

## Research Article

# Pulse Diagnosis Signals Analysis of Fatty Liver Disease and Cirrhosis Patients by Using Machine Learning

Wang Nanyue,<sup>1</sup> Yu Youhua,<sup>1</sup> Huang Dawei,<sup>1</sup> Xu Bin,<sup>2</sup> Liu Jia,<sup>1</sup> Li Tongda,<sup>3</sup> Xue Liyuan,<sup>4</sup> Shan Zengyu,<sup>1</sup> Chen Yanping,<sup>1</sup> and Wang Jia<sup>5</sup>

<sup>1</sup>Experimental Research Center, China Academy of Chinese Medical Sciences, Beijing 100700, China

<sup>2</sup>Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

<sup>3</sup>GULOU Hospital of TCM of Beijing, Beijing 100009, China

<sup>4</sup>Tai Yang Gong Health Care Center, Chaoyang District, Beijing 100102, China

<sup>5</sup>Shi Xuemin Academician Office, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China

Correspondence should be addressed to Yu Youhua; yuyh62@126.com.cn

Received 21 January 2015; Revised 28 May 2015; Accepted 15 November 2015

Academic Editor: Zhaohui Liang

Copyright © 2015 Wang Nanyue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objective.** To compare the signals of pulse diagnosis of fatty liver disease (FLD) patients and cirrhosis patients. **Methods.** After collecting the pulse waves of patients with fatty liver disease, cirrhosis patients, and healthy volunteers, we do pretreatment and parameters extracting based on harmonic fitting, modeling, and identification by unsupervised learning Principal Component Analysis (PCA) and supervised learning Least squares Regression (LS) and Least Absolute Shrinkage and Selection Operator (LASSO) with cross-validation step by step for analysis. **Results.** There is significant difference between the pulse diagnosis signals of healthy volunteers and patients with FLD and cirrhosis, and the result was confirmed by 3 analysis methods. The identification accuracy of the 1st principal component is about 75% without any classification formation by PCA, and supervised learning's accuracy (LS and LASSO) was even more than 93% when 7 parameters were used and was 84% when only 2 parameters were used. **Conclusion.** The method we built in this study based on the combination of unsupervised learning PCA and supervised learning LS and LASSO might offer some confidence for the realization of computer-aided diagnosis by pulse diagnosis in TCM. In addition, this study might offer some important evidence for the science of pulse diagnosis in TCM clinical diagnosis.

## 1. Introduction

Pulse diagnosis had played an important role in clinical diagnosis and therapeutic evaluation of TCM for several thousand years. Modern researches of pulse diagnosis based on the modern technology, such as the signal analysis, are very important for the development of TCM.

Machine learning builds empirical models on data for analysis and forecasting, which has recently been used for TCM data analysis [1], especially for the diagnosis data of TCM. Some research about modeling and symptom selection for multilabel data in the inquiry diagnosis of coronary heart disease (CHD) [1, 2] and multiclass support vector machines in lip diagnosis [3] offered some available new methods in modern research of TCM. Pulse diagnosis is more difficult in standardization researches. Our work team had done some

work about it, including data analysis for pregnant women, animals with High Blood Pressure and Heart Failure, rats after nephrectomy, and patients with CHD, HBP, and so forth [4–7]. The results are encouraging.

Supervised learning and unsupervised learning are the primary methods of machine learning. In the research of pulse diagnosis in TCM, it is very difficult to collect large number of samples with high quality. Accordingly, in this study, we combined unsupervised (PCA) with supervised methods (LS and LASSO) to analyze the signals collected from patients with different diseases and healthy volunteers for cross-reference to achieve reliable results in identification by signals of pulse diagnosis in TCM.

Fatty liver disease (FLD) and cirrhosis are both common liver diseases in clinic with high incidence. FLD is generally

described as the build-up of fat in the liver cells. The prevalence of nonalcoholic fatty liver disease (NAFLD) ranges from 9 to 36.9% of the population in different parts of the world [8–10]; even in army, the incidence of nonalcoholic fatty liver disease is about 17.1% in navy flight crew and submariners [11].

Cirrhosis is a result of advanced liver disease. It is characterized by replacement of liver tissue by fibrosis and regenerative nodules. Cirrhosis is most commonly caused by alcoholism, hepatitis B and hepatitis C, and fatty liver disease. Cirrhosis is a leading cause of death in the world. In Europe, 95,609 males and 53,123 females died of cirrhosis in 2002, with large differences in age adjusted death rates among the different European geographical areas [12]. Complications such as ascites, esophageal variceal bleeding, hepatic encephalopathy, and hepatorenal syndrome are the main cause of death in this kind of disease.

Traditional Chinese Medicine plays an important role in the treatment of the two diseases, and pulse diagnosis can help clinic doctors during the diagnosis and the treatment, including prescription and evaluation. Experienced doctors can feel the difference between patients and healthy people just by pulse feeling; someone can even separate cirrhosis patients from FLD patients, but we get no evidence.

In this study, we would like to analyze the pulse signals collected by pulse-collecting instrument, and 3 groups of people are collected: healthy volunteers, patients with FLD, and patients with cirrhosis. Supervised learning and unsupervised learning are used in this study, and the results are encouraging.

## 2. Objects

We collected the pulse waves of 100 healthy volunteers in the graduated students' institute of China Academy of Chinese Medical Sciences and Tsinghua University. 50 patients with FLD were collected in China Academy of Chinese Medical Sciences and 50 patients with cirrhosis were collected in Guang'anmen Hospital of China Academy of Chinese Medical Sciences from December 2012 to July 2013. All the volunteers and patients were asked to fill the questionnaires. According to the quality of the signals and the integrity of their information we chose 98 cases from the healthy volunteers, 38 cases from the patients with FLD, and 27 cases from the patients with cirrhosis.

## 3. Methods

**3.1. Signal Collection.** Volunteers were asked to sit and keep silence to adjust their breath for 15 min firstly; then we collect the pulse waves in 3 places of both left and right sides of the radial artery called “cun,” “guan,” and “chi” in TCM (Figure 1 shows the details) for 40 s by the instrument called “Collection and Analysis System of Pulse Diagnosis Signals in TCM (patent number of pulse-collecting instrument: 200810225717.0).

This study had been demonstrated by the ethics committee of Experimental Research Center of China Academy

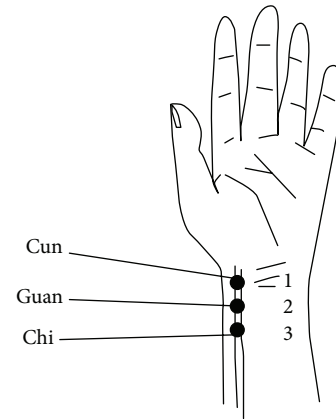


FIGURE 1: Places of “cun,” “guan,” and “chi” in the left hand.

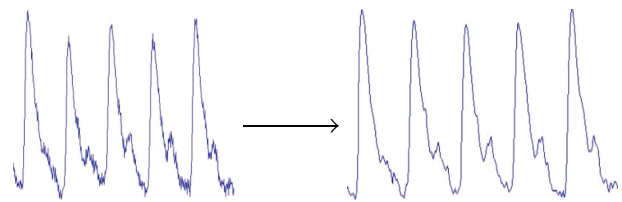


FIGURE 2: Pretreatment result: the figure on the left side is the original signal, and the figure on the right side is the signal after pretreatment; the main characters remain unchanged but the noise effectively reduced.

of Chinese Medical Sciences, and each of the patients and volunteers had read and signed the informed consent.

### 3.2. Pretreatment and Parameters Extracting Based on the Harmonic Fitting

**3.2.1. Pretreatment.** We do pretreatment [4] for the signals by the database administration and analysis system of pulse diagnosis signals in TCM. The first step is to filter the frequencies out of the range from 0.5 Hz to 48 Hz by DCT and then IDCT to recover the pulse waves without those components. The pretreatment results are shown in Figures 2, 3, and 4.

**3.2.2. Periodic Division.** We do periodic division for the pulse waves by the speed and acceleration of signals' intension and changes. Regarding the fact that the periods of human pulse are not exactly the same no matter the status, the lengths and the figures of each cycle are also not completely similar.

**3.2.3. Building Mathematics Model to Extract Parameters Based on Harmonic Fitting and Recursive Total Least Squares.** Firstly, format trigonometric functions by the lengths of every period and do LSQ for all cycles; then build the models fit to 12 harmonics. The model can match all the signals with different period cycles in one sample precisely. Because we built the models based on multicycle signals,  $T_i$  of each cycle can be dissimilar.

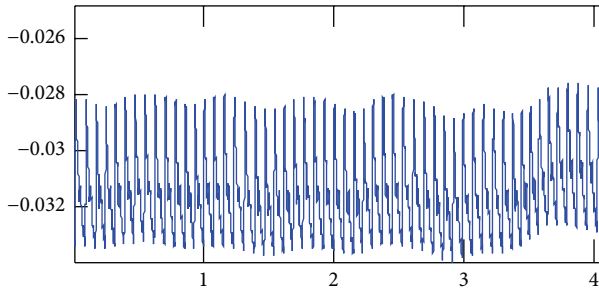


FIGURE 3: Original signals: the coordinate values of the start point of the signals regularly change because of the human breath.

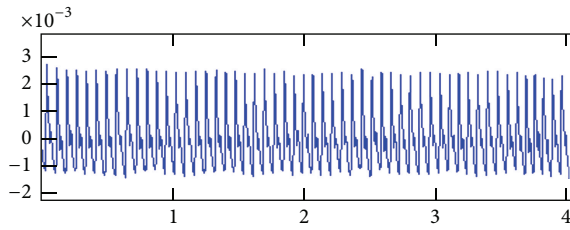


FIGURE 4: Pretreated signals: the coordinate values of the start point of the signals had been adjusted to zero, without any influence on the amplitude and other characters.

When the values of  $y_t = a_0 + \sum_{k=1}^p (a_k \cos 2kt\pi/L + b_k \sin 2kt\pi/L)$  arrive minimum, the difference of models and the original signals also be the minimum. And  $a_k$  and  $b_k$  can carry the information of all the cycles. For any cycle whose period length is  $L$ , we can use

$$y_t = a_0 + \sum_{k=1}^p \left( a_k \cos \frac{2kt\pi}{L} + b_k \sin \frac{2kt\pi}{L} \right) \quad (1)$$

to be the model of pulse waves, so  $a_k$  and  $b_k$  are the parameters we need, and the formula is

$$a_k = \sum_{i=1}^n \frac{T_i}{T} a_{i,k}, \quad (2)$$

$$b_k = \sum_{i=1}^n \frac{T_i}{T} b_{i,k}.$$

In fact we cannot construct a figure of a cycle even if we have all the characters, but we can use  $a_k$  and  $b_k$  to construct one by the formula (the fitting result is shown in Figures 5 and 6).

**3.2.4. Parameters Extracting.** To build models for the pulse waves collected from 6 places (3 places in radial artery of each side) called “cun,” “guan,” and “chi” in TCM and account for the amplitude and phase of 12 harmonics (C1–C12, F1–F12), 9 time domain parameters [13] were added such as  $h1$ ,  $t1$ ,  $h4/h1$ ,  $h5/h1$ , and  $w$  (Figure 7 shows the details). Because the period of waves from each place is similar, we can get 193 parameters from every volunteer: 32 parameters  $\times$  6 places + 1.

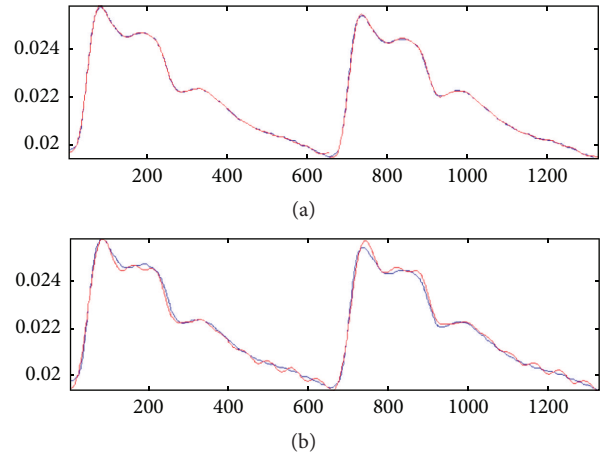


FIGURE 5: Reduction of error by recursive total least squares: the red signals are original signals and the blue ones are fitting signals. The upper figure shows the fitting effect of the two cycles by separate fitting, and the lower one shows the effect by normal fit. The separate fitting can get a better result.

### 3.3. Classification, Identification, and Features Mining

**3.3.1. Principal Component Analysis (PCA).** Principal Component Analysis [14, 15] is a method of multivariate statistical analysis to show the reason of variance of data by linear combination of the parameters. In this study we have extracted 193 parameters from pulse diagnosis signals collected from the 6 places and decomposed all the data to different orthogonal components by PCA, which means reducing the high-dimensional data to several orthogonal independent one-dimensional arrays (Figure 8 shows the details). The first principal component reflects the greatest change direction.

**3.3.2. Least Squares Regression (LS).** The purpose of classification and identifications is to establish a method to distinguish two or more groups of known data, and this method can be used to identify some new data. We do it as follows:

- (a) Design a one-dimensional vector as the optimal regressand in the meaning of canonical correlation and then do Least squares Regression:

$$\text{the regressand is } Y = \left[ \underbrace{n_2, \dots, n_2}_{n_1}, \underbrace{-n_1, \dots, -n_1}_{n_2} \right]^T. \quad (3)$$

- (b) Subtract the mean of regressors consisting of the extracted parameters in Section 3.2.
- (c) Select fewer regressors from the 193 parameters by Extended Forward Backward Least Square Regression (EFBS) [4].
- (d) For constructing the histogram of predicted indices by the regression model, we classify the samples by their indices. According to the histogram, we obtain  $p$  values of the samples belonging to either group.

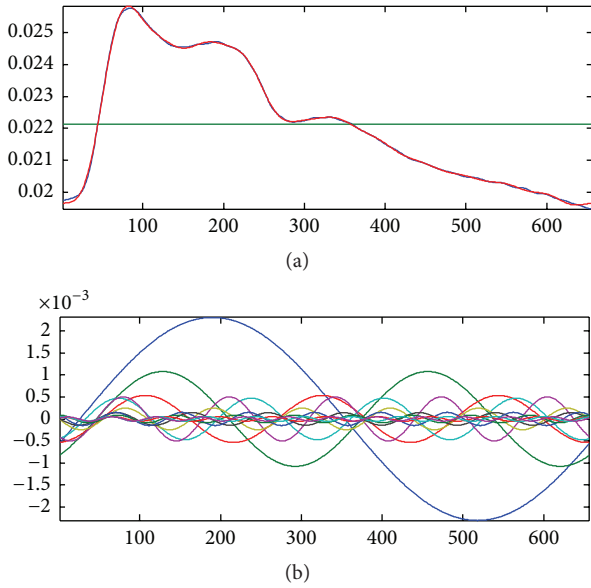


FIGURE 6: Harmonic fitting: the upper figure shows the signal of one cycle; the red one is the original signal and the blue one is the fitting result by the 12 harmonics shown in the lower figure. The fitting effect is optimal.

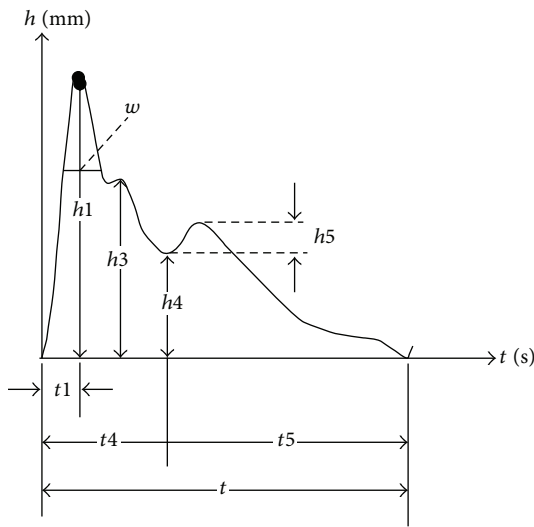


FIGURE 7: Time domain parameters.

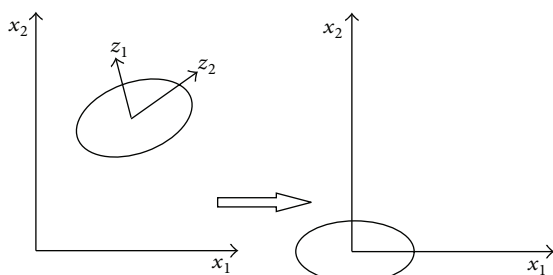


FIGURE 8: Diagram of data decomposition in PCA.

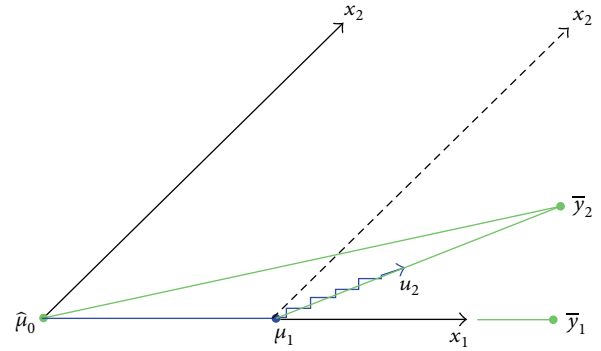


FIGURE 9: Schematic diagram of LAR: the LAR direction  $u_2$  makes an equal angle with  $x_1$  and  $x_2$ .

3.3.3. *Least Absolute Shrinkage and Selection Operator (LASSO)*. LASSO [16] is a new variable-choosing method created by Tibshirani in 1996 [17] on the basis of Bridge Regression by Frank and Friedman [18] and Nonnegative Garrote by Breiman [19]. The algorithm is summarized as follows. Suppose  $\beta$  is the coefficient of the model, the corresponding function is  $l(\beta)$ , and  $\beta$  is a  $d$ -dimensional vector. The equation of parameters penalty is. When  $l(\beta) = (y - X\beta)^2$ ,  $p_{\lambda_i}(|\beta_i|) = \lambda|\beta_i|^q$ , that is, the Bridge Regression. When  $q = 1$ , that is the LASSO (Figure 9 shows the details). In this study, we use LASSO with cross-validation to choose the regressors. It usually selects fewer numbers of the regressors and trade-offs between bias and mean squared error. So it may increase the accuracy of the model for coming new data.

3.3.4. *Comprehensive Comparison and Analysis*. We compare the results by the 3 methods. According to the result of PCA, the unsupervised learning, we can make sure if there is innate difference between the groups. The result of supervised learning will help the features selection to mine the most important features during the classification. Based on the cross-validation of 3 methods in data analysis, a reliable conclusion can be given in pulse diagnosis signals analysis.

## 4. Results

We have used 163 samples in this research (98 healthy volunteers, 38 patients with FLD, and 27 patients with cirrhosis). The results were based on the 193 parameters extracted from the signals in 3 places of each side of the radial artery called “cun,” “guan,” and “chi” in TCM by PCA, LS, and LASSO.

### 4.1. Signal Analysis between Healthy Volunteers and Patients with FLD

4.1.1. *Principal Component Analysis (PCA)*. By using the unsupervised learning without any information to guide the classification, we found that there is obvious difference between the pulse waves of healthy volunteers and patients with FLD. The accuracy to classify the signals of the two groups only by the 1st principal component is 83% (Figure 10 shows the details). The result suggests that it is feasible to

TABLE 1: Main parameters used in the LS model.

Main parameters	Parameters		
	C10 youguan	C2 youchi	C2 zuocun
Used times	20	15	14
Coefficient (importance)	1.12	1.35	1.28
Mean	-516.72	1412.28	1102.34
SD	55.10	162.67	98.03

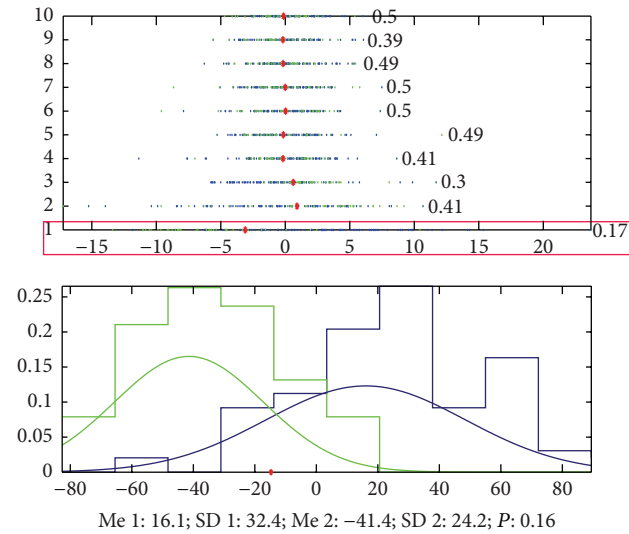


FIGURE 10: Classification by PCA. The accuracy that we cannot classify the signals of the 2 groups by the 1st principal component is 0.17, which means that the accuracy to classify this 2 groups is 83%. The blue points (patients with FLD) mainly appeared in the left side of the red point and the green ones (healthy volunteers) appeared in the right side. And the accuracy by 1st principal component and the 3rd principal component is 84%.

separate these two groups without any supervising due to physiological changes.

4.1.2. *Supervised Learning: LS and LASSO.* According to the character of clinical data, small samples with large dimensionality, we built a program to avoid the false classification in LS method: we test our method by 23 simulated samples to decide the upper limit number of selected regressors according to the number of samples. In this study, we can use 7 regressors at most.

(a) *Results by LS.* Doing the analysis by LS, we found that it is very easy to classify the two groups of signals and the accuracy is 91% by 7 parameters and 82% by only 2 parameters. The most important parameters we mined by a method we built named EFBL (Extended Forward Backward Least Square Regression) mainly appeared at zuocun, youguan, and youchi (Figure 11 and Table 1 show the details).

(b) *Results by LASSO.* Doing the analysis by LASSO, we found that there is obvious difference between the two groups of

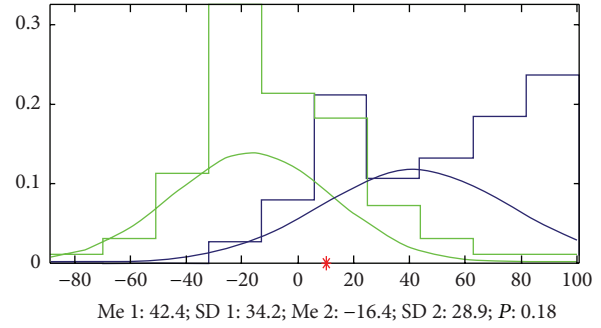


FIGURE 11: Classification by LS. The accuracy that we cannot classify the signals of the 2 groups by only 2 parameters is 0.18, which means that the accuracy to classify this 2 groups is 82%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones (patients with FLD) appeared in the right side.

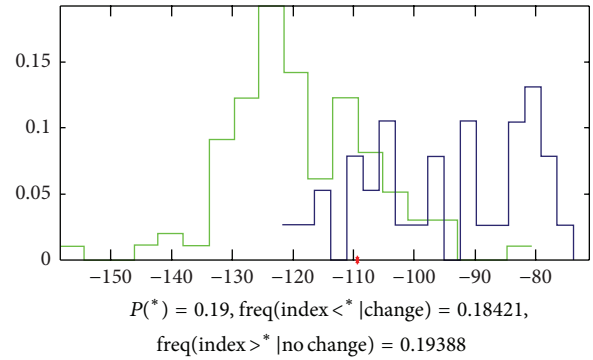


FIGURE 12: Classification by LASSO. The accuracy that we cannot classify the signals of the 2 groups by 3 parameters is 0.19, which means that the accuracy to classify this 2 groups is 81%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones (patients with FLD) appeared in the right side.

signals. The accuracy of classification is 81% by 3 parameters, and the equation of the model is as follows:

$$Y = -0.01C2\ zuocun - 0.11t4\ zuochi - 0.19F2\ youchi \quad (4)$$

(Figure 12 shows the details).

4.1.3. *Comprehensive Comparison and Analysis.* Comparing the results from the 3 methods, based on the combination of unsupervised learning and supervised learning, we can make a conclusion that there is obvious difference between the pulse signals of healthy volunteers and FLD patients, and the accuracy of classification is about 85%. The features we mined were mainly focused on zuocun, youguan, and youchi. In the theory of pulse diagnosis of TCM, youguan always represents the function of digestive. If some problem happened on digestive, doctors can feel the pulse in youguan changed. In this study, the result was partly matched with the theory of TCM. However, we need more data to confirm it.



TABLE 2: Main parameters used in the LS model.

Main parameters	Parameters		
	C1 zuoguan	C2 zuoguan	C9 zuochi
Used times	20	20	20
Coefficient (importance)	1.29	1.35	1.11
Mean	-1456.89	1417.54	647.32
SD	74.20	87.70	44.19

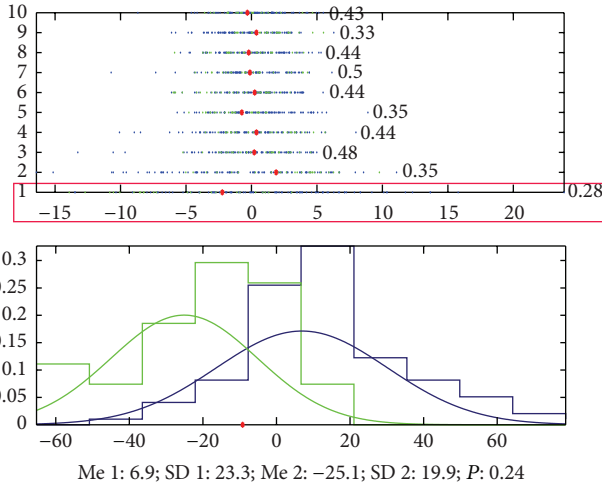


FIGURE 13: Classification by PCA. The accuracy that we cannot classify the signals of the 2 groups by the 1st principal component is 0.28, which means that the accuracy to classify this 2 groups is 72%, and the blue points (patients with cirrhosis) mainly appeared in the left side of the red point and the green ones (healthy volunteers) appeared in the right side. And the accuracy by 1st principal component and the 9th principal component is 76%.

4.2. Signal Analysis between Healthy Volunteers and Patients with Cirrhosis

4.2.1. Principal Component Analysis (PCA). By using the unsupervised learning without any information to guide the classification, we found that there is obvious difference between the pulse waves of healthy volunteers and patients with cirrhosis. The accuracy to classify the signals of the two groups only by the 1st principal component is 72% (Figure 13 shows the details).

4.2.2. Supervised Learning: LS and LASSO. In this study, we can use up to 7 regressors.

(a) Results by LS. Doing the analysis by LS, we found that it is very easy to classify the two groups of signals and the accuracy is 93% by 7 parameters and 84% by only 2 parameters. The most important parameters we mined by a method we built named EFBS (Extended Forward Backward Least Square Regression) mainly appeared at zuoguan, zuochi, and youguan (Figure 14 and Table 2 show the details).

(b) Results by LASSO. Doing the analysis by LASSO, we found that there is obvious difference between the two groups of

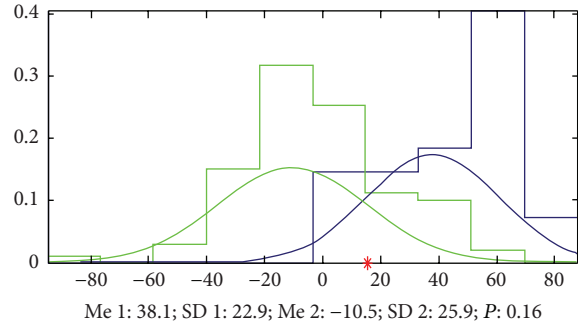


FIGURE 14: Classification by LS. The accuracy that we cannot classify the signals of the 2 groups by only 2 parameters is 0.16, which means that the accuracy to classify this 2 groups is 84%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones (patients with cirrhosis) appeared in the right side.

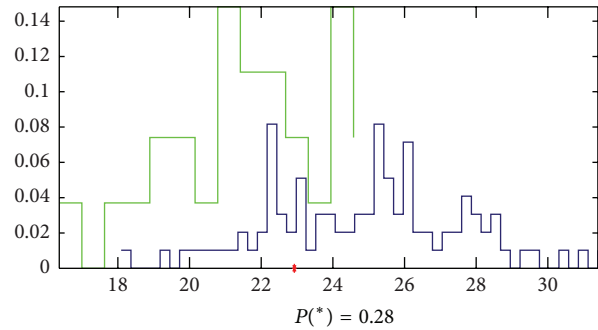


FIGURE 15: Classification by LASSO. The accuracy that we cannot classify the signals of the 2 groups by 2 parameters is 0.28, which means that the accuracy to classify this 2 groups is 72%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones (patients with cirrhosis) appeared in the right side.

signals. The accuracy to classify is 72% by 2 parameters and the equation of the model is as follows:

$$Y = -0.05C2 \text{ youguan} - 0.011F1 \text{ youchi} \quad (5)$$

(Figure 15 shows the details).

4.2.3. Comprehensive Comparison and Analysis. Comparing the results from the 3 methods, based on the combination of unsupervised learning and supervised learning, we can make a conclusion that there are differences between the pulse signals of healthy volunteers and cirrhosis patients, and the accuracy of classification is about 75%. The features we mined were mainly focused on zuoguan. In the theory of pulse diagnosis of TCM, zuoguan always represents the function of liver and gallbladder. If some problem happened on liver and gallbladder, doctors can feel the pulse in youguan changed. In this study, the result was matched with the theory of TCM.

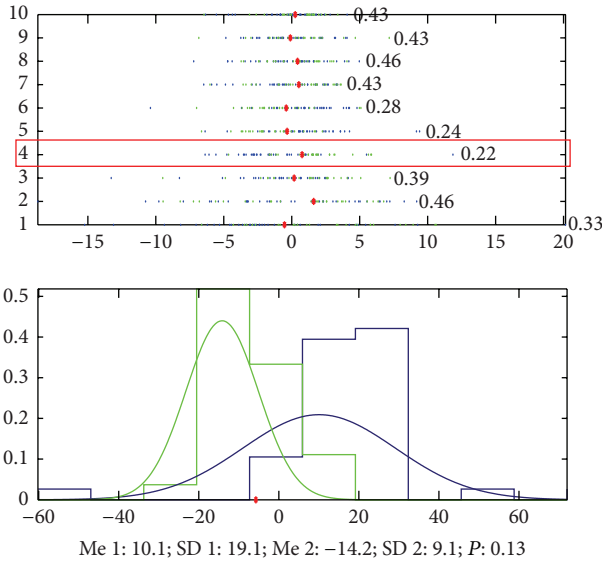


FIGURE 16: Classification by PCA. The accuracy that we cannot classify the signals of the 2 groups by the 1st principal component is 0.22, which means that the accuracy to classify this 2 groups is 78%, and the blue points (patients with cirrhosis) mainly appeared in the left side of the red point and the green ones (patients with FLD) appeared in the right side. And the accuracy by the 4th principal component and the 5th principal component is 87%.

4.3. Signal Analysis between Patients with FLD and Cirrhosis

4.3.1. Principal Component Analysis (PCA). By using the unsupervised learning without any information to guide the classification, we found that there is obvious difference between the pulse waves of healthy volunteers and patients with cirrhosis. The accuracy to classify the signals of the two groups only by the 4th principal component is 78% (Figure 16 shows the details).

4.3.2. Supervised Learning: LS and LASSO. In this study, we can use up to 7 regressors.

(a) Results by LS. Doing the analysis by LS, we found that it is very easy to classify the two groups of signals and the accuracy is 91% by 7 parameters and 73% by only 2 parameters. The most important parameters we mined by a method we built named EFBLs (Extended Forward Backward Least Square Regression) mainly appeared at zuocun, youguan, and youchi (Figure 17 and Table 3 show the details).

(b) Results by LASSO. Doing the analysis by LASSO, we found that there is obvious difference between the two groups of signals. The accuracy to classify is 72% by 3 parameters and the equation of the model is as follows:

$$Y = 0.02C1 \text{ youchi} - 0.025C5 \text{ youchi} - 0.05F8 \text{ youchi} \tag{6}$$

(Figure 18 shows the details).

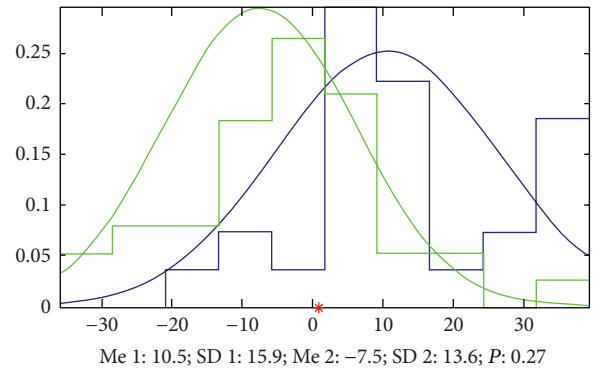


FIGURE 17: Classification by LS. The accuracy that we cannot classify the signals of the 2 groups by only 2 parameters is 0.27, which means that the accuracy to classify this 2 groups is 73%, and the green samples (patients with FLD) mainly appeared in the left side of the red point and the blue ones (patients with cirrhosis) appeared in the right side.

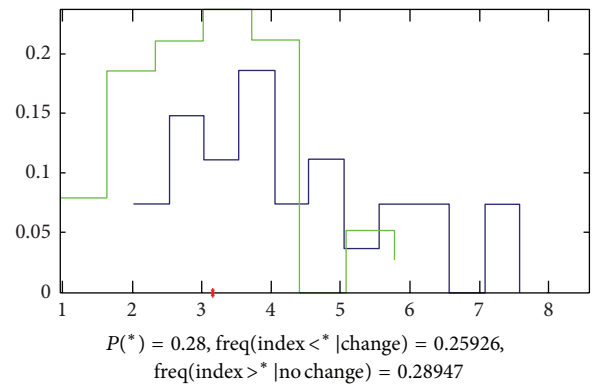


FIGURE 18: Classification by LASSO. The accuracy that we cannot classify the signals of the 2 groups by 2 parameters is 0.28, which means that the accuracy to classify this 2 groups is 72%, and the green samples (patients with FLD) mainly appeared in the left side of the red point and the blue ones (patients with cirrhosis) appeared in the right side.

4.3.3. Comprehensive Comparison and Analysis. Comparing the results from the cross-validation of the 3 methods, we can make a conclusion that there are differences between the pulse signals of FLD patients and cirrhosis patients, and the accuracy of classification is about 70%. The features we mined were mainly focused on youguan and youchi. This result is not simply matched with the theory of TCM mentioned in Section 4.2.3; zuoguan always represents the function of liver and gallbladder. As we know that FLD and cirrhosis are two liver diseases in different stage and the pathologic changes are much more serious in not only liver but also blood vessels and digestive system. So the features appear on other point instead of zuoguan.

5. Conclusions

There is a significant difference between the pulse diagnosis signals of healthy volunteers and patients with FLD and

TABLE 3: Main parameters used in the LS model.

Main parameters	Parameters				
	C3 youguan	C1 youchi	C4 youchi	F1 youchi	C9 zuocun
Used times	20	20	18	18	18
Coefficient (importance)	1.67	1.67	1.31	1.32	1.23
Mean	1218.55	-965.47	-708.04	2.42	366.40
SD	97.44	88.77	99.54	0.26	30.25

cirrhosis, and the result was confirmed by 3 analysis methods. The identification accuracy of the 1st principal component is about 75% without any classification formation by PCA, and supervised learning's accuracy (LS and LASSO) was even more than 93% when 7 parameters were used and 84% when only 2 parameters were used. From the results, we can have some conclusions.

(1) The machine learning method we built based on the combination of unsupervised learning, PCA, and supervised learning, LS and LASSO, is feasible in analyzing the pulse diagnosis signals. Moreover, according to the result of cross-reference by 3 methods and the equation established by LASSO, we can achieve a reliable result by signals of pulse diagnosis in TCM to identify the healthy volunteers and the patients. This method can help offer some objective data to prove the important role of pulse diagnosis in TCM.

(2) The features we mined by LS and LASSO to classify the healthy volunteers and patients with FLD and cirrhosis appear in specific places we called "cun," "guan," and "chi" when feeling the pulse. For example, the features to classify the healthy volunteers and patients with FLD mainly appear in zuocun, youguan, and youchi. The features to classify the healthy volunteers and patients with cirrhosis mainly appear in zuoguan and zuochi. However, youguan and youchi are the main place where features appear between the patients of FLD and cirrhosis patients. This result is similar to the theory of pulse diagnosis in Traditional Chinese Medicine (TCM) which can support the modern research of pulse diagnosis. But more research is needed to confirm the conclusion.

(3) FLD and cirrhosis are the most common liver diseases, and there are high incidences among middle-aged men. These two kinds of diseases have not only common points in the pathology but also different points. Pulse diagnosis of TCM can diagnose disease by feeling the pulse of radial artery, but we have no evidence to prove that. In this study, according to the result, we can find that from the pulse wave collected in different places called "cun," "guan," and "chi" in radial artery humans in different health condition can be classified with high accuracy even when the diseases affected the same organ, liver. From the results, we can offer some important evidence for the science of pulse diagnosis in TCM clinical diagnosis. Of course, we need more data to confirm it.

In a word, the machine learning method we built in this study based on the combination of unsupervised learning, PCA, and supervised learning, LS and LASSO, might offer some confidence for the realization of computer-aided diagnosis by pulse diagnosis in TCM. In addition, this study

might offer some important evidence for the science of pulse diagnosis in TCM clinical diagnosis.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Wang Nanyue wrote this paper and also did most of the work concerning data acquisition and some work of analysis; Yu Youhua designed the whole research and helped in drawing the conclusion; Huang Dawei developed all the software of data analysis and did some work of analysis; Shan Zengyu adjusted the instrument for data collection; Liu Jia, Li Tongda, Xue Liyuan, Chen Yanping, and Wang Jia did work of data acquisition in clinic and input the data from papers into computers. All authors read and approved the final paper.

## Acknowledgments

This study was supported by National 11th Five-Year Plan Project (no. 2006BAI08B01-02) and Fundamental Science Research Project of CACMS (no. ZZ2010009). Some work was supported by the key project of MOST basic work 2013 (no. 2013FY114400). Thanks are due to Dr. Lv Wenliang for the help in diagnosis of cirrhosis patients in Guang'anmen Hospital of China Academy of Chinese Medical Sciences and are also due to Dr. Li caifen for the help in diagnosis of FLD patients in Acupuncture Hospital of China Academy of Chinese Medical Sciences.

## References

- [1] G.-Z. Li, S. Sun, M. You, Y.-L. Wang, and G.-P. Liu, "Inquiry diagnosis of coronary heart disease in Chinese medicine based on symptom-syndrome interactions," *Chinese Medicine*, vol. 7, article 9, 2012.
- [2] H. Shao, G.-Z. Li, G. Liu, and Y. Wang, "Symptom selection for multi-label data of inquiry diagnosis in traditional Chinese medicine," *Science China Information Sciences*, vol. 56, no. 5, Article ID 052118, pp. 1-13, 2013.
- [3] F. Li, C. Zhao, Z. Xia, Y. Wang, X. Zhou, and G.-Z. Li, "Computer-assisted lip diagnosis on traditional Chinese medicine using multi-class support vector machines," *BMC Complementary and Alternative Medicine*, vol. 12, article 127, 2012.



- [4] N.-Y. Wang, Y.-H. Yu, D.-W. Huang et al., "Research of features in the pulse waves of women during pregnancy," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '10)*, pp. 730–732, IEEE, Hong Kong, December 2010.
- [5] N.-Y. Wang, Y.-H. Yu, D.-W. Huang et al., "Research of pulse wave's change of rats after kidney resection," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 24, no. 6, pp. 796–798, 2009.
- [6] N.-Y. Wang, Y.-H. Yu, D.-W. Huang et al., "Research of pulse wave's features about rats of CHF model," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 12, pp. 1953–1955, 2010.
- [7] L.-Y. Xue, N.-Y. Wang, Y.-H. Yu et al., "Research of features in the pulse waves of HBP patients based on principal component analysis and LS,Lasso," *Chinese Journal of Basic Medicine in Traditional Chinese Medicine*, vol. 6, pp. 660–663, 2013.
- [8] K. Omagari, Y. Kadokawa, J.-I. Masuda et al., "Fatty liver in non-alcoholic non-overweight Japanese adults: incidence and clinical characteristics," *Journal of Gastroenterology and Hepatology*, vol. 17, no. 10, pp. 1098–1105, 2002.
- [9] M. Hilden, P. Christoffersen, E. Juhl, and J. B. Dalgaard, "Liver histology in a 'normal' population—examinations of 503 consecutive fatal traffic casualties," *Scandinavian Journal of Gastroenterology*, vol. 12, no. 5, pp. 593–597, 1977.
- [10] L. Shen, J.-G. Fan, Y. Shao et al., "Prevalence of nonalcoholic fatty liver among administrative officers in Shanghai: an epidemiological survey," *World Journal of Gastroenterology*, vol. 9, no. 5, pp. 1106–1110, 2003.
- [11] B.-L. Lou, E.-L. Xue, N. Fu et al., "Incidence of non-alcoholic fatty liver disease and its risk factors in navy flight crew and the submariners," *Journal of Preventive Medicine of Chinese People's Liberation Army*, vol. 1, pp. 20–22, 2013.
- [12] W. A. Zatoński, U. Sulkowska, M. Mańczuk et al., "Liver cirrhosis mortality in Europe, with special attention to Central and Eastern Europe," *European Addiction Research*, vol. 16, no. 4, pp. 193–201, 2010.
- [13] Z.-F. Fei, *Chinese Pulse Diagnosis Research*, Shanghai Publishing Company of Traditional Chinese Medicine, 1991.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, World Publishing Company, New York, NY, USA, 2009.
- [15] G.-Z. Li, H.-L. Bu, M. Q. Yang, X.-Q. Zeng, and J. Y. Yang, "Selecting subsets of newly extracted features from PCA and PLS in microarray data analysis," *BMC Genomics*, vol. 9, supplement 2, article S24, 2008.
- [16] G. Jianchao, *Application of Lasso and related methods in generalized linear models selection [M.S. thesis]*, Central South University, Changsha, China, 2001.
- [17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [18] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.
- [19] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

