.* **61** (2012) 2200.
- 34 W. Khreich, E. Granger, A. Miri, and R. Sabourin: *Pattern Recognit. Lett.* **31** (2010) 91.
- 35 W. Turin and R. Jana: *Proc. IEEE Wireless Communications and Networking Conf. (Sydney, NSW, 2010)* p. 1.
- 36 L. R. Welch: *IEEE Inf. Theory Soc. Newsl.* **53** (2003) 194.
- 37 Z. Huang: *Data Min. Knowl. Discovery* **2** (1998) 283.
- 38 X. Li and S. Kwong: *Int. J. Speech Technol.* **13** (2010) 61.
- 39 L. He and B. Xuan: *Comput. Eng. Design* **29** (2008) 1253.
- 40 B. Juang, W. Hou, and C. Lee: *IEEE Trans. Speech Audio Process.* **5** (1997) 257.
- 41 L. Xu, M. Q. Meng, K. Wang, W. Lu, and N. Li: *Expert Syst. Appl.* **36** (2009) 3805.
- 42 D. Jia, D. Zhang, and N. Li: *Comput. Math. Methods Med.* **2014** (2014) 1.
- 43 D. Zhang, W. Zuo, D. Zhang, H. Zhang, and N. Li: *EURASIP J. Adv. Signal Process.* **2010** (2010) 1.