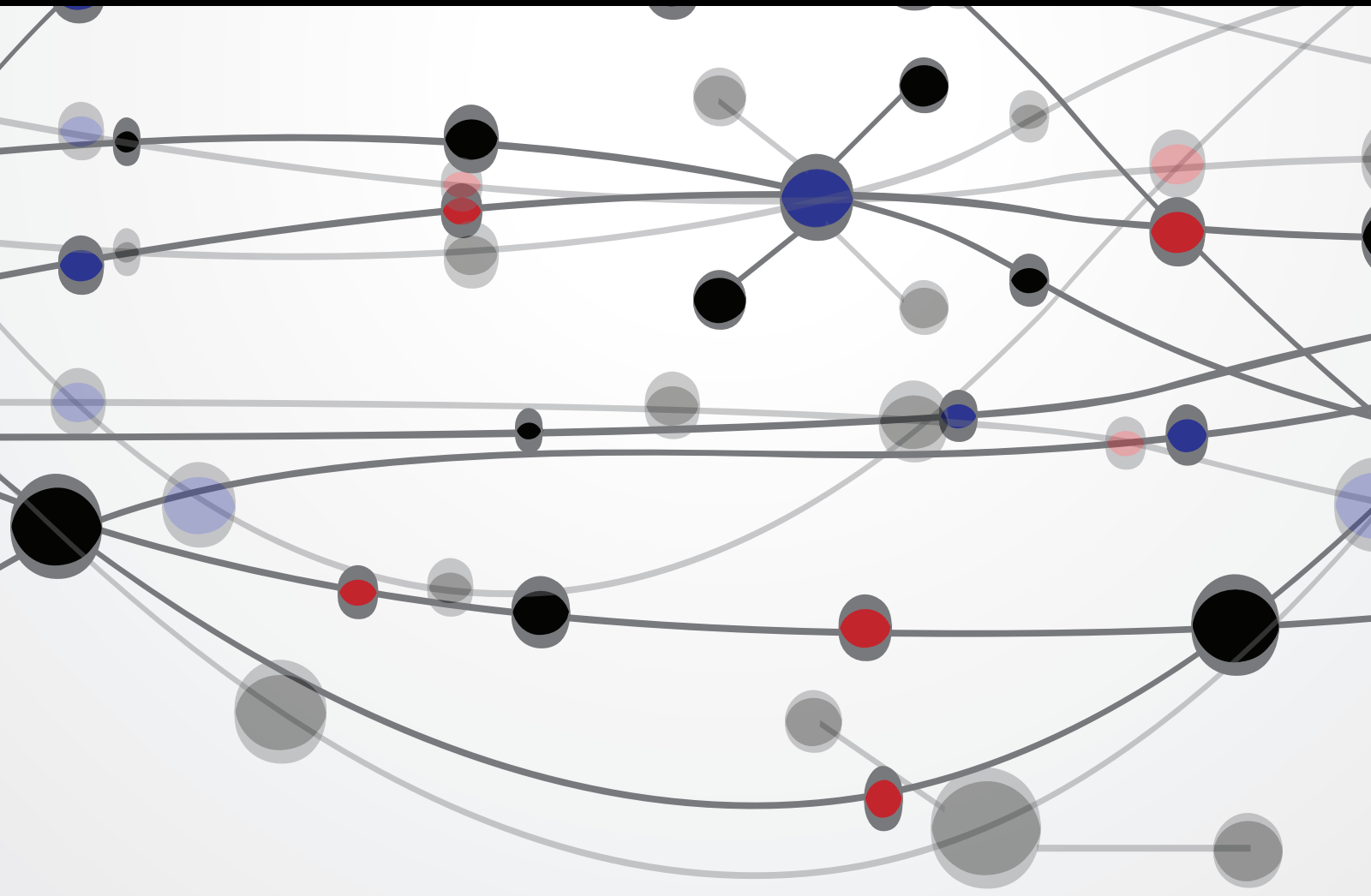
The Scientific World Journal

Bioinformatics/Medical Informatics in Traditional Medicine and Integrative Medicine

Guest Editors: Zhaohui Liang, Xiangji Huang, Byeongsang Oh, and Josiah Poon



Bioinformatics/Medical Informatics in Traditional Medicine and Integrative Medicine

The Scientific World Journal

Bioinformatics/Medical Informatics in Traditional Medicine and Integrative Medicine

Guest Editors: Zhaohui Liang, Xiangji Huang, Byeongsang Oh, and Josiah Poon

Contents

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 460490, 2 pages
http://dx.doi.org/10.1155/2015/460490

Editorial

Bioinformatics/Medical Informatics in Traditional Medicine and Integrative Medicine

Zhaohui Liang,[1] Xiangji Huang,[2,3] Byeongsang Oh,[4] and Josiah Poon[5]

[1] *York University, Toronto, ON, Canada*
[2] *School of Computer Science, Central China Normal University, Wuhan, China*
[3] *School of Information Technology, York University, Toronto, ON, Canada*
[4] *Sydney Medical School, The University of Sydney, Sydney, Australia*
[5] *School of Information Technologies, Faculty of Engineering and Information Technologies, The University of Sydney, Sydney, Australia*

Correspondence should be addressed to Zhaohui Liang; stan79@yorku.ca

Received 27 September 2015; Accepted 13 October 2015

Traditional Chinese Medicine (TCM) and integrative medicine are key components of the cultural heritage from Eastern Asia with thousands-of-years history in research and healthcare delivery. Traditional oriental medicine contributes significantly to the prosperity of Chinese and Eastern Asian culture. After the introduction of western biomedicine to Asia, traditional medicine still plays an important role in the healthcare system of many Asian countries and integrated with the mainstream medical treatments as a new track of healthcare named as integrative medicine. With the current trend of globalization, traditional medicine and integrative medicine are receiving gradual acceptance in the Western world. As a result, studies on traditional medicine attract more and more attention from researchers with various knowledge backgrounds and technologies.

Medical informatics is a new interdisciplinary branch in medical science when computer science and information technology are combined with research of health science. The application of medical informatics that has extended to the studies of traditional medicine and other therapies of complementary and alternative medicine (CAM).

The special issue supported by this journal provides a forum for traditional and integrative medical researchers and practitioners to share and exchange their new ideas on using computer science and information technology to explore and solve problems in healthcare. It is proposed with the Fifth International Workshop on Information Technology for Chinese Medicine (ITCM 2014) in Guangzhou, China, on 12 to 14 December 2014. The workshop is in conjunction with the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM'14), which was held in Belfast, UK, on 2 to 5 November 2014. Professor Xusheng Liu, Professor Honglai Zhang, and Professor Guozheng Li cochaired the workshop. The conference invited top experts from the US, UK, Australia, and Hong Kong to present their inspiring research outcomes and prospect the future of traditional and integrative medicine. However, numerous scientists and researchers were unable to introduce their excellent idea due to time limit of the workshop.

The ITCM 2014 received about 100 submissions. All papers were anonymously reviewed by members of the IEEE conference organization committee. The accepted papers were published in the Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine Workshops (IEEE-BIBMW) (ISBN 978-1-4799-1309-1). Just a few excellent papers were later invited to submit the extension version to the special issue alongside external submissions for consideration of publishing. This special issue has received 37 submissions. All papers have gone through rigorous view, and only 10 of them (27%) are finally accepted for publication.

This special issue reflects the up-to-date progress in applications of information technology to traditional and integrative medicine. The papers are categorized to represent the four aspects of medical informatics research of

the discipline. In the paper entitled "Standardization of Syndrome Differentiation Defined by Traditional Chinese Medicine in Operative Breast Cancer: A Modified Delphi Study," Q. Guo and Q. Chen present their research on TCM syndromes. Five papers are selected to demonstrate the research progress in disease diagnosis and treatment. G.-X. Shi et al. report a clinical study on vascular dementia. Z. Chen presents a new mathematics method to explore the classical theory of five elements in TCM in his work "Researches on Mathematical Relationship of Five Elements of Containing Notes and Fibonacci Sequence Modulo 5." In "Syndrome Differentiation Analysis on MARS500 Data of Traditional Chinese Medicine," Y.-Z. Li et al. succeed to use MARS500 to process the data of traditional medicine. The paper entitled "Detecting Disease in Radiographs with Intuitive Confidence" by S. Jaeger introduces the new idea to use informatics method to detect disease. Three papers are about information processing of traditional medicine. The paper entitled "Patterns Exploration on Patterns of Empirical Herbal Formula of Chinese Medicine by Association Rules" by L. Huang et al. used association rules to retrieve patterns from classical traditional medical formula. B. Zhang et al. proposed a bioinformatics approach to explore the latent patterns from conventional formula Shuang-Huang-Lian in their work "Using Bioinformatics Approach to Explore the Pharmacological Mechanisms of Multiple Ingredients in *Shuang-Huang-Lian*." The paper entitled "Pulse-Diagnosis Signals Analysis of Fatty Liver Disease and Cirrhosis Patients by Using Machine Learning" by N. Wang et al. introduces new data mining method to process diagnostic data of liver disease. Finally, the paper entitled "An Ensemble Learning Based Framework for Traditional Chinese Medicine Data Analysis with ICD-10 Labels" by G. Zhang et al. and the paper entitled "ISMAC: An Intelligent System for Customized Clinical Case Management and Analysis" introduce the applications of machine learning to electronic data analysis of traditional medicine.

Acknowledgments

The editors would like to extend their gratitude to the authors for their contribution to this special issue, especially to the corresponding authors, the Chair of the steering committee of BIBM, Professor Xiaohua Tony Hu, and the Program Chairs of ITCM: Professor Guozheng Li, Professor Xusheng Liu, and Professor Honglai Zhang. The research and editorial work is supported by Natural Science Foundation of China (Grants nos. 81573827, 81373883, and 81274003) and by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Zhaohui Liang
Xiangji Huang
Byeongsang Oh
Josiah Poon

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 859192, 9 pages
http://dx.doi.org/10.1155/2015/859192

Research Article

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

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

[1]*Experimental Research Center, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[2]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[3]*GULOU Hospital of TCM of Beijing, Beijing 100009, China*
[4]*Tai Yang Gong Health Care Center, Chaoyang District, Beijing 100102, China*
[5]*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

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.

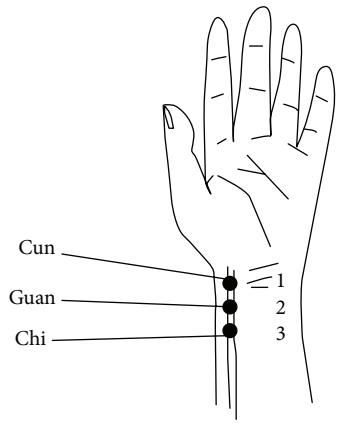


Figure 1: Places of "cun," "guan," and "chi" in the left hand.

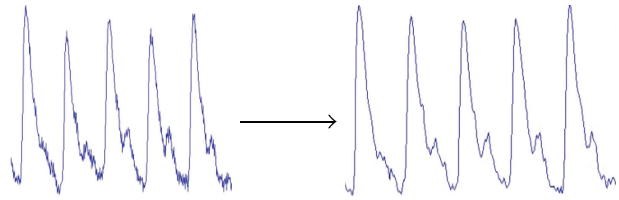


Figure 2: Pretreatment result: the figure on the left side is the original signal, and the figure on the right side is the signal after pretreatment; the main characters remain unchanged but the noise effectively reduced.

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 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.

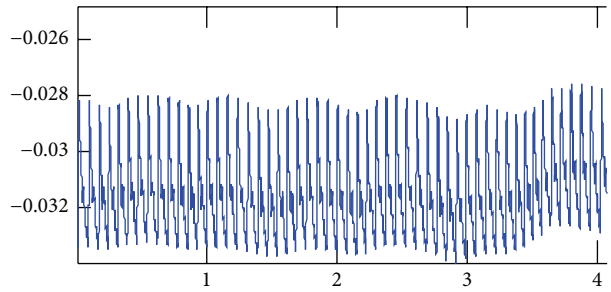


Figure 3: Original signals: the coordinate values of the start point of the signals regularly change because of the human breath.

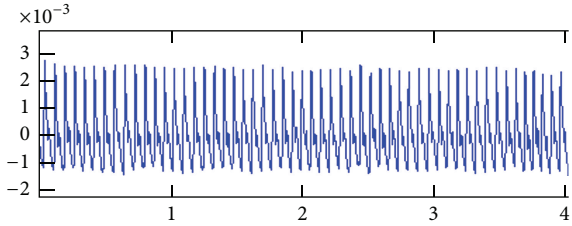


Figure 4: Pretreated signals: the coordinate values of the start point of the signals had been adjusted to zero, without any influence on the amplitude and other characters.

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

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

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 × 6 places + 1.

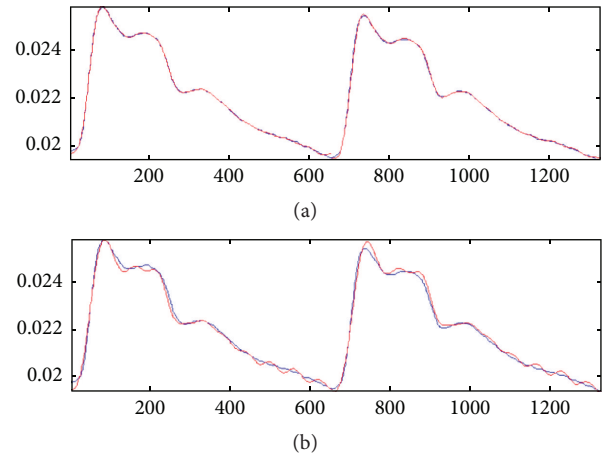


Figure 5: Reduction of error by recursive total least squares: the red signals are original signals and the blue ones are fitting signals. The upper figure shows the fitting effect of the two cycles by separate fitting, and the lower one shows the effect by normal fit. The separate fitting can get a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

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

(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 (EFBLS) [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.

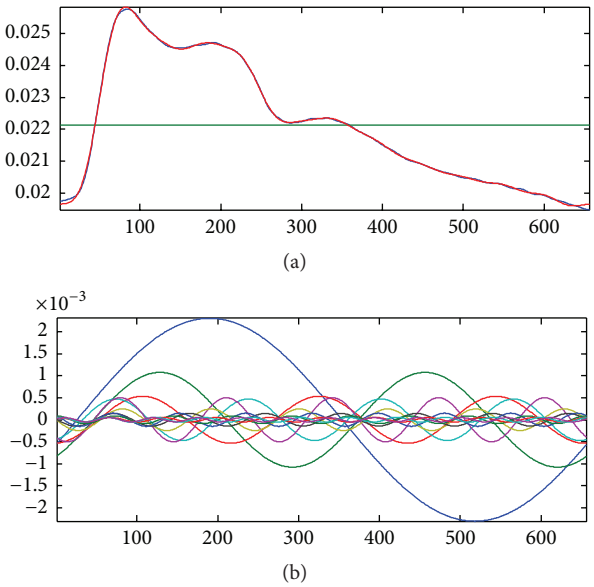


Figure 6: Harmonic fitting: the upper figure shows the signal of one cycle; the red one is the original signal and the blue one is the fitting result by the 12 harmonics shown in the lower figure. The fitting effect is optimal.

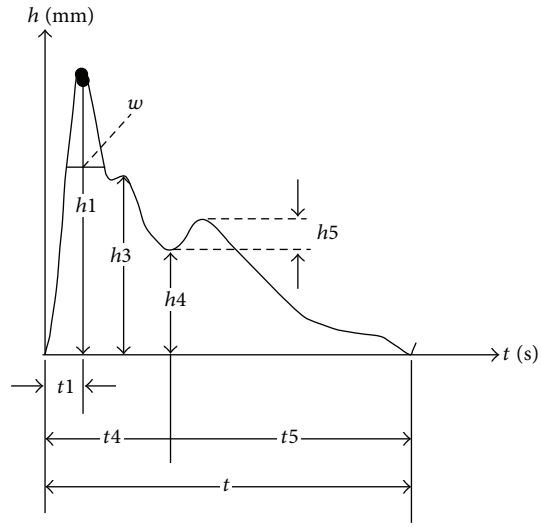


Figure 7: Time domain parameters.

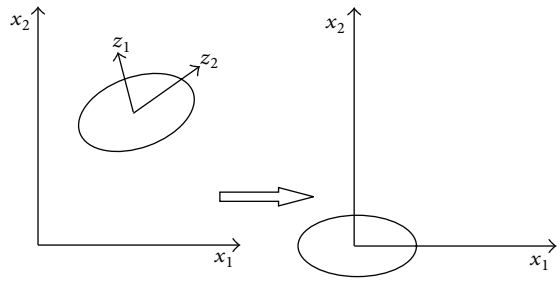


Figure 8: Diagram of data decomposition in PCA.

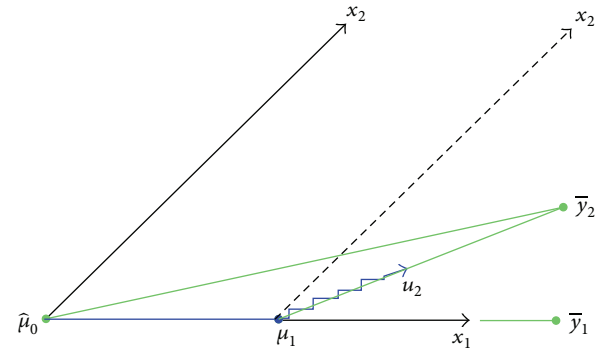


Figure 9: Schematic diagram of LAR: the LAR direction u_2 makes an equal angle with x_1 and x_2.

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 β is the coefficient of the model, the corresponding function is $l(\beta)$, and β 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		
	*C*10 youguan	*C*2 youchi	*C*2 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

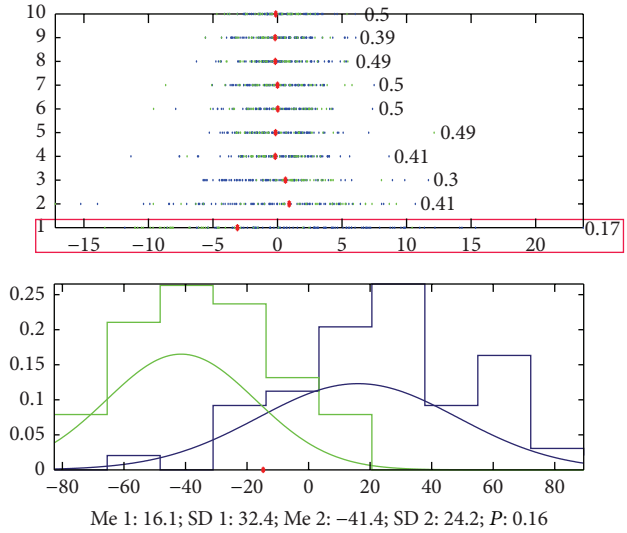


Figure 10: Classification by PCA. The accuracy that we cannot classify the signals of the 2 groups by the 1st principal component is 0.17, which means that the accuracy to classify this 2 groups is 83%. The blue points (patients with FLD) mainly appeared in the left side of the red point and the green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

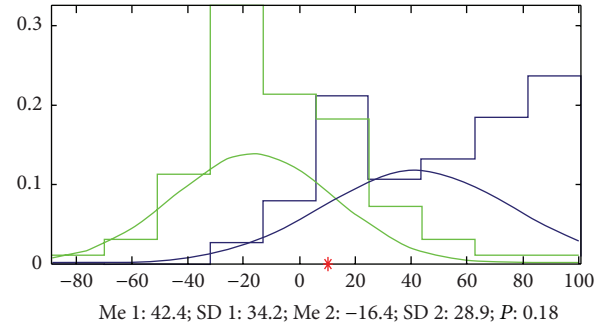


Figure 11: Classification by LS. The accuracy that we cannot classify the signals of the 2 groups by only 2 parameters is 0.18, which means that the accuracy to classify this 2 groups is 82%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones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

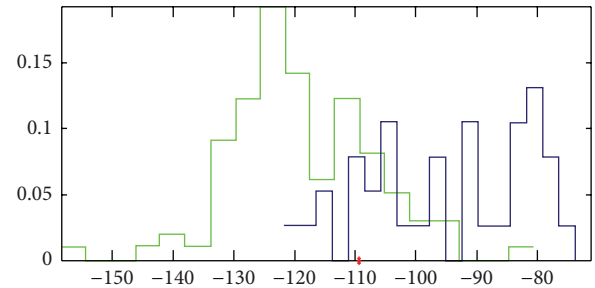


Figure 12: Classification by LASSO. The accuracy that we cannot classify the signals of the 2 groups by 3 parameters is 0.19, which means that the accuracy to classify this 2 groups is 81%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones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

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

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

(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		
	*C*1 zuoguan	*C*2 zuoguan	*C*9 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

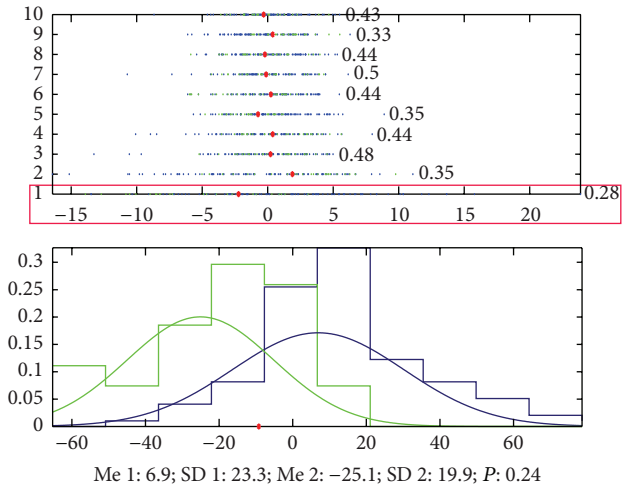


Figure 13: Classification by PCA. The accuracy that we cannot classify the signals of the 2 groups by the 1st principal component is 0.28, which means that the accuracy to classify this 2 groups is 72%, and the blue points (patients with cirrhosis) mainly appeared in the left side of the red point and 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 EFBLS (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 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.

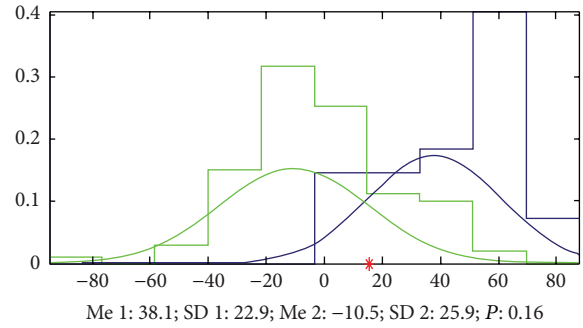


Figure 14: Classification by LS. The accuracy that we cannot classify the signals of the 2 groups by only 2 parameters is 0.16, which means that the accuracy to classify this 2 groups is 84%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones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

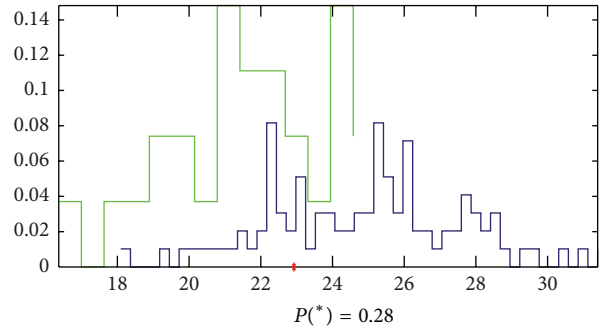
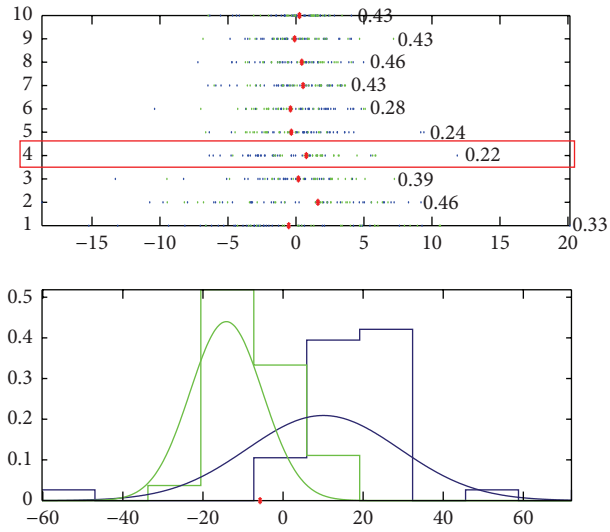


Figure 15: Classification by LASSO. The accuracy that we cannot classify the signals of the 2 groups by 2 parameters is 0.28, which means that the accuracy to classify this 2 groups is 72%, and the green samples (healthy volunteers) mainly appeared in the left side of the red point and the blue ones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



Me 1: 10.1; SD 1: 19.1; Me 2: −14.2; SD 2: 9.1; P: 0.13

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).

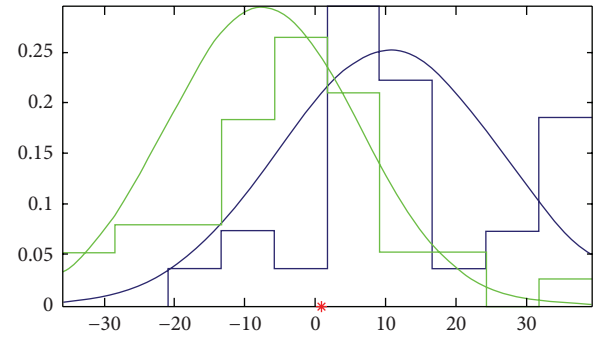
4.3.2. Supervised Learning: LS and LASSO. In this study, we can use up to 7 regressors.

(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:

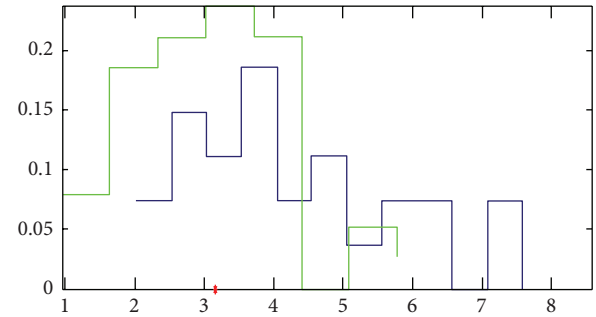
$$Y = 0.02C1 \text{ youchi} - 0.025C5 \text{ youchi} - 0.05F8 \text{ youchi} \quad (6)$$

(Figure 18 shows the details).



Me 1: 10.5; SD 1: 15.9; Me 2: −7.5; SD 2: 13.6; P: 0.27

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.



$P(^*) = 0.28$, freq(index $<^*$ |change) = 0.25926,
freq(index $>^*$ |no change) = 0.28947

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				
	*C*3 youguan	*C*1 youchi	*C*4 youchi	*F*1 youchi	*C*9 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 diseaseandits 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 Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 507925, 8 pages
http://dx.doi.org/10.1155/2015/507925

Research Article

An Ensemble Learning Based Framework for Traditional Chinese Medicine Data Analysis with ICD-10 Labels

Gang Zhang,[1] Yonghui Huang,[1] Ling Zhong,[1] Shanxing Ou,[2] Yi Zhang,[3] and Ziping Li[4]

[1]*School of Automation, Guangdong University of Technology, Guangzhou 510006, China*
[2]*Department of Radiology, Guangzhou General Hospital of Guangzhou Military Command, Guangzhou 510010, China*
[3]*Department of Plastic and Reconstructive Surgery, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou 510080, China*
[4]*The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510120, China*

Correspondence should be addressed to Ziping Li; lzip_008@163.com

Received 21 January 2015; Accepted 25 June 2015

Academic Editor: Josiah Poon

Objective. This study aims to establish a model to analyze clinical experience of TCM veteran doctors. We propose an ensemble learning based framework to analyze clinical records with ICD-10 labels information for effective diagnosis and acupoints recommendation. *Methods.* We propose an ensemble learning framework for the analysis task. A set of base learners composed of decision tree (DT) and support vector machine (SVM) are trained by bootstrapping the training dataset. The base learners are sorted by accuracy and diversity through nondominated sort (NDS) algorithm and combined through a deep ensemble learning strategy. *Results.* We evaluate the proposed method with comparison to two currently successful methods on a clinical diagnosis dataset with manually labeled ICD-10 information. ICD-10 label annotation and acupoints recommendation are evaluated for three methods. The proposed method achieves an accuracy rate of 88.2% ± 2.8% measured by zero-one loss for the first evaluation session and 79.6% ± 3.6% measured by Hamming loss, which are superior to the other two methods. *Conclusion.* The proposed ensemble model can effectively model the implied knowledge and experience in historic clinical data records. The computational cost of training a set of base learners is relatively low.

1. Introduction

In the study of Traditional Chinese Medicine (TCM), clinical experience of veteran doctors plays an important role in both theoretical research and clinical research [1]. The clinical experience is often recorded in a semistructural or unstructured manner, since most of them have a relatively long history. Some of them are manually organized in simple categories or even in plain text. In data mining and machine learning applications, structural inputs are required for computational models [2]. However, there is valuable knowledge in these clinical experience records; for example, they can be used for classification or association rule mining to find patterns of disease diagnosis and Chinese medical ZHENG diagnosis, or for identification of core elements of ZHENG, the relation between herbal medicine formula and different ZHENG and disease, and the common law of clinical diagnosis [3, 4].

There are at least three challenges in building the computational model for analysis clinical records of veteran TCM doctors. The first is that the target data record set for analysis is multimodal with many correlated factors, which means that the data samples are not generated from a single model, but several unknown models or their combination. Hence a simple parameter model cannot capture the generative laws of such data [5, 6]. The second is that the prior knowledge from TCM theory and clinical treatment is available, and they are totally informally organized and even ambiguous, which cannot be directly used in building analysis models. The third is that the data is unstructured, which means that effective feature representations are often unavailable [7].

Currently there some studies on TCM data analysis with machine learning models. We briefly review some work closely related to this work. Di et al. [8] proposed a clinical outcome evaluation model based on local learning for the efficacy of acupuncture neck pain caused by cervical

spondylosis. They introduced a local learning method, by defining a distance function between treatment records of each patient. When evaluating the efficacy of acupuncture for a patient, the model selects p samples most close to the test sample. The model significantly reduces the computational cost when the dataset is large. However, their model requires a structural input and cannot process data stored in plain text. Liang et al. [9] proposed a multiview KNN method for subjective data of TCM acupuncture treatment to evaluate the therapeutic effect of neck pain. They regard the clinical records as data samples with multiple view, each of which refers to a subset of attributes. And different views are disjointed from each other. The model fully makes use of information from different views. A boosting-style method is used to combine models associated with different views together. Zhang et al. [10, 11] proposed a kernel decision tree method for TCM data analysis. Their model processes data in a feature space induced by a kernel function, which is effective for the multimodal data. However, the prior knowledge cannot be explicitly expressed in the feature space, which limits its further application.

To tackle the aforementioned challenges, in this paper, we propose to adopt the recently proposed deep ensemble learning method to build our analysis model. Deep ensemble learning is an extension of ensemble learning, which is a famous topic in machine learning research [12–14]. Ensemble learning makes a weighted combination of a set of base learners to form a combined learner as the final model. Equation (1) shows the general form of ensemble of base learners:

$$h_{\text{ens}}(x) = \sum_{i=1}^{m} w_i \cdot h_i(x), \tag{1}$$

where h_i is a set of base learners of at least some difference and w is a weight vector with constraints $\sum_i w_i = 1,\ w_i \geq 0$. To avoid the overfitting problem of the ensemble learner h_{ens}, a regularization prior should be imposed on w [15]. A common regularization prior is the sparsity of w, meaning that more 0 in w is preferable. Or one can impose a normal distribution on w.

The quality of the set of base learners and w fully controls the performance of the ensemble learner [16]. There are three methods to determine the best ensemble of a set of base learners [17]. The first is the selective ensemble, which selects small parts of base learners by some criteria and combines them using a majority voting strategy. This kind of method in fact imposes a prior on w that only a small number of elements in w can be nonzero, as well as the equal weight for each remaining learner. The second method finds the optimal w through solving an optimization as follows:

$$\min_{w} \text{Loss}\left(\sum_{i} w_i \cdot h_i, D\right) + \Omega(w). \tag{2}$$

This kind of method finds the optimal w such that the ensemble achieves the minimal loss and the best regularization on the evaluation set parameterized by w. Since the optimization problem is not convex for most loss evaluation functions, it may not be solved analytically. The third method is an iterative method that initializes the weights randomly and adjusts them through a iterative procedure. The famous Adaboost algorithm falls into this kind [18]. The Adaboost algorithm adopts very simple principle when finding the optimal weights; that is, if a candidate base learner has a good performance on the training dataset and is different from others, its weight can be increased by the algorithm. The idea of Adaboost is to find a subset of base learners of high quality whose diversity is also high [19].

However, the above three methods do not fully meet the requirement of the problem of TCM data analysis. The concept class implied in our dataset is of complex structure, or in another word, its VC dimension is extremely large, leading to a complex class boundary. Simple or shallow function classes may suffer from lack of representation capability. Motivated by the current research process of ensemble learning and deep learning, we propose to use deep ensemble learning for our analysis task. Different from classical ensemble learning, deep ensemble learning tries to tackle the problem of multimodal analysis and extends the bound of generalization ability of the ensemble learner. A key advantage of deep ensemble learning is that deep models can be used as base learners, which extends the representation capability to a great extent [20, 21].

Figure 1 shows the main idea of this paper, as well as an example of the clinical data to be analyzed.

Deep ensemble learning method adopts a capacity-conscious criterion to evaluate the quality of base learners. Different from the famous accuracy-diversity selective ensemble framework, the deep ensemble learning methods try to directly minimize the error bound according to the current training dataset.

The remainder of this paper is organized as follows. In Section 2 we present the main methods, including the self-adaptive region cutting method, stacked autoencoder training algorithm, and MIML model. In Section 3 we present the settings of evaluation of the proposed method and report the evaluation results on a real clinical dataset at different multiple-label classification criteria. And finally we conclude the paper in Section 4.

2. Deep Ensemble Learning

2.1. Problem Definition. Before going further, we formally define the problem to be solved. Let $D = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\}$ be a set of TCM clinical records and the corresponding ICD-10 labels, where $x_i \in X \subseteq 0,\ 1^d$ is the representation of each data sample in D. Each element of x_i is denoted as x_{ij}, indicating whether an acupoint or ZHENG is included in the treatment plan. The acupoint and ZHENG information are extracted through a simple key word matching procedure. $y_i \in L$ is an ICD-10 label associated with the ith record. $z_i \in Z$ is the TCM diagnosis of the ith clinical sample. When an acupoint or ZHENG is found, the correspond element in x_i is set to 1, and 0 otherwise. The goal is to find a function $h : X \to L \times Z$ that achieves minimal loss on a training dataset D_{train}, given a predefined loss function, for example, zero-one loss.

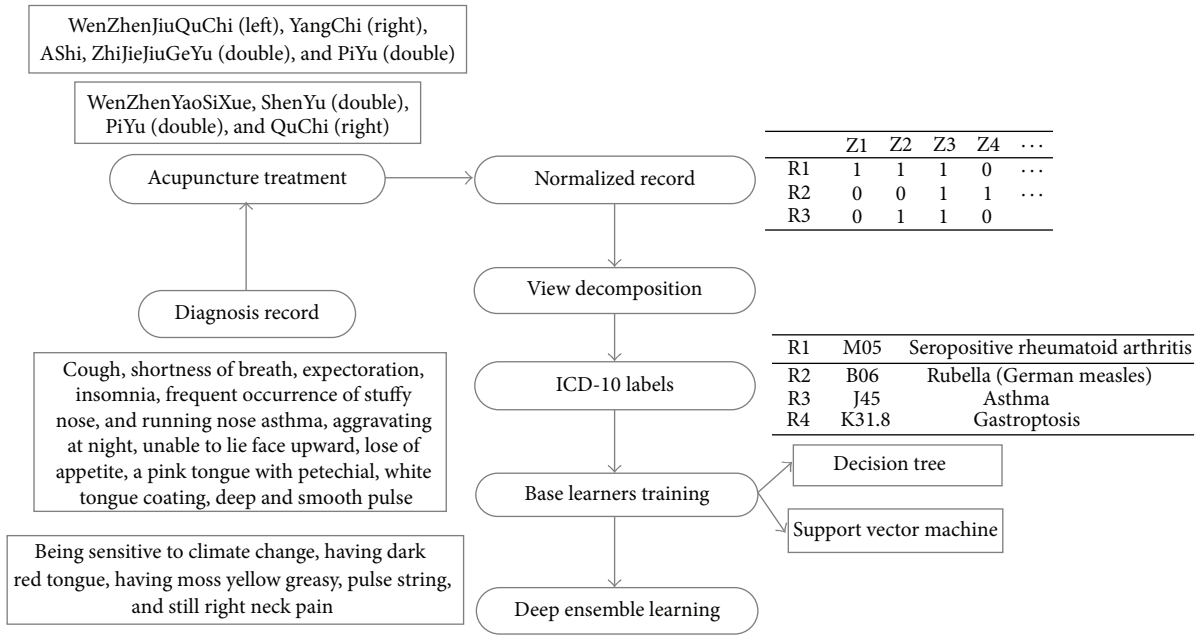


Figure 1: The main idea of this paper.

2.2. Selective Ensemble and Learners Sorting. Selective ensemble is an ensemble strategy that sorts the base learners with some criteria and then selects the learners at the top of the list to ensemble. Three criteria are used in this study. The first is accuracy, which evaluates how the model output matches the ground truth label [22]. Since the output of h is a pair of labels, that is, $h(x_i) = (y_i, z_i)$, the simple zero-one loss is not suitable in this case. We define a new accuracy as follows:

$$\mathrm{Acc}_h(D) = \frac{1}{|D|}\sum_{i=1}^{|D|}\alpha\cdot\delta\left(h(x_i)_y, y_i\right) + \beta \cdot\delta\left(h(x_i)_z, z_i\right) + \gamma\cdot\Delta\left(h(x_i), y_i, z_i\right), \quad (3)$$

where α, β, and γ are parameters controlling the importance of TCM diagnosis, ICD-10, and both. The indicator function $\delta(x, y) = 1$ if $x = y$ and 0 otherwise. Δ is also an indicator function that evaluates two tuples.

The second criterion is diversity which evaluates the difference between base learners. According to the theory of ensemble learning, an ensemble of learners that are different from each other may achieve better performance. There are some diversity definitions proposed in the literature of ensemble learning [23]. A simple way is to compare the results of each learner on the whole evaluation dataset. In this study, there are two target variables for prediction and the diversity for a learner h given that a dataset D is defined as follows:

$$\mathrm{Div}_h = \frac{\sum_{i=1}^{|D|}\left(h(x_i) - h_{\mathrm{ens}}(x_i)\right)^2\cdot\phi\left(h(x_i), (y_i, z_i)\right)}{|D|^2}; \quad (4)$$

since $h(x_i)$ returns a tuple, in the definition we use Hamming distance when evaluating the difference between two tuples. $\phi(\cdot,\cdot)$ is an indicator function in which $\phi(h(x_i), y_i) = 1$ if $h(x_i) = y_i$, and 0 otherwise. h_{ens} stands for the ensemble learner of majority voting. The intuition of this definition is that if the output of a learner h is away from that of the ensemble learner h_{ens}, it is assigned with large diversity [24].

To this end, we are able to sort all learners by both their accuracy and diversity. We use a sorting strategy named nondominated sort (NDS) to get a reasonable sorting. The rule NDS is that if the accuracy and diversity of h_i can dominate those of h_j, h_i should be ahead of h_j in the queue. When the accuracy and diversity of h_i and h_j cannot dominate each other, we add the rank of accuracy and diversity to form a single rank r. And the learner of small r should be ahead of the other [25]. Table 1 shows an example of 6 learners sorted by NDS.

In Table 1, the column Sum Rank stands for the sum of rank of accuracy and diversity of an individual learner. And the column NDS Rank stands for the ranking by NDS algorithm. Learner 1 dominates Learner 2 at both the rankings of accuracy and diversity. Hence the ranking of Learner 1 is prior to Learner 2. But Learner 3 and Learner 4 cannot dominate each other. In such case, NDS uses the Sum Rank for sorting, which adds the ranking of accuracy and diversity together. Finally, we get a fully sorted list of all learners in the base set, and we select the top b% of the base set size to form an ensemble learner.

2.3. Deep Boosting. With the definition of accuracy and diversity of the base learners, we can sort the learners based on their quality. To further get an optimal weight for combination, an iterative procedure can be applied to search valuable data samples in the training dataset as well as updating the weights. Adaboost is a famous algorithm to find optimal ensemble weights. Algorithm 1 shows the main steps of Adaboost.

Require:
m: The size of ensemble D: The training data set
Ensure:
α: The weight vector for ensemble
(1) Define a uniform distribution V_1 on all samples in D
(2) **for** $i = 1$ to m **do**
(3) train h_i with a V_i
(4) Calculate $s_i = p_D(h_i(x) \neq y)$
(5) **if** $s_t \geq 1/2$ **then**
(6) break
(7) **end if**
(8) Set $\alpha_i = 1/2 \ln((1 - s_i)/s_i)$
(9) Update
(10) **for** $k = 1$ to $|D|$ **do**
(11) $V_{i+1}(k) = \dfrac{V_i \exp(-\alpha_i y_k h_t(x_k))}{Z_i}$
(12) **end for**
(13) **end for**
(14) **return** α

Algorithm 1: Adaboost.

Table 1: An example of 6 learners and their rankings.

Learner number	Accuracy rank	Diversity rank	Sum rank	NDS rank
Learner 1	1	2	3	1
Learner 2	2	3	5	3
Learner 3	5	4	9	4
Learner 4	4	6	10	5
Learner 5	3	1	4	2
Learner 6	6	5	11	6

In Adaboost, a uniform distribution V is imposed on the training dataset D. Each round the combination weights α and the distribution V are both updated according to the performance of the current learner on the whole training dataset. If a sample is misclassified by some learners, it would be chosen again with high probability, which is controlled by the distribution V.

When it comes to deep ensemble learning, a different sample selection and weight update strategy is implemented. The main idea of deep ensemble learning is described as follows. Firstly the initial distribution V is set to $V_i = 1/|D|$. Then try to solve the optimization problem as follows:

$$\min_{\alpha \geq 0} \quad \frac{1}{n}\Phi\left(1 - y_j \sum_{i=1}^{n} \alpha_i h_i\left(x_j\right)\right) + \lambda \sum_{i=1}^{n} \alpha_i r_i$$
$$\text{s.t.} \quad \sum i = 1^n \alpha_i \leq \frac{1}{n}. \tag{5}$$

Cortes et al. [26] proposed an algorithm to solve the above optimization problem, and a vector of optimal weights can be determined. Finally, for a test example x', the result can be $y' = (1/n) \sum_{i=1}^{m} h_i(x')$. For a binary output, a sign function s can be applied on y', in which $s(y') = 1$ if $y' \geq 0.5$ and 0 otherwise.

2.4. Base Learners. The quality of base learners affects the performance of the ensemble significantly. In this study, we use two kinds of base learners. The first is decision tree (DT) and the second is support vector machine (SVM). Note that both types of learners implement shallow models with two layers. For DT, a path from a leaf to the root is in fact a conjunctive normal form (CNF), and the root performs an OR operation of all paths in the tree; that is, $h_{\mathrm{DT}} = \bigcup_{i=1}^{p} c_i$. For SVM, the model is structured with a kernel operation between the test sample x_t and the samples of the training dataset D and then summarizes with a normalized weight vector; that is, $h_{\mathrm{SVM}}(x_t) = \sum_{i=1}^{|D|} \alpha_i k(x_t, x_i)$. For either DT or SVM, a three-layer model can be obtained by ensemble the trained base learners with a vector of learned weights.

The DT and SVM models are implemented by the famous WEKA project [27]. And in order to be invoked in MATLAB environment, we use the Spider project to generate a MATLAB interface for WEKA. To train each learner, a sampling procedure is launched on the training dataset D with replacement, resulting in some difference between the training datasets of each learner. The size of the set of base learners is denoted as m, including m_{DT} DTs and m_{SVM} SVMs with default parameter settings. In our evaluation, we set $m_{\mathrm{DT}} = 500$ and $m_{\mathrm{SVM}} = 500$ to build a relative large set of base learners, leading to a sufficient ensemble.

3. Evaluations

3.1. Dataset and Settings. We evaluate the proposed on a real clinical dataset gathered from some veteran TCM doctors, composing 2835 records. There are 21 different types of diseases in the dataset attached with 4 kinds of feature groups.

Table 2: Description of the evaluation dataset.

Number	Name	Type	Description
1	ICD-10 label	Boolean vector	The ICD-10 labels associated with the record
2	BasicInfo	Real vector	Patient's basic information, 11-ary
3	Diagnosis text feature	Boolean vector	4000-ary
4	Acupoints	Boolean vector	Acupoints in the patient's acupuncture plan, 53-ary

Table 3: Description of the evaluation dataset.

No.	Name	Sample data
1	ICD-10 labels	Fibromyalgia: M79.7
2	BasicInfo	Age: 33, gender: male, weight: 68, height: 171, and job type: heavy
3	Diagnosis text	The sequela of stroke hemiplegia: there has been some recovery, for many years has not double knee joint pain, were migratory, Jigzhi healed, recently accompanied by low back pain, pale tongue slightly red, and moss white veins fine strings
4	Acupoints	Huantiao, Yinmen, Taixi, Yaoyangguan, Changqiang, and YangChi (right)

The first group is the ICD-10 label vector. There are 31 ICD-10 labels concerning this study. But for each data record, there is only one ICD-10 label that can be attached. We use a boolean vector with 31 elements to indicate which ICD-10 label is attached among all labels. The second group contains the patient's information, including age, gender, job type, history of disease, weight, and height. All this information is placed in a real vector with 11 elements. The third group contains the diagnosis and ZHENG description of the patient in Chinese. The raw data of this field is in plain text which is not easy to process directly. We process them with a key word matching procedure. 4000 key words including the name of diseases, name of acupoints, ZHENG description, and severity description are predefined. And the diagnosis description text is matched with the set of key words. A boolean vector records the matching result whose element indicates whether the corresponding word exists in the text description. Finally the fourth group describes the acupoints proposed by the doctor for acupuncture treatment. In this study 53 acupoints are considered for analysis. Table 2 shows the feature of the evaluation dataset.

To make a clear presentation, Table 3 shows some examples of the dataset. Note that the name of acupoints and diagnosis description are originally in Chinese. We translate them into English for presentation in the table.

3.2. Evaluation Criteria and Methods for Comparison. To evaluate the effectiveness of the proposed method, we perform two types of evaluation. The first is to evaluate the prediction of ICD-10 labels given a diagnosis description and patient's basic information, as well as the acupoints for treatment. A zero-one loss function is adopted to evaluate the accuracy of the model output. Equation (6) shows the accuracy evaluated by a zero-one loss function:

$$\mathrm{Acc}\,(h, D) = \frac{1}{n}\sum_{i=1}^{n}\delta\left(h\left(x_i\right), y_i\right), \tag{6}$$

where $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. $\delta(\cdot, \cdot)$ is an indicator function where $\delta(y_i, y_j) = 1$ if $y_i = y_j$ and 0 otherwise.

For the second type of evaluation, we want to illustrate the effect of acupoint recommendation for a treatment plan given the basic information of a patient. This type of evaluation can be regarded as a multilabel classification problem. In this case, we adopt a Hamming loss to evaluate the accuracy. Equation (7) gives the definition of the Hamming loss:

$$\mathrm{Loss}_H\left(h\left(x\right), y\right) = \frac{1}{|y|}\sum_{i=1}^{|y|}\left(h\left(x\right)\Delta y\right). \tag{7}$$

In (7), y is the ground truth labels associated with x, and h is the learner to be evaluated. The Hamming loss function evaluates how many sample-label pairs are misclassified by the learner h.

We also implement two current state-of-the-art methods for the problem to be tackled in this paper and evaluate them on the same dataset, to further show the effectiveness of the proposed method. The first method is the multiview KNN method proposed by Liang et al. [9]. The second is a deep learning based method, which proposed a convolutional neural network for healthcare data decision making [28]. The motivation of choosing these two methods is twofold. The first is that both of them (Liang et al. [9, 28]) are proposed for TCM data analysis, which is similar to the theme of this study. And the evaluation dataset is the same as that used in this study. The second is that these two methods reflect two different directions for medical data analysis. The multiview KNN method in fact obeys the local learning and ensemble learning principles, leading to shallow model and transductive learning, which means that it is not necessary to derive a general model for the problem. The convolutional neural network method attempts to derive a classification function of powerful ability so as to express arbitrary complex classification boundary. For brevity, we denote these two methods as MV-KNN and CNN. The parameters of MV-KNN and CNN are set to default as they are proposed.

Table 4: ICD-10 annotation accuracy of each type of disease (%).

No.	Name	Size	DEL	MV-KNN	CNN
1	Arthralgia syndrome	481	82.4 ± 2.7	83.1 ± 2.8	**84.2 ± 3.4**
2	Acne	75	**90.2 ± 2.1**	86.4 ± 2.3	85.3 ± 3.1
3	Epilepsy	26	**89.1 ± 2.5**	88.0 ± 1.9	86.9 ± 2.4
4	Tinnitus and deafness	68	83.1 ± 2.6	81.0 ± 3.1	**85.2 ± 3.6**
5	Abdominal pain	96	**84.7 ± 2.7**	81.3 ± 2.9	82.8 ± 3.4
6	Allergic rhinitis	376	**89.2 ± 2.1**	84.1 ± 2.8	85.3 ± 3.0
7	Neck and shoulder pain	110	**91.4 ± 1.9**	88.4 ± 2.1	86.0 ± 2.5
8	Cervical spondylosis	33	**92.6 ± 2.2**	87.7 ± 2.9	90.5 ± 3.1
9	Cough	96	**88.5 ± 2.7**	86.9 ± 3.1	87.1 ± 3.9
10	Facial paralysis	89	**82.7 ± 1.3**	78.8 ± 2.5	79.1 ± 3.0
11	Traumatic brain injury	47	85.8 ± 2.1	**86.0 ± 2.9**	85.1 ± 2.6
12	Migraine	33	**93.0 ± 2.9**	88.7 ± 3.2	89.4 ± 3.6
13	Ankylosing spondylitis	33	**91.9 ± 2.0**	90.0 ± 3.6	91.1 ± 3.9
14	Insomnia	47	**90.2 ± 2.2**	84.5 ± 3.3	88.5 ± 3.6
15	Headache	145	86.6 ± 2.5	87.1 ± 3.2	**89.2 ± 3.8**
16	Flaccidity syndrome	124	**87.2 ± 1.9**	83.1 ± 2.8	84.4 ± 3.1
17	Stomachache	145	**89.2 ± 2.4**	86.5 ± 2.8	87.2 ± 3.1
18	Asthma	355	**90.6 ± 2.1**	88.2 ± 2.9	88.6 ± 2.9
19	Palpitation	33	**90.2 ± 2.5**	89.9 ± 3.1	86.5 ± 3.4
20	Lumbocrural pain	397	**88.1 ± 2.3**	82.1 ± 3.2	87.2 ± 3.8
21	Urticaria and rubella	26	**85.4 ± 2.3**	84.2 ± 3.1	84.9 ± 3.0
22	Total	**2835**	**88.2 ± 2.8**	85.6 ± 3.4	86.4 ± 3.9

3.3. Evaluation Results. We use a tenfold validation strategy for evaluation. The whole dataset is randomly divided into 10 parts with equal sizes. In each round, 9 parts are used to train the model and the remainder for test. We randomly divide the dataset 20 times. For each time a tenfold validation is run. Totally there are 200 runs. The mean loss and stand derivation are recorded in either kind of evaluation. Table 4 shows the ICD-10 annotation accuracy of each type of disease.

The column DEL stands for the accuracy of the proposed method. In Table 4, we boldface the best result in each row. At the last of the table, we summarize the accuracy of three methods. It can be seen that the proposed method has best performance in the annotation of 17 (totally 21) types of diseases. Moreover, in a multiple-label classification perspective, the proposed method also achieves the best result for all diseases to be annotated, as shown in the last row of Table 4. It can be concluded that the proposed method is effective for the annotation of the concerned diseases. The proposed method achieves best performance among all three methods for $17/21 \approx 81.0\%$ types of disease and for average results of all diseases, which indicates that the proposed method is statistically better than the other two methods.

For the second part of evaluation, we want to see the accuracy of acupoints recommendation for treatment. We compare the ground truth acupoints suggested by experienced doctors with the model output. Note that in this part MV-KNN and CNN are not suitable for this case. Henceforth we only report the accuracy measured by Hamming loss and the variance of the whole accuracy of the proposed method. Table 5 shows the results of this session of evaluation.

4. Conclusions

In this paper, we proposed an ensemble learning framework for ICD-10 label annotation and acupoints recommendation. The model analyzes the clinical diagnosis records in plain text, acupoints for acupuncture treatment, and the patient's basic information and performs multilabel classification to annotate correct ICD-10 labels for each clinical record. At the same time, the model recommends acupoints for personal treatment, which provides valuable support for doctor's diagnosis decision. The proposed method adopts the recently proposed deep ensemble learning to find the optimal weight vector for combination of base learners. Different from the traditional Adaboost method, the deep ensemble learning can achieve better generalization ability when given a set of base learners with powerful representation ability. Decision tree and support vector machine classifiers are implemented as the base learners. We set up our evaluation on a real clinical dataset gathered from several veteran doctors, with comparison to two previously proposed successful methods. We achieve an accuracy of 88.2% in ICD-10 labels annotation evaluated by the zero-one loss function and 79.6% in acupoints recommendation evaluated by the Hamming loss function, either of which is superior to the two previous methods.

Conflict of Interests

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

Table 5: Acupoints recommendation accuracy of each type of disease (%).

No.	Name	Accuracy	No.	Name	Accuracy
1	Arthralgia syndrome	77.9	2	Acne	82.3
3	Epilepsy	80.6	4	Tinnitus and deafness	75.6
5	Abdominal pain	76.8	6	Allergic rhinitis	77.2
7	Neck and shoulder pain	81.0	8	Cervical spondylosis	80.5
9	Cough	82.4	10	Facial paralysis	76.5
11	Traumatic brain injury	79.1	12	Migraine	78.4
13	Ankylosing spondylitis	78.3	14	Insomnia	81.4
15	Headache	80.0	16	Flaccidity syndrome	75.9
17	Stomachache	83.0	18	Asthma	81.2
19	Palpitation	82.2	20	Lumbocrural pain	80.1
21	Urticaria and rubella	77.9	—	—	—
22	Total	**79.6 ± 3.6**	—	—	—

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 81373883), the Natural Science Foundation of Guangdong Province of China (no. S2013040012898), the Science and Technology Planning Project of Guangdong Province of China (no. 2013B010404019), the College Student Career and Innovation Training Plan Project of Guangdong Province (yj201311845015, yj201311845023, yj201311845031, and xj201411845025), the Higher Education Research Funding of Guangdong University of Technology (no. GJ2014Y17), and the Science and Technology Project of Huadu District of Guangzhou (no. 14ZK0024).

References

[1] S. Qiao, C. Tang, H. Jin, J. Peng, D. Davis, and N. Han, "KISTCM: knowledge discovery system for traditional Chinese medicine," *Applied Intelligence*, vol. 32, no. 3, pp. 346–363, 2010.

[2] Y. Wang, Z. Yu, L. Chen et al., "Supervised methods for symptom name recognition in free-text clinical records of traditional Chinese medicine: an empirical study," *Journal of Biomedical Informatics*, vol. 47, pp. 91–104, 2014.

[3] Y. Feng, Z. Wu, X. Zhou, Z. Zhou, and W. Fan, "Methodological review: knowledge discovery in traditional Chinese medicine: state of the art and perspectives," *Artificial Intelligence in Medicine*, vol. 38, no. 3, pp. 219–236, 2006.

[4] N. L. Zhang, S. Yuan, T. Chen, and Y. Wang, "Latent tree models and diagnosis in traditional chinese medicine," *Artificial Intelligence in Medicine*, vol. 42, no. 3, pp. 229–245, 2008.

[5] Y. Wang, Z. Yu, Y. Jiang, Y. Liu, L. Chen, and Y. Liu, "A framework and its empirical study of automatic diagnosis of traditional Chinese medicine utilizing raw free-text clinical records," *Journal of Biomedical Informatics*, vol. 45, no. 2, pp. 210–223, 2012.

[6] F. Zou, D. Yang, S. Li, J. Xu, and C. Zhou, "Level set method based tongue segmentation in traditional chinese medicine," in *Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering (GVE '07)*, pp. 37–40, ACTA Press, Anaheim, Calif, USA, 2007, http://dl.acm.org/citation.cfm?id=1712936.1712945.

[7] S. Lukman, Y. He, and S.-C. Hui, "Computational methods for traditional Chinese medicine: a survey," *Computer Methods and Programs in Biomedicine*, vol. 88, no. 3, pp. 283–294, 2007.

[8] Z. Di, H.-L. Zhang, G. Zhang et al., "A clinical outcome evaluation model with local sample selection: a study on efficacy of acupuncture for cervical spondylosis," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '11)*, pp. 829–833, IEEE Computer Society, November 2011.

[9] Z. Liang, G. Zhang, G. Li, and W. Fu, "An algorithm for acupuncture clinical assessment based on multi-view KNN," *Journal of Computational Information Systems*, vol. 8, no. 21, pp. 9105–9112, 2012.

[10] G. Zhang, Z. Liang, J. Yin, W. Fu, and G.-Z. Li, "A similarity based learning framework for interim analysis of outcome prediction of acupuncture for neck pain," *International Journal of Data Mining and Bioinformatics*, vol. 8, no. 4, pp. 381–395, 2013.

[11] N. Rooney, H. Wang, and P. S. Taylor, "An investigation into the application of ensemble learning for entailment classification," *Information Processing and Management*, vol. 50, no. 1, pp. 87–103, 2014.

[12] V. Kuznetsov, M. Mohri, and U. Syed, "Multi-class deep boosting," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., vol. 27, pp. 2501–2509, Curran Associates, 2014.

[13] D. Akdemir and J.-L. Jannink, "Ensemble learning with trees and rules: supervised, semi-supervised, unsupervised," *Intelligent Data Analysis*, vol. 18, no. 5, pp. 857–872, 2014.

[14] M. Uchida, Y. Maehara, and H. Shioya, "Unsupervised weight parameter estimation method for ensemble learning," *Journal of Mathematical Modelling and Algorithms*, vol. 10, no. 4, pp. 307–322, 2011.

[15] Q. L. Zhao, Y. H. Jiang, and M. Xu, "Incremental learning by heterogeneous bagging ensemble," in *Advanced Data Mining and Applications: 6th International Conference, ADMA 2010, Chongqing, China, November 19–21, 2010, Proceedings, Part II*, vol. 6441 of *Lecture Notes in Computer Science*, pp. 1–12, Springer, Berlin, Germany, 2010.

[16] V. Pisetta, P.-E. Jouve, and D. A. Zighed, "Learning with ensembles of randomized trees: new insights," in *Machine Learning and Knowledge Discovery in Databases: European Conference,*

ECML PKDD 2010, Barcelona, Spain, September 20–24, 2010, Proceedings, Part III, vol. 6323 of *Lecture Notes in Computer Science*, pp. 67–82, Springer, Berlin, Germany, 2010.

[17] H. Yu and J. Ni, "An improved ensemble learning method for classifying high-dimensional and imbalanced biomedicine data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 4, pp. 657–666, 2014.

[18] F. Bellal, H. Elghazel, and A. Aussem, "A semi-supervised feature ranking method with ensemble learning," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1426–1433, 2012.

[19] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*, Springer, New York, NY, USA, 2012.

[20] X. S. Zhang, B. Shrestha, S. Yoon et al., "An ensemble architecture for learning complex problem-solving techniques from demonstration," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 4, article 75, 2012.

[21] J. Tian, M. Li, F. Chen, and J. Kou, "Coevolutionary learning of neural network ensemble for complex classification tasks," *Pattern Recognition*, vol. 45, no. 4, pp. 1373–1385, 2012.

[22] I. Choi, K. Park, and H. Shin, "Sharpened graph ensemble for semi-supervised learning," *Intelligent Data Analysis*, vol. 17, no. 3, pp. 387–398, 2013.

[23] C.-X. Zhang, J.-S. Zhang, N.-N. Ji, and G. Guo, "Learning ensemble classifiers via restricted Boltzmann machines," *Pattern Recognition Letters*, vol. 36, no. 1, pp. 161–170, 2014.

[24] J. Liao, *Totally corrective boosting algorithms that maximize the margin [Ph.D. thesis]*, University of California, Santa Cruz, Calif, USA, 2006.

[25] Q. Wang, L. Zhang, M. Chi, and J. Guo, "MTForest: ensemble decision trees based on multi-task learning," in *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI '08)*, pp. 122–126, IOS Press, Amsterdam, The Netherlands, 2008.

[26] C. Cortes, M. Mohri, and U. Syed, "Deep boosting," in *Proceedings of the 31st International Conference on Machine Learning (ICML '14)*, Beijing, China, June 2014.

[27] R. R. Bouckaert, E. Frank, M. A. Hall et al., "WEKA—experiences with a java open-source project," *Journal of Machine Learning Research*, vol. 11, pp. 2533–2541, 2010.

[28] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu, "Deep learning for healthcare decision making with EMRs," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM '14)*, pp. 556–559, November 2014.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 473168, 12 pages
http://dx.doi.org/10.1155/2015/473168

Research Article

ISMAC: An Intelligent System for Customized Clinical Case Management and Analysis

Mingyu You,[1] Chong Chen,[1] Guo-Zheng Li,[1] Shi-Xing Yan,[1] Sheng Sun,[1] Xue-Qiang Zeng,[2] Qing-Ce Zhao,[1] Liao-Yu Xu,[3] and Su-Ying Huang[3]

[1]*Department of Control Science and Engineering, Tongji University, Shanghai 200000, China*
[2]*Computer Center, Nanchang University, Nanchang 330000, China*
[3]*Shanghai Institute of Traditional Chinese Medical Literatures, Shanghai 200000, China*

Correspondence should be addressed to Mingyu You; myyou@tongji.edu.cn

Received 22 January 2015; Revised 29 March 2015; Accepted 28 April 2015

Academic Editor: Josiah Poon

Clinical cases are primary and vital evidence for Traditional Chinese Medicine (TCM) clinical research. A great deal of medical knowledge is hidden in the clinical cases of the highly experienced TCM practitioner. With a deep Chinese culture background and years of clinical experience, an experienced TCM specialist usually has his or her unique clinical pattern and diagnosis idea. Preserving huge clinical cases of experienced TCM practitioners as well as exploring the inherent knowledge is then an important but arduous task. The novel system ISMAC (Intelligent System for Management and Analysis of Clinical Cases in TCM) is designed and implemented for customized management and intelligent analysis of TCM clinical data. Customized templates with standard and expert-standard symptoms, diseases, syndromes, and Chinese Medince Formula (CMF) are constructed in ISMAC, according to the clinical diagnosis and treatment characteristic of each TCM specialist. With these templates, clinical cases are archived in order to maintain their original characteristics. Varying data analysis and mining methods, grouped as Basic Analysis, Association Rule, Feature Reduction, Cluster, Pattern Classification, and Pattern Prediction, are implemented in the system. With a flexible dataset retrieval mechanism, ISMAC is a powerful and convenient system for clinical case analysis and clinical knowledge discovery.

1. Introduction

Traditional Chinese Medicine (TCM) is one of the main approaches for disease diagnosis and treatment in China [1, 2]. The basic theories were formed more than 2000 years ago [3], and according to the reported data of Chinese National Bureau of Statistics [4], there were 2688 TCM hospitals, receiving 8.89 million inpatients and 275 million outpatients in China 2008. It is widely accepted in China that TCM is safer and more efficient for some chronic and intractable illnesses. In past decades, TCM has been increasingly adopted around the world as the complementary medical therapy for various diseases such as cancer [5], rheumatoid arthritis [6], leukemia [7], H1N1 Virus [8], and migraines [9]. In TCM diagnosis, discomforts in different parts of the body are taken into consideration together to find the root pathological cause. TCM also maintains that the health of individual human beings is intimately involved with the environment. Excellent TCM practitioners pursue the discovery of potential disease before it may be perceived by the patient or examined by medical instruments. The master goal of TCM is preventative: a common modern desire.

As opposed to modern biomedical science, diagnostic knowledge and TCM herb formulas are mostly developed through practice. Veteran TCM practitioners usually have more than 30 years of clinical experience. Moreover, all TCM experts have a unique understanding of TCM philosophy, which results in unique diagnosis and treatment patterns. Therefore, to archive clinical data as well as maintain different characteristics in one system is a difficult but vital task in TCM informatics.

TCM clinical cases archive the important experience of veteran TCM practitioners and store useful information for young TCM physicians. The analysis of clinical data can help

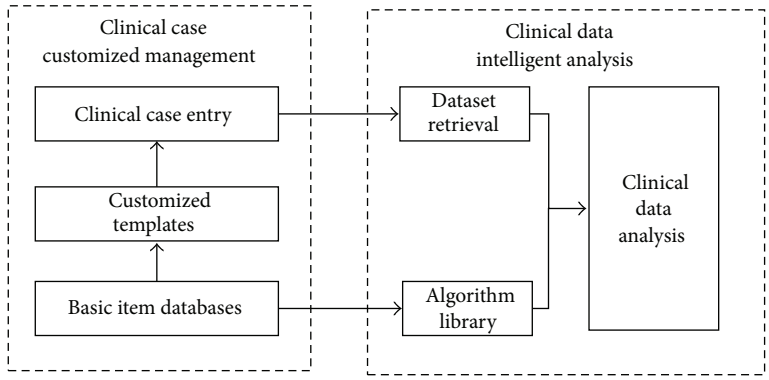


Figure 1: Functional overview of the platform.

in the extraction of vital clinical strategies for TCM knowledge accumulation and promotion [10]. Clinical records are important for the understanding of theory and intellectual gain for new doctors. In a traditional sense, successors follow their teachers and study the in-depth knowledge while developing personal ability. But nowadays, with the unprecedented growth of clinical data, it is difficult to extract new knowledge from the data mountain. Data mining is a distinguished method of tracking underlying information. A great deal of research is dedicated to TCM data mining [11–13]. This indicates a promising future for TCM data mining, but most research contributes to TCM literature mining, herb knowledge modeling, gene network analysis, and the role of the understanding of herb components. None of this focuses on TCM clinical case analysis, most especially clinical records from highly experienced TCM practitioners.

A powerful, intelligent system is key for clinical case collection and analysis. Tian et al. [14] propose a platform for medical image processing and analyzing. It is powerful and useful but is designed specifically for medical image processing, which is a different task from clinical data management. Masseroli and Marchente [15] and Ku and Huang [16] propose efficient platforms for healthcare requirements, both of which are based on the web structure and provide convenient services to the patient. However, none of the platforms have been designed for TCM clinical cases. Chen et al. [17] demonstrate a grid-based TCM system, which also extends to a semantic-based database grid. Zhou et al. [18] developed TCMMDB, a unified, web-accessible multidatabase query system which has already integrated more than 50 databases, and CDW [11], a clinical data warehouse which incorporates the structured electronic medical record (SEMR) data for medical knowledge discovery and TCM clinical decision support (CDS). These systems provide platforms for TCM clinical data or herb collection but pay little attention to the clinical cases of high-experienced TCM practitioners. As opposed to strictly standardized modern biomedicine, the clinical cases of TCM practitioners, especially veteran practitioners, have obvious personal characteristics [19], which greatly distinguish many personal cases from others. The philosophy of syndrome differentiation from a patient's symptoms varies among different veteran practitioners. This variation may be the embodiment of the TCM veteran practitioner's thinking mode and knowledge system: more effort should be taken to preserve and analyze these systems. There are many tasks in the analysis of clinical records, such as symptom reduction, syndrome classification, and core formula mining. Since the data has the characteristics of being high dimension, multiclass, imbalance, the discovery of knowledge in clinical records is challenging.

In this paper, we present a novel system designed exactly for the customized management of TCM clinical cases and discuss the intelligent techniques for clinical data analysis and knowledge mining. The overview of our system has been proposed [20]. The rest of the paper is arranged as follows. An overview of this system is demonstrated in Section 2. The details for the design of the case management and analysis are introduced in Sections 3 and 4, respectively. An example for the system usage with clinical data from Mr. Zhang is provided in Section 5. Finally, Section 6 concludes the paper and presents a future perspective.

2. Platform Overview

Functional overview of ISMAC is illustrated in Figure 1. The customized management and intelligent analysis of TCM clinical cases are the two main parts in the system. The customized template is the critical element in data customized management. The template is defined as the entry format designed for clinical cases. It provides a predefined layout for specific doctor's medical records, which contain information like patient demographic information, symptoms, examinations, syndromes, and therapeutic principles. In ISMAC, we customize the entry template for each doctor, according to diagnosis methodology. The basic Item Database is set up to provide a foundation for the creation of the Customized Template, and Clinical Case Entry can be efficient and effective with the suitable Customized Template. As for intelligent analysis, algorithms and datasets are the two bases. ISMAC integrates not only popular algorithms in statistic analysis and data mining, but also some newly proposed methods in machine learning, in order to build a library for TCM clinical case intelligent analysis. In conjunction with

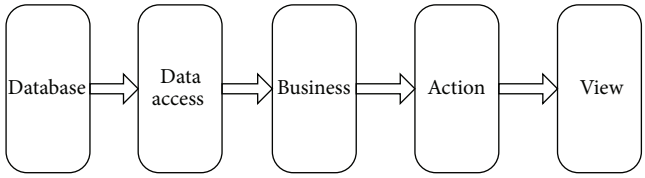


Figure 2: Framework of the platform.

the flexible Dataset Retrieval, the Algorithm Library makes various kinds of intelligent analysis possible in ISMAC.

ISMAC is implemented as a web-based application using the classical SSH (Struts + Spring + Hibernate) framework. As shown in Figure 2, it consists of 5 layers. The Database Layer contains tables storing various information, ranging from basic items such as symptom and syndrome to complex items, like knowledge models and learning algorithms. In the Data Access Layer, the Object-Relational Mapping (ORM) mechanism is used for convenient data manipulation. The top three layers, Business, Action, and View, are configured as the Model-View-Controller (MVC) pattern, in which the Business Layer manages the behavior and data in the application domain, the View Layer manages the display of different information, and the Action Layer interprets the mouse and keyboard inputs from the user and informs the Business Layer and the View Layer to react appropriately.

3. Custom Management of Clinical Case

The customized management of clinical cases is the core function in the ISMAC system, as archiving the clinical cases of different veteran TCM practitioners using the same entry form may submerge important knowledge in the clinical data.

3.1. Overall Design. The overall design of the Customized Management of Clinical Case is shown in Figure 3, which illustrates the left part of Figure 1 in detail. Members of the Basic Item Database are listed. Varying libraries in the Basic Item Database possess the same structures but store different kinds of items: standard items and expert-standard items. A Customized Template consists of three parts: fixed, standard, and expert. These three parts are built up with different basic items. Clinical Case Entry depends on not only Customized Template but also the CMF and Herb item libraries.

Figure 3 also tells us that the task of customized management is divided into four subtasks, and charged by three roles. This is to increase work efficiency and protect private information. The three roles are Item Management Assistant (IMA), Case Entry Assistant (CEA), and the Case Management Operator (CMO), which is, respectively, assigned tasks of basic item maintenance and customized template maintenance, case input, and case management.

More details about the function of each part in customized management and its implementation technologies are introduced in the following.

3.2. Introduction of Functions

3.2.1. Basic Item Database. Basic Item Database targets supporting template creation and case entry. As illustrated in Figure 3, libraries included in it can be classified into three categories for distinct usages: fixed part, standard part, and expert customized part. The last two categories may store two different kinds of items: standard items and expert-standard items.

The fixed Item Library in the left contains items such as patient name and created date. This is entitled "fixed" because these items appear in every template, and the library determines the fixed part of templates.

The Herb Library and CMF Library on the right store items about herbs and CMFs. Each CMF is composed of several herbs and corresponding dosages. The contents of these two libraries have two sources: one is *the Pharmacopoeia of the People's Republic of China* and the other contains each specialist's experience. Items from the first source are entitled standard items in ISMAC, and those from second source are tagged as expert-standard. In ISMAC, the two libraries directly serve Case Entry. CEAs choose items from this when filling a specific template.

The third category of item libraries is shown in the middle of the Basic Item Database. They work for the customization of templates. A customized itemset for a specific disease or syndrome is picked up from these libraries, to constitute a template together with fixed items. Six libraries are embodied in this category, including the Symptom Library, Examination Library, TCM Syndrome Library, TCM Disease Library, WM Disease Library, and Therapeutic Principle Library, as they are normal parts of a TCM clinical case. Similarly, both Herb and CMF libraries have standard and expert-standard items. The standard part comes from national or sectorial standards. The expert-standard part is the summary of veteran TCM practitioners experience.

3.2.2. Customized Templates. Customized Templates are the core and basis for clinical case customized management.

The characteristics of TCM clinical cases prompt us to propose customized templates for a particular specialist. There are thousands of items carried in TCM diagnosis. However, almost all of the TCM practitioners take a subset of these items into consideration in clinical practice. Moreover, with different clinical experience and personal understanding of TCM theory, each practitioner has a different approach to diagnosis and clinical explanation. Providing a uniform sheet for data entry would cause two significant problems: clinical case entry would be inefficient (CEAs have to go through many unused items), and a uniform sheet cannot retain the clinical thinking of veteran TCM doctors in original taste. The knowledge of the TCM clinic is vital to TCM theory development and therefore deserves more attention and protection.

Three categories of basic items are used in Customized Template creation. Fixed items will be added to a template automatically. Standard items can be easily chosen from corresponding libraries in the Basic Item Database. For expert-standard items, those previously saved in the database

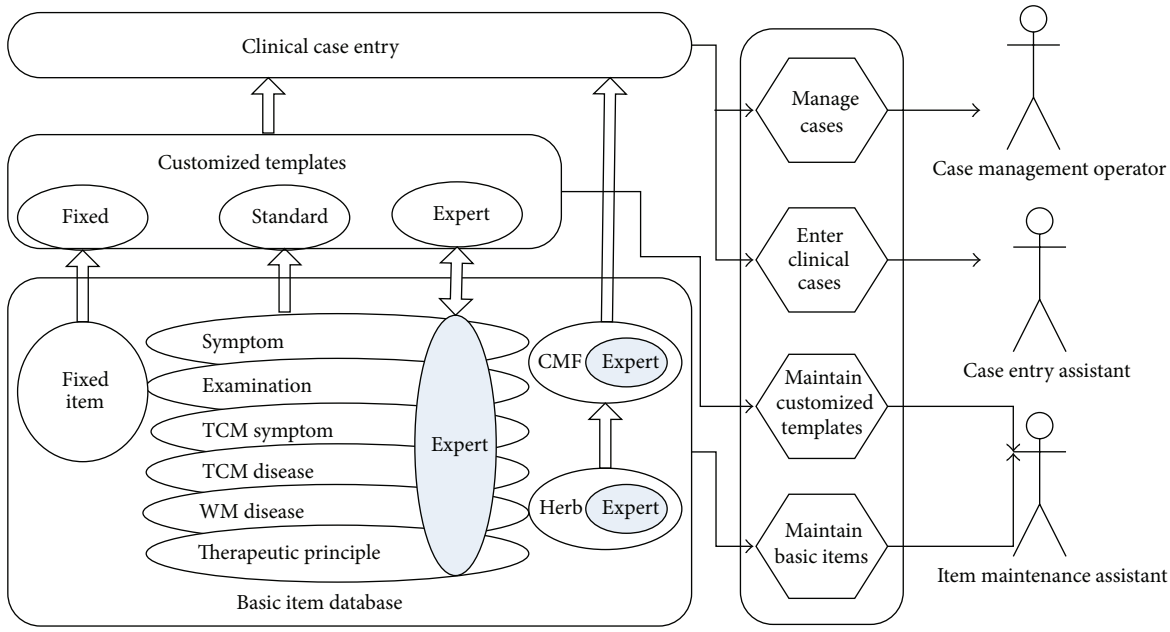


Figure 3: Overall design of the customized management of clinical case.

can be selected like standard items. Otherwise, manual entry in a matched box is needed.

Customized Templates contribute largely to ISMAC's customization feature. However, this does not weaken its standardization. Most items in templates are standard or expert-standard items from the database. Items possessing the same names and IDs must refer to the same database. Multi-items for the same symptom or syndrome are forbidden. Such mechanisms ensure customization while retaining standardization.

3.2.3. Clinical Case Collection and Management. With the establishment of Customized Templates and Basic Item Database, ISMAC will autogenerate a form with corresponding textboxes or checkboxes for the CEA to type in cases. A case usually contains the basic information of the patient and his diagnosis information, such as symptoms, syndromes, and CMF in every clinical visit. In some other systems, diagnoses resulting from different visits are usually treated as different clinical cases, which should repeat basic information entry and cut the strong relation between clinical visits. ISMAC links information of all visits together. With accumulated entry cases, more advanced data management operations can be created. We maintain that case management can be executed only by a CMO, who is often a doctor. Usually, he can check how many diagnoses he has conducted and review some important ones. He is also responsible for the verification of cases. Moreover, he can export his case records for backup, or other purposes.

3.2.4. Access Control. As mentioned above, four tasks are assigned to three roles in ISMAC. Technically, the function is realized by the system's access control mechanism, which includes both role-based permission assignment and specialist-specific data access.

The access control is set up mainly for information security, which is important to the field of medicine and health care. Patients usually do not want clinical data viewed by others, except for physicians. On the other hand, the division of labor increases work productivity.

Role-based permission assignment in customized management part involves IMA, CMO, and CEA. IMA is created as an individual role, taking charge of maintenance (creation, update, etc.) for basic items and templates. They help build up the system but have no permission to input or read any clinical cases. DMO, always the physician, can check their own cases and execute jobs like case verification and export. The CEA helps enter the clinical cases and has no access to any stored data.

Specialist-specific data access is intended to limit all operators to access data, except for the appropriate physician. This means that a physician can only view their own clinical cases, and IMAs can only maintain items for a specific specialist, and CEAs can only enter cases with templates belonging to a particular physician.

4. Clinical Data Intelligent Analysis

Clinical case intelligent analysis is important for expert knowledge understanding and TCM theory development. The diagnosis methods and therapeutic means in TCM are the summation of clinical trials by intelligent specialists over many generations. However, in modern society, it is difficult and inefficient for physicians to learn TCM practical knowledge from solely clinical and personal experience. Intelligent analysis and advanced methods can provide a

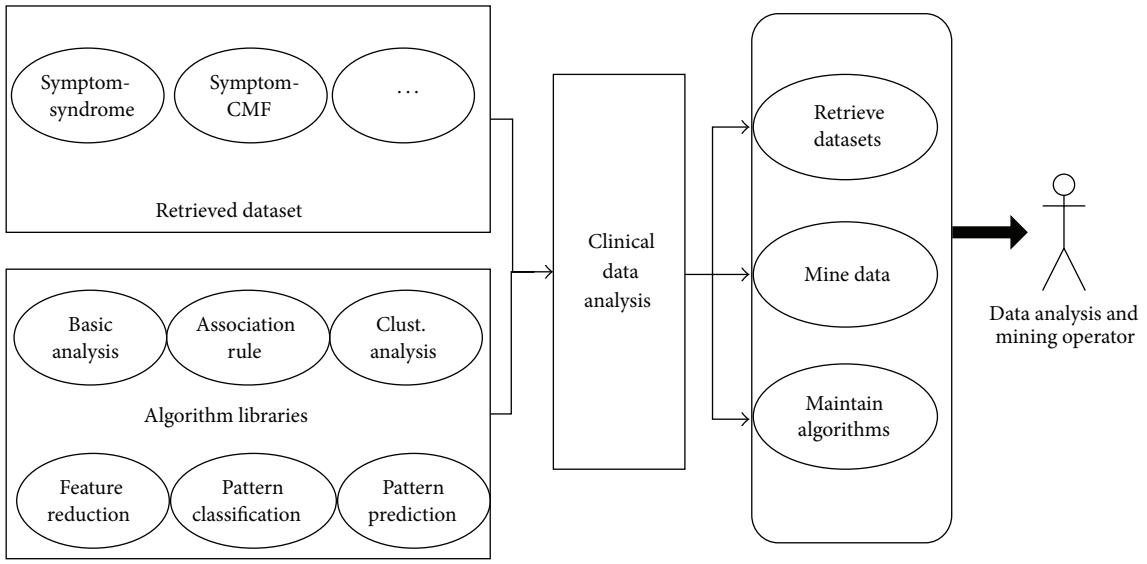


Figure 4: Overview design of intelligent data analysis.

rapid investigation of clinical data. This knowledge and models may be difficult to retrieve with human intelligence.

4.1. Overview Design. Figure 4 details the right part of Figure 1. Dataset Retrieval is flexible, and various kinds of attribution combinations can be retrieved from the clinical case database. For example, datasets with symptom and syndrome attributes can be retrieved for the purpose of syndrome classification, while a dataset with CMF information can only be retrieved for core CMF mining. The Algorithm Library contains algorithms categorized into six groups: Basic Analysis, Association Rule, Feature Reduction, Cluster, Pattern Classification, and Pattern Prediction. With datasets and algorithms, varied intelligent analysis tasks for TCM clinical cases can then be conducted.

Similar to clinical case customized management, the work of intelligent analysis is also divided into several tasks. All the tasks are associated closely and assigned by the Data Analysis and Mining Operator (DAMO). Specialist-specific data access functions in access control mechanisms are also limiting. Each DAMO can only access the data (including cases and templates) of a specific TCM specialist.

4.2. Introduction of Functions

4.2.1. Dataset Retrieval. The first step of data analysis is to retrieve the target dataset. A target dataset is the one that belongs to a target specialist and a certain template and contains necessary record items. The data stream diagram is shown in Figure 5. Because each DAMO is only capable of accessing one specialist, Access Control helps to ensure the correct choice of specialist. Then, the target template and items can be chosen through a corresponding user interface. And finally, the formed dataset is filtered out from the massive clinical case data. Information of each formed dataset is stored in ISMAC's database for multiple usage. It may be

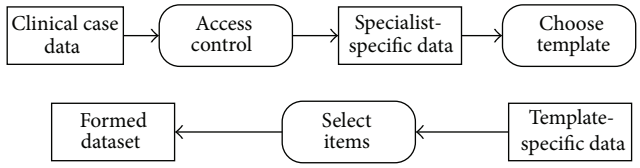


Figure 5: Data stream diagram of dataset retrieval.

conveyed into .arff files and used by *Weka*-based algorithms. Furthermore, this can also be exported in a .csv format for sharing, or further investigation with other tools.

4.2.2. Algorithm Library. The next step in data analysis is to choose the proper algorithms with suitable parameters. Many algorithms based on Weka are integrated in ISMAC. *Weka* [21] is one of the most popular toolkits in the field of machine learning and data mining, which is written in Java. To avoid reinventing the wheel, we wrap *Weka* into ISMAC. However, algorithms in *Weka* are not enough to deal with TCM clinical cases. Thus the extension of *Weka* has been made to include our newly proposed methods. All algorithms built in ISMAC are summarized as follows.

Algorithms Organized as Six Groups

Basic Analysis

- Statistics computing
- Visualization tools

Feature Reduction

- PCA
- ICA
- CFS

IG
Im-IG

Pattern Classification

NaiveBayes
SVM
J48
PLSC
APLSC
MAPLSC

Association Rule

Apriori
AprioriTid
Fp-tree
AVM

Cluster

EM
*K*Means

Pattern Prediction

Linear Regression
Logistic Regression.

Training Section of MAPLSC

Training dataset D, the number of classes k

MAPLSC classifier:

(1) Set up $k(k-1)/2$ subsets from the dataset D, each subset S_{ij} is composed of the examples from class C_i and class C_j.

(2) Train classifier $APLSC_{ij}$ with the examples in S_{ij}, and obtain output f_m from $APLSC_{ij}$ for each example x_m in S_{ij}.

(3) Calculate posterior probability parameters A_{ij} and B_{ij} with (1) and f_m.

(4) Output *MAPLSC* classifier with $k(k-1)/2$ $APLSC_{ij}$ and the corresponding A_{ij} and B_{ij}.

Testing Section of MAPLSC

MAPLSC classifier and corresponding posterior probability parameters, test sample x_t, predicted label $\widetilde{y_t}$ for test sample x_t:

(1) Calculate the output f_t^{ij} from $APLSC_{ij}$ for test example x_t.

(2) Calculate the posterior probability output $r_{ij} = \text{Prob}(C_i \mid x_t)$ by (1), with parameters A_{ij}, B_{ij} and f_t^{ij}.

(3) Combine the $k(k-1)/2$ r_{ij} by (2), and get the probabilities on k classes $(\widetilde{p_1}, \widetilde{p_2}, \widetilde{p_3}, \widetilde{p_4})$.

(4) Predict label for x_t by $\widetilde{y_t} = \text{argmax}_i p_i$.

Basic Analysis Algorithm Group contains algorithms for basic statistical analysis and data visualization. The cluster algorithm is intended to cluster instances with similar properties. The Association Rule Library provides methods that help find frequent patterns and association rules from a data set. *Apriori* [22] is an important algorithm for frequent pattern mining, and many other methods are based upon it.

Feature Reduction helps to reduce the volume of the feature set yet maintains the majority of the original information. It can be divided into two branches, feature extraction and feature selection. Feature extraction applies data encoding or transformation to obtain the "compressed" presentation of original data, while feature selection tries to remove irrelevant, weakly relevant, and redundant features. One new feature of selection algorithms proposed by us is integrated in ISMAC, known as *Im-IG* [23], an improved Information Gain (IG) for imbalance (Im) problem.

Pattern Classification and Pattern Prediction algorithms are used to build classification and prediction models. Pattern classification assigns one or several categorical (discrete, unordered) labels to a given instance. Pattern Prediction algorithms assign input instances to continuous prediction values. Three of our proposed novel methods are provided in the Pattern Classification Algorithm Group. PRIFEAB adopts asymmetric bagging and feature selection to address the imbalanced problem. *APLSC* [24] and *MAPLSC* [25] also attempt to address the imbalance problem for binary and multiple problems, respectively, based on the Partial Least Squares Classifier (PLSC).

4.2.3. Novel Method, MAPLSC. Here we would like to introduce in detail one of the novel methods, MAPLSC, which is integrated in ISMAC. MAPLSC is based on the APLSC algorithm. It combines multiple one-versus-one APLSC binary classifiers to solve the multiclass imbalance problem. The pseudocode of MAPLSC is illustrated in the previous section.

In the training section, one-versus-one binary classifiers for all classes are first trained with the APLSC algorithm. After that, (1) is used to transfer output f of APLSC to posterior possibility parameters A and B, used for classifier combinations:

$$\text{Prob}\left(C_i \mid x\right) = \frac{1}{1 + \exp\left(Af + B\right)}. \tag{1}$$

In the testing section, a simplified version of the pairwise coupling strategy as proposed by Hastie and Tibshirani is used to combine the probabilistic outputs of all the APLSC classifiers as well as the output estimates of the posterior probabilities for all candidate classes. The combination equation is illustrated in

$$\widetilde{p_i} = 2\sum_j \frac{r_{ij}}{k\,(k-1)}. \tag{2}$$

4.2.4. Our Framework of TCM Data Mining. In TCM, hierarchical data analysis demands have risen. Figure 6 demonstrates our framework analysis. The top layer includes items from clinical cases. Symptom, syndrome, and CMF are the

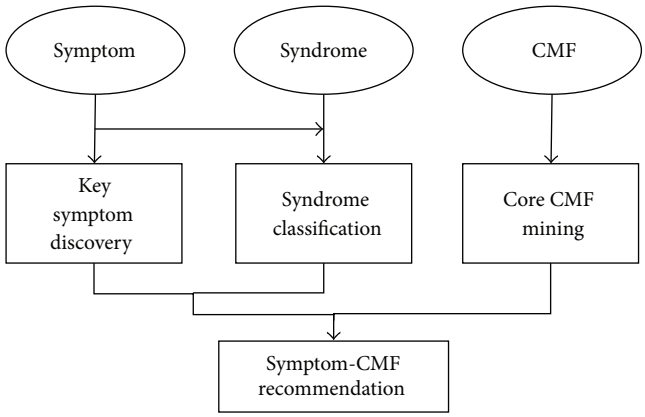


Figure 6: Framework of TCM clinical data analysis.

three most important clinical items in which researchers are typically involved. The middle layer and bottom layer consist of four analysis directions that attract great attention in the field of TCM.

(i) Key Symptom Discovery. The Key Symptom Discovery is used to find the remarkable symptoms, which may be biomarkers for a specific disease or syndrome. As mentioned previously, items in templates are the summary of doctors' clinical knowledge. However, in the previous process, items are selected based on doctor's personal understanding and experience, which tends to be subjective and one-sided. Key Symptom Discovery discovers significant symptoms with objective algorithms, whereas the Feature Selection Group is usually used.

(ii) Syndrome Classification. ISMAC helps doctors to diagnose disease/syndrome using the input symptoms and classification models. This speeds up knowledge dissemination in TCM by offering students more chances to study the clinical thoughts of the TCM veteran, as compared with traditional apprenticeship. Furthermore, ISMAC with a syndrome classification function may be deployed in outlying areas, helping junior doctors to improve diagnostic skills and treatment effects.

(iii) Core CMF Mining. Core CMF is the basic Chinese Medicine Formula for a specific disease or syndrome. It is worthy of discovery since it reflects doctoral thought in clinical practice and the core of clinical knowledge. CMFs are also usually a valuable source of patent medicine. Mining CMFs could also bring great commercial value. Several methods embedded in ISMAC may be adopted to the core CMF mining. Algorithms in the Association Rule Group help find the closely associated herb subsets which are the candidates of core CMF. Clustering methods may be attempted to find herb clusters for CMF analysis.

(iv) Symptom-CMF Recommendation. Symptom-CMF Recommendation is the combination and improvement of all three analyses in the middle layer. After a new patient inputs his symptoms, the corresponding classification model is called to judge the syndrome. The Key Symptom Discovery is embedded into the syndrome prediction operation for precision and speed. Core CMF for that syndrome can be recommended roughly to the patient. The customized treatment theory in TCM indicates that the CMF should either continue or remove some of the herbs, according to special symptoms of the patient. Thus, the Symptom-CMF Recommendation needs more models to learn the relationship between symptom difference and CMF modification.

Table 1: Information for template design (2).

(a) TCM diseases

Stuffiness of liver
Vertigo

(b) WM diseases

Fatty liver
Hyperlipaemia
Hypertension

5. Example

After three years of observation of 11 TCM veteran practitioners, more than three thousand clinical records have been collected in ISMAC. Here, we take clinical data of fatty liver from Mr. Zhang Yunpeng, a known TCM liver specialist, as an example of demonstrating how to use ISMAC to manage and analyse clinical cases.

5.1. Template Design. The first step is to design the template. Usually, a template is designed by a specific TCM doctor and an expert team, as they are experienced in TCM diagnosis. Here, the itemset for a fatty liver template is supplied by Mr. Zhang's case research team, which has been conducting years of research on Mr. Zhang's clinical cases. The contents of the itemset are illustrated in Tables 1, 2, and 3.

5.2. Template Creation. With the summarized symptoms, syndromes, and therapeutic principles mentioned above, the corresponding IMA now can set up the template of Mr. Zhang for fatty Liver as the following steps.

Step 1. Select the entry *Inquiry and Diagnosis Template Maintenance* in the menu on the left side to go to its main page. Click the button *Add* to pop up the New Template Dialog.

Step 2. Assign the template a name; here *Fatty Liver* is entered. The serial number will be generated automatically.

Step 3. Input the name of TCM diseases or syndromes associated with this template. Notice that standard and expert-standard TCM disease have different ways of entry. Here in Mr. Zhang's template for fatty liver, expert-standard TCM diseases and syndromes are used, which are listed in Tables 1(a) and 2(a), respectively.

Table 2: Information for template design (3).

(a) TCM syndromes

Syndrome class number
Intermingled phlegm and blood stasis and retained dampness of heat toxin
Intermingled phlegm and blood stasis
Intermingled phlegm and blood stasis and ascendant hyperactivity of liver yang

(b) TCM therapeutic principle

Blood-activating and stasis-dissolving, dispersing stagnated liver qi and removing obstruction in the channels
Blood-activating and stasis-dissolving, clearing away heat and toxic materials
Blood-activating and stasis-dissolving, clearing toxic materials and resolving hard lump
Blood-activating and stasis-dissolving, calming Liver to stop endogenous wind
Blood-activating and stasis-dissolving, nourishing the heart and restoring the pulse

(i) Standard TCM diseases or syndromes: check the option *Use standard TCM disease*. Click the right button below to pop up the *Choose TCM Disease Dialog*. Select a disease from the *Disease Tree*.

(ii) Expert-standard TCM diseases or syndromes: uncheck the option *Use standard TCM disease*. Two new text input boxes will replace the original text box for the manual entry of the TCM disease or respective syndrome.

Step 4. Input the therapeutic principles of this template. The ways of entering standard and expert-standard therapeutic principle are also different. For this template, we enter expert-standard therapeutic methods as shown in Table 2(b).

(i) Standard therapeutic principles: check the option *Use Standard Therapeutic Principle*. Select an item from the popup *Choose Therapeutic Principle Dialog*.

(ii) Expert-standard therapeutic methods: uncheck the option *Use Standard Therapeutic Principle* and input the therapeutic principles manually.

Step 5. Input the standard or expert-standard WM diseases that this template deals with. The different ways of entering standard and expert-standard WM diseases are similar to those of the therapeutic principles mentioned above. For the template of fatty liver, we input the expert-standard WM diseases, which are shown in Table 1(b).

Step 6. Select the default dosage forms and default route of administration. The dosage forms might be injection, enema, decoction, and so forth. The administration routes could be "intravenous injection," "enema," "swallow," and so forth. As summarized by the clinical experience of Mr. Zhang, decoction and swallow are chosen as default forms and are routed, respectively.

Table 3: Information of template design for fatty liver (1).

(a) Main symptoms

Whether in first-visit
Tired or not
State of tongue
State of tongue coat
State of tongue proper
Pulse condition
Trigger of hypochondriac pain
Type of hypochondriac pain (left side)
Type of hypochondriac pain (right side)
Trigger of liver pain
Type of liver pain
Level of liver pain

(b) Estabilished symptoms

Nausea
Loose stool
Constipation
Sou huang
Backaches
Hline nai cha
Lusterless in face
Uncomfortable in epigastrium
Thirsty
Acid regurgitation
Yellow eyes
Abdominal distention
Depression
Sallow complexion
Nausea
Belch
Vomit
Edema of lower extremity
Bitter taste of mouth
Restless sleep at night

Step 7. Select the *First-visit Symptoms*, and *Examination Items*. The operation process is firstly, click the button *First-visit Symptoms*, *Return-visit Symptoms*, or *Examination Items* to switch the mode for data entry. Secondly, click the button *Add* to pop up the item selection dialog. Finally, check the needed items in the popup dialog. Here, for the template offatty liver, the items listed in Tables 3(a) and 3(b) are added

to both first-visit and return-visit tables, while there are no items for the examination table.

Step 8. Submit the form. A new template is created.

5.3. Data Collection and Analysis. Using the template created above, we have collected 63 clinical records for fatty liver. In the following, we will take syndrome classification and core CMF mining on the fatty liver data as an example of demonstrating how to use ISMAC for clinical case intelligent analysis.

5.3.1. Syndrome Classification of Fatty Liver Data. After years of research, we have discovered that the TCM clinical diagnoses have specific characteristics like multiple possible syndromes, the complex classification of diseases, imbalanced data distribution, and small sample size, which make the classification of TCM syndrome based on symptoms quite challenging. In response to these problems, we use the MAPLSC algorithm: a pattern classification algorithm proposed to settle the imbalanced multiclass problem, and build a model on fatty liver data and compare its performance with other state-of-the-art algorithms.

(a) Experiment Procedure. First, target dataset needs to be retrieved. In this task, all 63 cases are involved, but not all attributions are required. We want a model that can help us classify patients' syndromes through the analysis of symptoms and examinations. So in this target dataset, 32 items listed in Tables 3(a) and 3(b) act as feature attributions. Each record in the target dataset has one class label that indicates which syndrome the patient suffers. The type of syndrome can be one listed in Table 2(a).

Next, a comparative approach should be carefully designed. Our object is to check whether MAPLSC outperforms other state-of-the-art methods, as it analyzes imbalance and multiclass problems. In this case, the first question is which methods it should be compared with. After careful consideration, APLSC PWC, APLSC Vote, SVM Prob, SVM Vote, and J48 are chosen. APLSC PWC and APLSC Vote are used to check whether our method of extending the APLSC algorithm to a multiclass domain is the best one. SVM related multiclass algorithms are chosen, because SVM is very powerful in building a classification model which does not consider data imbalance. Therefore, comparisons between MAPLSC and others can demonstrate whether data imbalance should be taken into consideration when building a classification model. J48 is also a powerful and general classification method which can be included. The second question regards which criterion should be exploited in the experiment, including microaverage accuracy, macroaverage, and macro-$F1$-measure. More details about these algorithms and criterions can be referred to in the original paper [25].

Finally, the learning task is set up. All the six algorithms mentioned above are added in the Algorithm Maintenance module with specific parameters. And then these algorithms, the three criterions, and the target dataset are collected in a New Learning Task Dialog to create the target learning task.

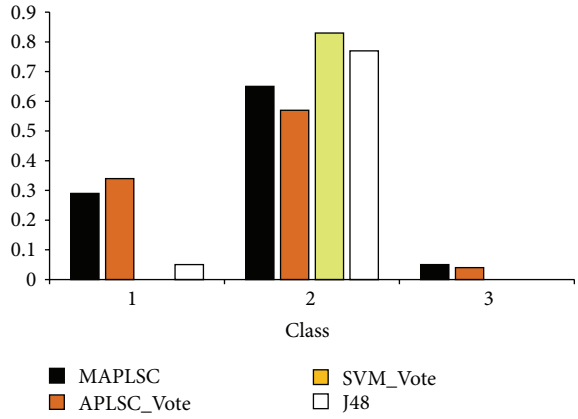


Figure 7: $F1$-measure of each class syndrome classification.

Cross-validation is also configured to report more convincing results.

(b) Experiment Results and Analysis. Table 4 illustrates the comparison results, from which we can observe that

(1) APLSC_PWC and APLSC_Vote are worse than MAPLSC on all the measurement rules.

(2) SVM related methods demonstrate poorer performance than APLSC related techniques on macro_average accuracy and macro_average $F1$-measure. However, higher micro_average accuracy values are recorded.

(3) J48 has no advantage on any metrics.

Observation (1) proves that our way of combining multiple classifiers is best. Observation (2) shows that MAPLSC is better with macro criterion, but worse with micro criterion. As macro criterion weighs equally all the classes while micro average weighs equally all the examples, MAPLSC can balance performance among classes. We draw Figure 7 to see performance distributions of different methods and different classes. In Figure 7, MAPLSC successfully classifies some examples of the difficult class 3. However, SVM Vote and J48 cannot identify any examples, and the $F1$-Measure diverges greatly among classes. This verifies that the MAPLSC has the ability to balance performance among classes. In the task of syndrome classification, MAPLSC performs well on almost all classes. And for some difficult class, the MAPLAC can also work.

5.3.2. Core Formula Mining on Fatty Liver Clinical Records. Referring to the TCM theory of Mr. Zhang, a core CMF for fatty liver therapy exists. From Section 4.2.3, we know that core CMF mining is very valuable. Occurrence frequency computing and Apriori are two popular algorithms for finding frequent patterns. The experimental procedure is similar to that of syndrome classification. Here, we omit the description of experimental procedure and elaborate the experiment results and analysis.

Table 4: Comparative results on syndrome classification.

Metrics	APLSC_Vote	MAPLSC	APLSC_PWC	SVM_Vote	SVM_Prob	SVM_PWC	J48
Macro avg_acc	38.33 ± 6.81	40.07 ± 6.69	38.74 ± 7.22	33.33 ± 0.00	32.79 ± 0.32	32.79 ± 0.32	30.11 ± 2.06
Micro avg_acc	45.78 ± 5.00	52.42 ± 5.49	51.8 ± 5.42	71.88 ± 0.00	70.70 ± 0.69	70.70 ± 0.69	62.73 ± 3.49
Macro avg_f1meas	34.73 ± 5.19	38.14 ± 5.57	37.13 ± 5.72	27.88 ± 0.00	27.61 ± 0.16	27.61 ± 0.16	27.36 ± 2.27

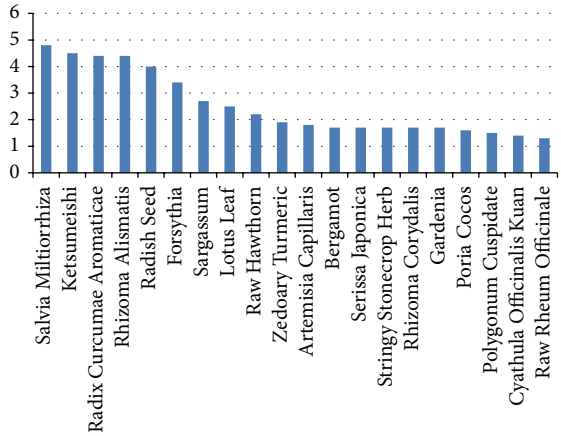


Figure 8: The first 20 most frequently used herbs and their normalized frequency.

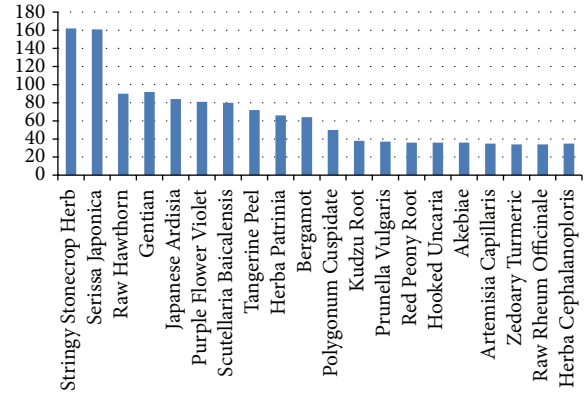


Figure 9: The first 20 most frequently used herbs and their normalized frequency except 8 herbs in basic core formula.

(a) Experiment Result and Analysis

(i) Core formula for fatty liver: using the *occurrence frequency computing* method and other visualization functions provided by ISMAC, the first 20 herbs most frequently used in the CMFs for fatty liver can be found in Figure 8. The first 8 herbs should be the basic core formula for fatty liver, which are *salvia miltiorrhiza*, *ketsumeishi*, *radix curcumae aromaticae*, *rhizoma alismatis*, *radish seed*, *forsythia*, *sargassum,* and *lotus leaf*.

(ii) Core formula for syndrome of intermingled phlegm and blood stasis and retained dampness heat toxin: dataset in this task contains only clinical records belonging to the syndrome of intermingled phlegm and blood stasis and retained dampness heat toxin. By removing the 8 herbs belonging to the basic core formula, there will be 137 herbs used for this syndrome. Once again, *occurrence frequency computing* is called to list the first 20 frequent herbs among 137 herbs. The result is illustrated in Figure 9, which shows us that the occurrence of the third herb *born hawthorn* goes down obviously compared with the first two herbs *stringy stonecrop* and *serissa japonica*. Thus, these two herbs are considered to be additional herbs of core formula for this syndrome.

Employing *Apriori* to find frequent patterns in this dataset, we get the results in Table 5. Agreeing with the result using *occurrence frequency computing*, the most frequent 3-itemset is far less than the most frequent 2-itemset as indicated in that table.

Table 5: The most frequent n-itemsets and their frequency.

Number of items	Frequency	The most frequent n-itemset
1	162	Stringy Stonecrop Herb
2	158	Stringy Stonecrop Herb, Serissa Japonica
3	89	Stringy Stonecrop Herb, Serissa Japonica, Gentian
4	56	Stringy Stonecrop Herb, Serissa Japonica, Gentian, Japanese Ardisia
5	37	Stringy Stonecrop Herb, Serissa Japonica, Hawthorn, Gentian, Japanese Ardisia
6	25	Stringy Stonecrop Herb, Serissa Japonica, Hawthorn, Gentian, Tangerine Peel, Japanese Ardisia

Synthetically taking results from these two methods into account, we draw a conclusion that *stringy stonecrop* and *serissa japonica*, together with the 8 herbs in basic core formula, constitute the core formula for the syndrome of intermingled phlegm and blood stasis and the retained dampness of heat toxin.

(iii) Core formula for syndrome of intermingled phlegm and blood stasis: with the same analysis method and procedure mentioned above, we find that the core formula for the syndrome of intermingled phlegm and blood stasis is *hawthorn*, *aloe*, and *snakegourd seed*.

(iv) Core formula for syndrome of intermingled phlegm and blood stasis and ascendant hyperactivity of liver

yang: using the same way of analysis, the core formula for the syndrome of intermingled phlegm and blood stasis and ascendant hyperactivity of liver yang is *patrinia*, *rhixoma gastrodiae*, *stringy stonecrop herb,* and *serissa japonica*.

6. Conclusion

The paper proposes an intelligent ISMAC System for TCM clinical case management and analysis. Customized Templates for data entry may be designed exactly according to the clinical diagnosis and treatment characteristics of a certain physician. State-of-the-art data analysis and mining methods have been integrated and improved to discover hidden knowledge. ISMAC is dedicated to TCM clinical case archiving, TCM theory verification, latent knowledge discovering, young doctor tutoring, and so forth. Having cooperated with many TCM veteran practitioners, ISMAC has collected more than three thousand clinical records, for which a great deal of clinical records mining has been conducted. And some of the data analysis results are verified in clinical examination. In the future, more clinical cases are necessary, and the customized management function of ISMAC will be improved by considering more clinical requirements and real-life problems. More data analysis methods should be experimented with for TCM clinical cases. For example, more than one syndrome would be tagged to a clinical diagnosis, indicating that syndrome prediction is a multilabel problem. We preliminarily attempted multilabel k nearest neighbor algorithm to analyze the syndrome classification of TCM coronary heart disease dataset [13], but more investigation is needed. What is more, thousands of symptoms may be carried in TCM clinical examination, symptoms contributing mostly to syndrome diagnosis are of vital concern. Selecting appropriate symptoms for clinical diagnosis may be modeled as a multilabel feature selection problem in machine learning. HOML may succeed in the selection of relevant symptoms by using wrapper methods. It is a time-consuming technique, and symptom reduction remains a challenging topic.

Conflict of Interests

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

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant nos. 61273305 and 81274007, as well as the Fundamental Research Funds for the Central Universities.

References

[1] C. Zhu, "Chen Zhu interview. China's modern medical minister. Interview by Richard Stone," *Science (New York, N.Y.)*, vol. 319, no. 5871, pp. 1748–1749, 2008.

[2] J.-L. Tang, B.-Y. Liu, and K.-W. Ma, "Traditional Chinese medicine," *The Lancet*, vol. 372, no. 9654, pp. 1938–1940, 2008.

[3] Anonymous, *The Inner Canon of Emperor Huang*, Chinese Medical Ancient Books Publishing House, Beijing, China, 2003.

[4] Chinese National Bureau of Statistics, "The 2008 National Social Statistical Data of China," Tech. Rep., 2008, http://www.stats.gov.cn/tjsj/ndsj/2009/indexch.htm.

[5] V. B. Konkimalla and T. Efferth, "Evidence-based Chinese medicine for cancer therapy," *Journal of Ethnopharmacology*, vol. 116, no. 2, pp. 207–210, 2008.

[6] J. K. Rao, K. Mihaliak, K. Kroenke, J. Bradley, W. M. Tierney, and M. Weinberger, "Use of complementary therapies for arthritis among patients of rheumatologists," *Annals of Internal Medicine*, vol. 131, no. 6, pp. 409–416, 1999.

[7] L. Wang, G.-B. Zhou, P. Liu et al., "Dissection of mechanisms of Chinese medicinal formula Realgar-Indigo naturalis as an effective treatment for promyelocytic leukemia," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 12, pp. 4826–4831, 2008.

[8] T.-T. Chang, M.-F. Sun, H.-Y. Chen et al., "Screening from the world's largest TCM database against H1N1 virus," *Journal of Biomolecular Structure & Dynamics*, vol. 28, no. 5, pp. 773–786, 2011.

[9] H.-C. Diener, K. Kronfeld, G. Boewing et al., "Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial," *Lancet Neurology*, vol. 5, no. 4, pp. 310–316, 2006.

[10] X. Zhou, S. Chen, B. Liu et al., "Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support," *Artificial Intelligence in Medicine*, vol. 48, no. 2-3, pp. 139–152, 2010.

[11] X. Zhou, Y. Peng, and B. Liu, "Text mining for traditional Chinese medical knowledge discovery: a survey," *Journal of Biomedical Informatics*, vol. 43, no. 4, pp. 650–660, 2010.

[12] S. K. Poon, J. Poon, M. McGrane et al., "A novel approach in discovering significant interactions from TCM patient prescription data," *International Journal of Data Mining and Bioinformatics*, vol. 5, no. 4, pp. 353–368, 2011.

[13] G.-P. Liu, G.-Z. Li, Y.-L. Wang, and Y.-Q. Wang, "Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning," *BMC Complementary and Alternative Medicine*, vol. 10, article 37, 2010.

[14] J. Tian, J. Xue, Y. Dai, J. Chen, and J. Zheng, "A novel software platform for medical image processing and analyzing," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 6, pp. 800–812, 2008.

[15] M. Masseroli and M. Marchente, "X-PAT: a multiplatform patient referral data management system for small healthcare institution requirements," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, pp. 424–432, 2008.

[16] H.-H. Ku and C.-M. Huang, "Web2OHS: a Web2.0-based omnibearing homecare system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 224–233, 2010.

[17] H. Chen, Z. Wu, and C. Huang, "TCM-Grid: weaving a medical grid for traditional Chinese medicine," in *Computational Science—ICCS 2003*, pp. 1143–1152, Springer, 2003.

[18] X. Zhou, Z. Wu, and W. Lu, "TCMMDB: a distributed multidatabase query system and its key technique implemention," in *Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace*, vol. 2, pp. 1095–1100, IEEE, 2001, (Cat.No.01CH37236).

[19] M. You, G. Z. Li, X. Q. Zeng et al., "A personalized traditional Chinese medicine system in the case of Cai's gynecology," *International Journal of Functional Informatics and Personalised Medicine*, vol. 1, no. 4, p. 419, 2008.

[20] M. You, S.-X. Yan, G.-Z. Li et al., "Customized management of clinical data in traditional Chinese medicine," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM '13)*, pp. 292–293, IEEE, Shanghai, China, December 2013.

[21] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[22] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, vol. 1215, pp. 487–499, Santiago, Chile, 1994.

[23] M. You, Y. Chen, and G. Li, "Im-IG: a novel feature selection method for imbalanced problems," *Journal of Shangdong University (Engineering Science)*, vol. 40, no. 5, pp. 123–128, 2010.

[24] H.-N. Qu, G.-Z. Li, and W.-S. Xu, "An asymmetric classifier based on partial least squares," *Pattern Recognition*, vol. 43, no. 10, pp. 3448–3457, 2010.

[25] M. You, R. W. Zhao, G. Z. Li, and X. Hu, "MAPLSC: a novel multi-class classifier for medical diagnosis," *International Journal of Data Mining and Bioinformatics*, vol. 5, no. 4, p. 383, 2011, http://www.inderscience.com/link.php?id=41555.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 161439, 8 pages
http://dx.doi.org/10.1155/2015/161439

Clinical Study

Acupuncture for Vascular Dementia: A Pragmatic Randomized Clinical Trial

Guang-Xia Shi,[1] Qian-Qian Li,[1] Bo-Feng Yang,[2] Yan Liu,[2] Li-Ping Guan,[2] Meng-Meng Wu,[2] Lin-Peng Wang,[1] and Cun-Zhi Liu[1,2]

[1] *Acupuncture and Moxibustion Department, Beijing Hospital of Traditional Chinese Medicine Affiliated to Capital Medical University, 23 Meishuguanhou Street, Dongcheng District, Beijing 100010, China*
[2] *Acupuncture and Moxibustion Research Institute, The First Hospital Affiliated to Tianjin College of Traditional Chinese Medicine, 314 West Anshan Avenue, Tianjin 300193, China*

Correspondence should be addressed to Cun-Zhi Liu; lcz623780@126.com

Received 19 August 2014; Revised 1 September 2014; Accepted 26 September 2014

Academic Editor: Zhaohui Liang

In this trial, patients who agreed to random assignment were allocated to a randomized acupuncture group (R-acupuncture group) or control group. Those who declined randomization were assigned to a nonrandomized acupuncture group (NR-acupuncture group). Patients in the R-acupuncture group and NR-acupuncture group received up to 21 acupuncture sessions during a period of 6 weeks plus routine care, while the control group received routine care alone. Cognitive function, activities of daily living, and quality of life were assessed by mini-mental state examination (MMSE), Activities of Daily Living Scale (ADL), and dementia quality of life questionnaire (DEMQOL), respectively. All the data were collected at baseline, after 6-week treatment, and after 4-week follow-up. No significant differences of MMSE scores were observed among the three groups but pooled-acupuncture group had significant higher score than control group. Compared to control group, ADL score significantly decreased in NR-acupuncture group and pooled-acupuncture group. For DEMQOL scores, no significant differences were observed among the three groups, as well as between pooled-acupuncture group and control group. Additional acupuncture to routine care may have beneficial effects on the improvements of cognitive status and activities of daily living but have limited efficacy on health-related quality of life in VaD patients.

1. Introduction

Vascular dementia (VaD) is thought to be resulted from various types of ischemic and hemorrhagic brain lesions which lead to intellectual and physical disability [1]. It is the second most common cause of dementia among the older people, representing 15–25% of all cases worldwide [2–4]. Projections indicate that progressive ageing of the populations and the high prevalence of cerebrovascular and cardiovascular pathologies capable of producing VaD are the most common form of dementia [5, 6]. This creates a difficult situation for those suffering from VaD, their caregivers, and healthcare providers. However, definitive medical or surgical treatments do not exist so far [7, 8].

Maintaining or improving quality of life in people with VaD is currently a key outcome of health services and the increasing number of psychosocial interventions targeting this population [9]. Being not able to perform basic activities of daily living reduces quality of life in dementia [10], including bathing, continence, dressing, feeding, toileting, and transfer. Since the need for assistance in these activities makes people with VaD dependent on informal (family) or formal (professional) care, quality of life is one of the most important outcomes in improving well-being of people with VaD [11, 12].

Acupuncture is a core component in traditional Chinese medicine (TCM) and can be traced back more than 3000 years in China. It is often used as a treatment for dementia

[13, 14]. Nevertheless, recent systematic review has shown inconclusive evidence because of low methodological quality [13]. The effectiveness of acupuncture for VaD has not been fully established. The majority of previous trials for acupuncture were designed for experimental studies [15]. Hence, there is currently very little information about the effectiveness of acupuncture in general medical practice.

Based on our previous study [16], we designed the present study as a pragmatic trial to investigate the effectiveness of acupuncture in addition to routine care among patients with VaD. In addition, we examined whether the effects of acupuncture differ between randomized and nonrandomized patients.

2. Experimental

2.1. Methods. Patients were eligible for this trial only after they had met rigorous criteria for probable VaD as defined by the National Institute of Neurological Disorders and Stroke-AIREN criteria (NINDS-AIREN). Inclusion criteria also included a score of 0 to 23 on the mini-mental state examination (MMSE), disease duration of more than 2 months, onset of the disease at age less than 80 years, and the availability of a reliable caregiver. All subjects included in the trial were evaluated with the Hachinski Ischemic Score (HIS). A HIS of $\geqslant 7$ has been validated in an autopsy study as an accurate indicator of VaD [17]. Participants were excluded if they had a prior diagnosis of Alzheimer's disease (AD), Parkinson disease, Huntington disease, and other neurodegenerative dementia. Participants with less than secondary education were ineligible in the present study. Exclusion criteria also included the presence of abnormal executive control function, severe enough to interfere with social or occupational functioning and inability to give consent because of impaired cognition or receptive aphasia. All patients have undergone a neurologic interview to determine their history of onset, symptoms, and recovery from stroke in the first interview.

We performed the study according to common guidelines for clinical trials (Declaration of Helsinki, International Conference on Harmonisation (ICH)/WHO Good Clinical Practice standards (GCP) including certification by an external audit). The trial protocol has been approved by the Research Ethical Committee of The First Hospital affiliated to Tianjin College of Traditional Chinese Medicine (20073055). All study participants provided written, informed consent.

2.2. Randomization. Patients who agreed to be randomly assigned were allocated to a randomized acupuncture group (R-acupuncture group) or control group. Block randomization with a block size of 4 was by sequential, sealed, opaque envelopes. It occurred after the acupuncturist's evaluation (concealed allocation) using a computer-generated, random-allocation sequence (random list generated with SAS 8.2). Furthermore, participants who did not consent to randomization were assigned to a nonrandomized acupuncture group (NR-acupuncture group). We ensured that the patients, data collection staff, and data analysts were blinded during the study period; they were all unaware of the randomization. The acupuncturists were not blinded to the treatments they delivered because acupuncture manipulation made this impossible. During the intervention, acupuncturist and the personnel who collected data were segregated by an opaque screen immediately after the treatment started and were instructed not to exchange information with each other.

2.3. Intervention. Patients in the R-acupuncture group and NR-acupuncture group received acupuncture treatment plus routine care, while those in the control group received routine care alone. Routine care here refers to the use of certain medications. These medicines include antiplatelet agents (aspirin or ticlopidine), antihypertensive, diuretics, and nimodipine and should be taken following the advice of physician. In addition, they received a weekly phone call to inquire of their health status to provide individual attention. Acupuncture was administrated by 5 therapists with more than 6 years of experience and a Chinese medicine practitioner license from the Ministry of Health of the People's Republic of China. Based on the TCM theory, the main acupoints we selected were as follows: GV20 (baihui), EX-HN1 (sishencong), GV24 (shenting), CV17 (tanzhong), PC6 (neiguan), CV12 (zhongwan), CV6 (qihai), SP10 (xuehai), and ST36 (zusanli). Moreover, the following acupoints could be added as auxiliary acupoints: GB 20 (fengchi), ST40 (fenglong), LR3 (taichong), SP6 (sanyinjiao), and ST25 (tianshu). The acupuncture point prescriptions used were individualized to each patient and were at the discretion of the acupuncturist. Acupuncture was performed by means of standard stainless-steel needles ($0.25\,\Phi \times 25$ mm and $0.25\,\Phi \times 40$ mm, Beijing Hanyi Medical Instrument Center) and manually stimulated to elicit needle sensation (de qi). The treatment consisted of 21 sessions of 30 minutes' duration, and each was administered once every other day over a period of 6 weeks.

2.4. Measurements. Demographic measures collected at the baseline evaluation included age, gender, common complications of VaD, HIS, and outcome variables.

Cognitive status, including orientation, memory, calculation, language, and constructional apraxia, of the VaD patients was assessed using the MMSE. Total scores for this measure range from 0 to 30, with lower scores indicating lower cognitive functioning.

Activities of daily living were determined by Activity of Daily Living Scales (ADL). It has been developed specifically for use with VaD, consisting of 20 daily-living abilities, where higher scores indicate lower levels of activities of daily living (scores range from 0 to 60).

Health-related quality of life is measured by dementia quality of life questionnaire (DEMQOL). It is a tool with which to evaluate whether the interventions and services achieve this. It covers five domains of quality of life and uses both self-reporting and rating by family guardian or staff member as proxy. Higher scores indicate better quality of life and vice versa. It has good internal consistency, interior reliability, and concurrent validity and can generate a measure of utility.

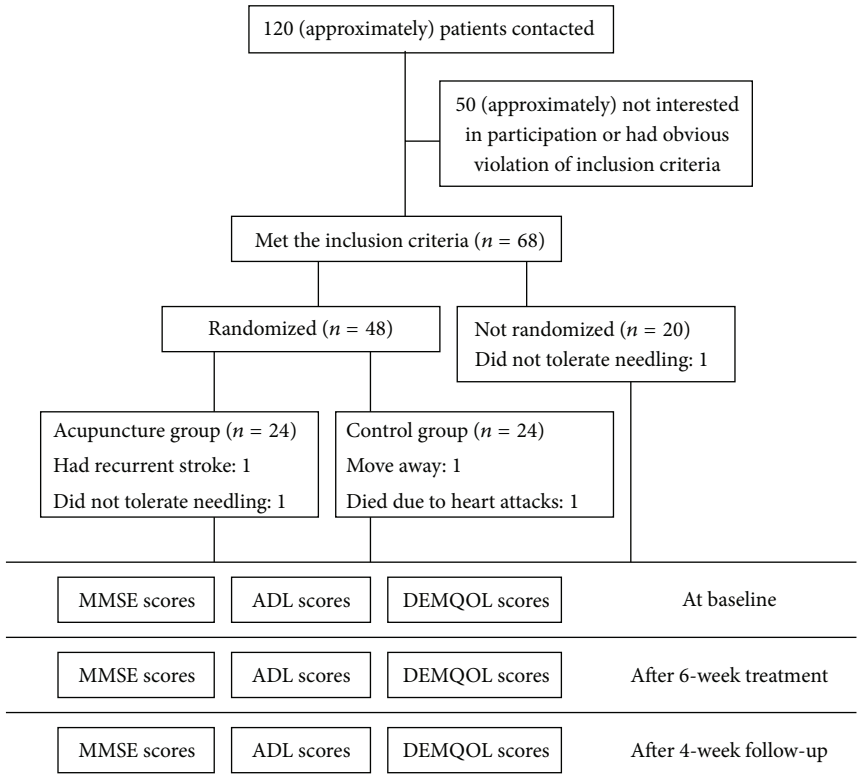


Figure 1: Trial flow chart.

All the outcomes were assessed at baseline, after 6-week treatment and after 4-week follow-up.

2.5. Sample Size Calculation. According to previous study, there was a significant difference between the pre- and post-treatment for the treatment group than that for the control group with an increase of 4.27 ± 2.05. We calculated the number of sample size, using the following formula:

$$n_1 = n_2 = 2\left[\frac{\left(t_\alpha + t_\beta\right)s}{\delta}\right]^2 \quad (1)$$

$$\left(\alpha = 0.05,\ \beta = 0.10\right) \quad \delta = 1.84, \quad s = 1.56.$$

As a result, we estimated that 20 patients were required in each group. Assuming a 20% drop-out rate, we planned to randomize 24 patients to each arm.

2.6. Statistical Analysis. A per protocol analysis was done based on patients with no major protocol violations by the end of 4-week follow-up after randomization. Data are expressed as mean ± standard deviation (SD), or frequencies and percentages, according to the type of variable. For the between-group comparisons at baseline, either the Chi-Square test or one-way ANOVA was used. An analysis of covariance with additional covariate of age was performed to account for potential baseline differences. Repeated-measures analysis of variance was used to determine whether significant differences exist across time. Furthermore, results were evaluated by pooling all patients who received acupuncture treatment into one group. Pooled-acupuncture group which actually contained patients in R-acupuncture group and NR-acupuncture group were compared with the control group. P values lower than 0.05 were considered to be statistically significant. Statistical analysis was performed with SPSS version 10.0 (SPSS Inc., Chicago, Illinois, USA).

3. Result

3.1. Patient Enrollment. In the present population-based study, patients managed in the community as well as those managed in hospital (Tianjin, china) were recruited from June 2007 to February 2010. Figure 1 showed the flow of participants through the trial. Of 120 screened patients, 52 could not be included in the study, mainly because they did not meet all eligibility criteria. A total of 68 patients were included. Of these, 24 were randomly assigned to R-acupuncture and 24 to control group. 20 patients who declined randomization were allocated to NR-acupuncture group. Five patients dropped out during the trial, accounting for a 7.4% dropout rate. Among the 5 dropouts, 1 patient had recurrent stroke, 1 patient died because of heart attacks, 2 patients did not tolerate needling, and 1 patient withdrew due to move from one city to another. Thus, the protocol analysis included 63 patients.

3.2. Analysis of Baseline Data. The demographic and clinical features at baseline are shown in Table 1. Three groups

Table 1: Subjects' baseline characteristics.

	R-acupuncture ($n = 22$)	NR-acupuncture ($n = 19$)	Control ($n = 22$)
Gender n (%)			
Men n (%)	12 (54.55)	6 (31.58)	11 (50.00)
Women n (%)	10 (45.45)	13 (68.42)	11 (50.00)
Age (M ± SD, y)	67.24 ± 9.33*	57.21 ± 10.80*	67.45 ± 10.14*
Hypertension, n (%)	17 (77.27)	17 (89.47)	15 (68.18)
Diabetes mellitus, n (%)	2 (9.10)	4 (21.05)	4 (18.18)
Coronary heart disease, n (%)	9 (40.91)	7 (36.84)	8 (36.37)
HIS (M ± SD)	9.55 ± 2.37	11.32 ± 2.16	11.00 ± 2.14
MMSE (M ± SD)	18.27 ± 4.08	17.74 ± 3.33	17.77 ± 3.99
ADL (M ± SD)	42.91 ± 13.97	43.84 ± 11.42	49.86 ± 14.97
DEMQOL (M ± SD)	72.41 ± 7.02	64.68 ± 6.68	72.27 ± 9.23

*Differences among the three groups are statistically significant $P < 0.05$. R-acupuncture = randomized acupuncture group; Control = control group; NR-acupuncture = nonrandomized acupuncture group; HIS = Hachinski Ischemic Score; MMSE = mini-mental state examination; ADL = activities of daily living; DEMQOL = health-related quality of life.

were comparable with regard to most baseline characteristics except for age. Patients in the randomization group were, on average, younger than those in the other two groups ($P = 0.003$).

3.3. Analysis of Outcome Variables. Repeated-measures analysis of variance on MMSE scores revealed a time effect ($P = 0.034$) and a treatment × time interaction ($P = 0.001$), indicating a favorable improvement in the cognitive evolution of VaD individuals as the extension of time (Figure 2(a)). No significant differences of MMSE were observed among the three groups. However, pooled-acupuncture group had significant higher score than control group ($P = 0.014$) (Figure 2(b)).

Figure 3(a) indicated no time effect ($P = 0.241$), but a treatment effect ($P = 0.027$) and treatment × time interaction ($P = 0.014$) on ADL score. Lower score was observed in the NR-acupuncture group compared to the control group ($P = 0.01$). Nonetheless, no significant differences were detected in R-acupuncture group versus NR-acupuncture group ($P = 1.00$), as well as R-acupuncture group versus control group ($P = 0.067$). In addition, lower score was found in pooled-acupuncture group compared to the control group ($P = 0.003$) (Figure 3(b)).

Figure 4 showed a time effect ($P = 0.000$) and treatment × time interaction ($P = 0.011$) and no treatment effect ($P = 0.05$) on DEMQOL score, indicating a significant improvement on health-related quality of life across time. However, no significant differences were observed among the three groups (Figure 4(a)). Besides, no significant differences existed between the pooled-acupuncture group and control group ($P = 0.283$) (Figure 4(b)).

3.4. Safety of Acupuncture. During the acupuncture treatment, 25% experienced discomfort at the sites of needle insertion or simulated needle insertion, and 20% had bruising. No serious adverse events were documented.

4. Discussion

The aim of our study was to examine whether acupuncture has additional value in patients with VaD compared to treatment with routine care alone. Results indicate that, compared to patients in control group, those in pooled-acupuncture group showed significant improvements in cognitive status and activities of daily living. Moreover, patients who declined randomization and therapeutic outcomes after acupuncture were better than those who consented to. Additional acupuncture to routine care is of limited efficacy in VaD patients whose health-related quality of life has already deteriorated.

Recent international policy guidelines aim to promote independence in dementia and show a rising interest in how nonpharmacological interventions could help maintain everyday functional independence as long as possible [18]. This information could contribute to shaping interventions to help VaD remain at home as long as possible and maintain a good health-related quality of life.

There is growing evidence implicating dementia is a risk factor for stroke and stroke is associated with an increased risk of subsequent dementia [2, 19, 20]. Although most stroke survivors go on to show some improvement over time, a large percentage eventually develops significant symptoms of dementia [21, 22]. Acupuncture is frequently advocated as an adjunct treatment during stroke rehabilitation. However, controversy remains regarding the effectiveness of acupuncture for recovery in activities of daily living and health-related quality of life after stroke [19, 23]. A recent meta-analysis of data from rigorous randomized sham-controlled trials did not show a positive effect of acupuncture as a treatment for functional recovery after stroke [14]. Park et al. [24] reported that acupuncture is not superior to sham treatment. Nonetheless, Moffet [25] argued that even if acupuncture is ineffective for stroke itself, it may nonetheless be helpful in treating pain, sleep disturbance, anxiety, depression, or other conditions that are common consequence of stroke and that are often barriers to rehabilitation and recovery.

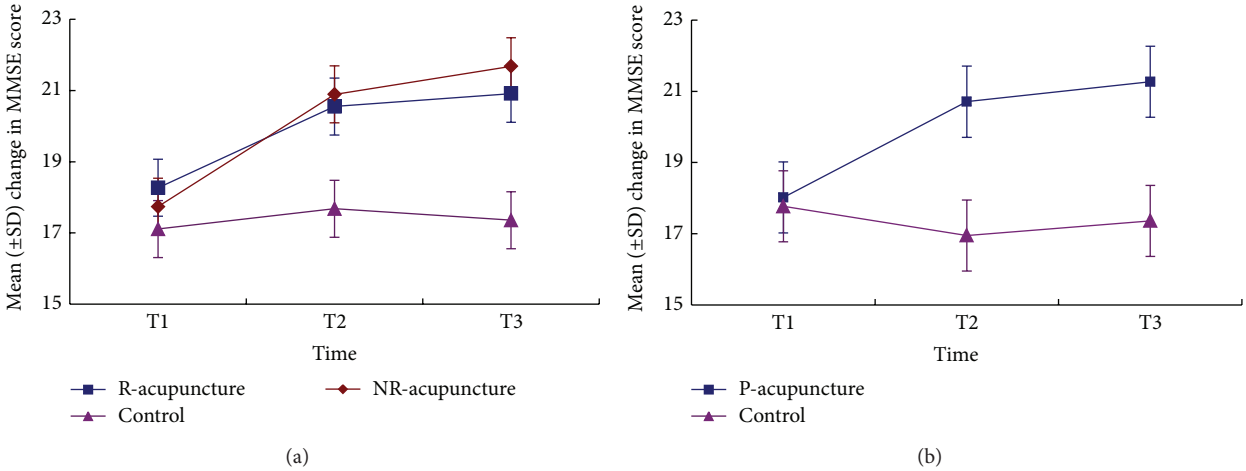


Figure 2: The changes of score in MMSE across time. Comparison among R-acupuncture group, NR-acupuncture group, and control group (a) and between pooled-acupuncture group and the control group (b). R-acupuncture = randomized acupuncture group, Control = control group, NR-acupuncture = nonrandomized acupuncture group, and P-acupuncture = pooled-acupuncture group. T1 = at baseline, T2 = after 6-week treatment, and T3 = after 4-week follow-up.

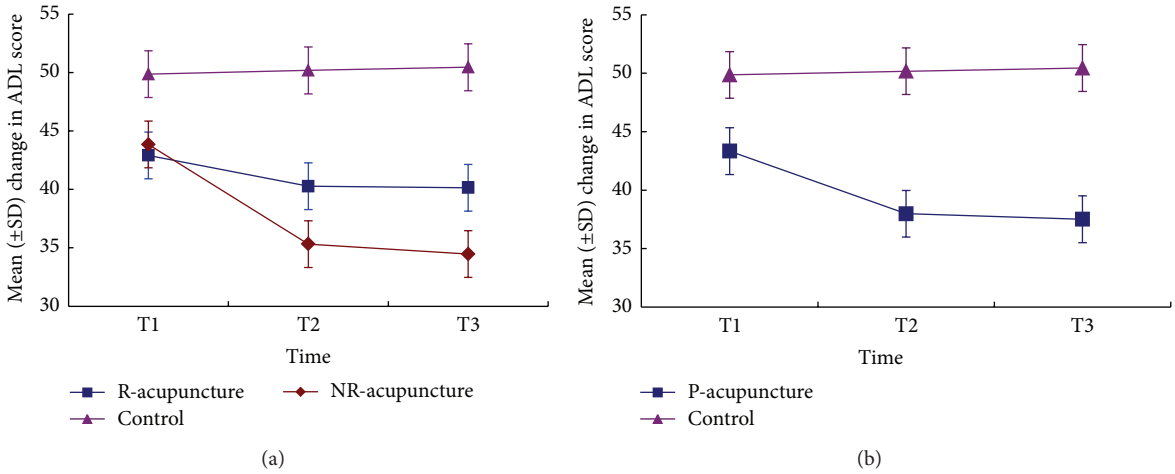


Figure 3: The changes of score in ADL across time. Comparison among R-acupuncture group, NR-acupuncture group, and control group (a) and between pooled-acupuncture group and the control group (b). R-acupuncture = randomized acupuncture group, Control = control group, NR-acupuncture = nonrandomized acupuncture group, P-acupuncture = pooled-acupuncture group. T1 = at baseline, T2 = after 6-week treatment, and T3 = after 4-week follow-up.

Kwok et al. [26] showed that acupuncture elicited significant improvement in sleep quality of elders with dementia in terms of significant gain in resting time as well as sleep time in the treatment period over the control period. Preliminary searches revealed more than 105 studies of acupuncture for treating vascular dementia. Benefit was reported in up to 70 to 91% of the treatment group. However, one review in 2007 suggested there is currently no evidence available from sufficiently high quality randomized controlled trials to allow assessment of the efficacy of acupuncture in the treatment of vascular dementia [27]. Our results showed a beneficial effect of additional acupuncture on cognitive status and activities of daily living for VaD patients.

We took a pragmatic approach, aiming to evaluate acupuncture in a manner that would reflect as closely as possible the conditions of daily medical practice. The additional inclusion of patients who declined randomization allowed us to investigate any potential selection effects. One explanation for the significant improvements in NR-acupuncture group may attribute to potent placebo effects. Previous studies have shown that expectancy, a crucial component of placebo, plays an important role in acupuncture treatment efficacy [28]. Linde et al. [29] suggested that expectancy is able to enhance acupuncture analgesia initiated by an inhibited brain response to calibrated pain stimuli. They showed strong association between better improvement and higher expectations by a pooled analysis of four randomized controlled trials of acupuncture in patients with chronic pain. Patients refusing randomization tend to believe that they were assuring to get an active or a more effective intervention.

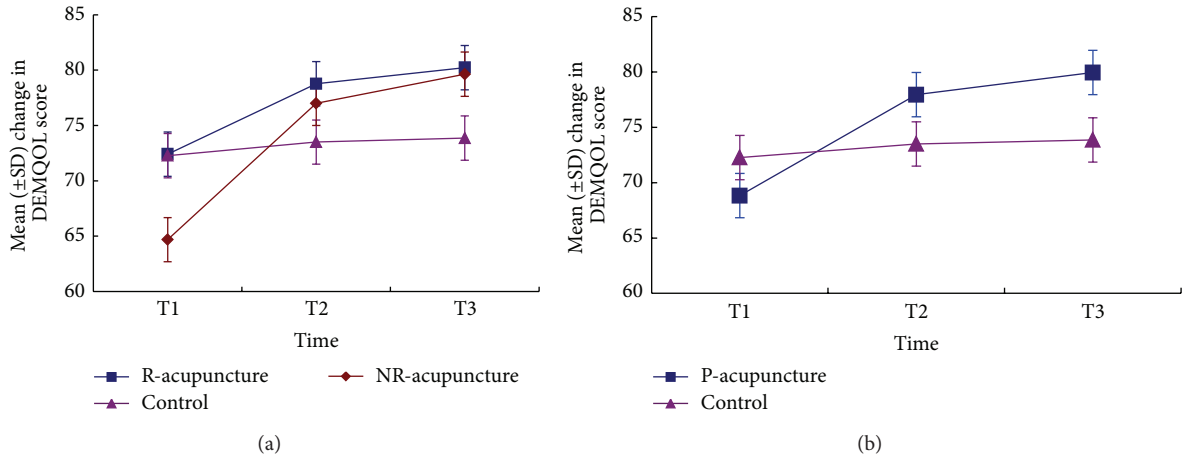


Figure 4: The changes of score in DEMQOL across time. Comparison among R-acupuncture group, NR-acupuncture group, and control group (a) and between pooled-acupuncture group and the control group (b). R-acupuncture = randomized acupuncture group, Control = control group, NR-acupuncture = nonrandomized acupuncture group, and P-acupuncture = pooled-acupuncture group. T1 = at baseline, T2 = after 6-week treatment, and T3 = after 4-week follow-up.

Hence, patient expectations may play a more prominent role than those of patients receiving randomization.

Treatment of VaD also includes prevention and attenuation of potential risk factors [4, 30]. There is evidence that the treatment of modifiable vascular risk factors, such as hypertension, diabetes mellitus, hypercholesterolemia, and heart disease, is likely to slow the progress of cognitive decline and a reduction of the risk of dementia [31, 32]. Park et al. indicating that the beneficial effect of acupuncture could be of clinical importance to prevent the progression of cardiovascular diseases [33]. The main finding in the present study was that additional acupuncture treatment was efficacious in reducing and controlling clinical syndrome associated with VaD. Therefore, future researches should provide answers regarding the beneficial effect of acupuncture on modifying risk factors of VaD.

In a pragmatic trial, it is not usually appropriate to use a placebo control and blinding, as these are likely to have a detrimental effect on the trial's ecological validity [34]. So, we compared the effect of acupuncture with another treatment, not with a placebo and the sham acupuncture. Our study has some limitations. Firstly, the treatment time is relatively short. This study demonstrated that the 6-week treatment of acupuncture in VaD patients improves cognitive function and activities of daily living but did not change the health-related quality of life. The potential role of acupuncture treatment for long-term therapy has not been examined. Further studies will be necessary to demonstrate whether long-term acupuncture treatment can sustain the improvement. In addition, we consider that a longer period of follow-up could be necessary to investigate the optimum timing for such an acupuncture treatment and to assess the value of repeated courses of acupuncture for patients experiencing VaD. Secondly, because the recruitment was limited due to administrative factors, our results need to be further investigated in the future studies with larger number of patients. Another potential limitation is that whether acupuncture had an independent beneficial effect on VaD remains unexplained in the present trial.

5. Conclusion

Our study shows that additional acupuncture to routine care may have beneficial effects on the improvements of cognitive status and activities of daily living but have limited efficacy on health-related quality of life in VaD patients.

Abbreviations

VaD:	Vascular dementia
R-acupuncture group:	Randomized acupuncture group
NR-acupuncture group:	Nonrandomized acupuncture group
MMSE:	Mini mental state examination
ADL:	Ability of daily life scale
DEMQOL:	Dementia quality of life questionnaire
TCM:	Traditional Chinese medicine
HIS:	Hachinski Ischemic Score
AD:	Alzheimer's disease
SD:	Standard deviation.

Conflict of Interests

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

Acknowledgments

The authors thank the research staff at The First Hospital affiliated to Tianjin College of Traditional Chinese Medicine (Wei Guan and Chuan Xiao), The Second Hospital affiliated to Tianjin College of Traditional Chinese Medicine (Zhan-kui Wang and Hui Shi), Tianjin Hospital of Traditional Chinese

Medicine (Lei Wang), and Tiantuo community Hospital (Hong Zhu). The study was supported by Program for New Century Excellent Talents in University (NCET-09-0007), Technology New Star Program of Beijing (2009B46), and the Bureau of Public Health of Tianjin (07059).

References

[1] A. D. Korczyn, V. Vakhapova, and L. T. Grinberg, "Vascular dementia," *Journal of the Neurological Sciences*, vol. 322, no. 1-2, pp. 2–10, 2012.

[2] M. E. Murray, D. S. Knopman, and D. W. Dickson, "Vascular dementia: clinical, neuroradiologic and neuropathologic aspects," *Panminerva Medica*, vol. 49, no. 4, pp. 197–207, 2007.

[3] G. C. Román, "Facts, myths, and controversies in vascular dementia," *Journal of the Neurological Sciences*, vol. 226, no. 1-2, pp. 49–52, 2004.

[4] L. Battistin and A. Cagnin, "Vascular cognitive disorder. A biological and clinical overview," *Neurochemical Research*, vol. 35, no. 12, pp. 1933–1938, 2010.

[5] D. E. Barnes, K. Yaffe, A. L. Byers, M. McCormick, C. Schaefer, and R. A. Whitmer, "Midlife vs late-life depressive symptoms and risk of dementia," *Archives of General Psychiatry*, vol. 69, no. 5, pp. 493–498, 2012.

[6] K. A. Jellinger and J. Attems, "Is there pure vascular dementia in old age?" *Journal of the Neurological Sciences*, vol. 299, no. 1-2, pp. 150–154, 2010.

[7] R. Sahathevan, A. Brodtmann, and G. A. Donnan, "Dementia, stroke, and vascular risk factors; a review," *International Journal of Stroke*, vol. 7, no. 1, pp. 61–73, 2012.

[8] S. Ankolekar, C. Geeganage, P. Anderton, C. Hogg, and P. M. W. Bath, "Clinical trials for preventing post stroke cognitive impairment," *Journal of the Neurological Sciences*, vol. 299, no. 1-2, pp. 168–174, 2010.

[9] E. Moniz-Cook, M. Vernooij-Dassen, R. Woods et al., "A European consensus on outcome measures for psychosocial intervention research in dementia care," *Aging and Mental Health*, vol. 12, no. 1, pp. 14–29, 2008.

[10] H. C. Beerens, S. M. G. Zwakhalen, H. Verbeek, D. Ruwaard, and J. P. H. Hamers, "Factors associated with quality of life of people with dementia in long-term care facilities: a systematic review," *International Journal of Nursing Studies*, vol. 50, no. 9, pp. 1259–1270, 2013.

[11] J. L. Conde-Sala, J. Garre-Olmo, O. Turró-Garriga, S. López-Pousa, and J. Vilalta-Franch, "Factors related to perceived quality of life in patients with Alzheimer's disease: the patient's perception compared with that of caregivers," *International Journal of Geriatric Psychiatry*, vol. 24, no. 6, pp. 585–594, 2009.

[12] J. L. Conde-Sala, J. Garre-Olmo, O. Turró-Garriga, J. Vilalta-Franch, and S. López-Pousa, "Quality of life of patients with Alzheimer's disease: differential perceptions between spouse and adult child caregivers," *Dementia and Geriatric Cognitive Disorders*, vol. 29, no. 2, pp. 97–108, 2010.

[13] W. N. Peng, H. Zhao, Z. S. Liu, and S. Wang, "Acupuncture for vascular dementia," *The Cochrane Database of Systematic Reviews*, no. 2, Article ID CD004987, 2007.

[14] M. S. Lee, B.-C. Shin, and E. Ernst, "Acupuncture for Alzheimer's disease: a systematic review," *International Journal of Clinical Practice*, vol. 63, no. 6, pp. 874–879, 2009.

[15] J. C. Kong, M. S. Lee, B.-C. Shin, Y.-S. Song, and E. Ernst, "Acupuncture for functional recovery after stroke: a systematic review of sham-controlled randomized clinical trials," *CMAJ*, vol. 182, no. 16, pp. 1723–1729, 2010.

[16] J. Yu, X. Zhang, C. Liu, Y. Meng, and J. Han, "Effect of acupuncture treatment on vascular dementia," *Neurological Research*, vol. 28, no. 1, pp. 97–103, 2006.

[17] J. T. Moroney, E. Bagiella, D. W. Desmond et al., "Meta-analysis of the Hachinski ischemic score in pathologically verified dementias," *Neurology*, vol. 49, no. 4, pp. 1096–1105, 1997.

[18] S. T. Pendlebury, "Stroke-related dementia: rates, risk factors and implications for future research," *Maturitas*, vol. 64, no. 3, pp. 165–171, 2009.

[19] S. T. Pendlebury and P. M. Rothwell, "Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis," *The Lancet Neurology*, vol. 8, no. 11, pp. 1006–1018, 2009.

[20] R. F. Gottesman and A. E. Hillis, "Predictors and assessment of cognitive dysfunction resulting from ischaemic stroke," *The Lancet Neurology*, vol. 9, no. 9, pp. 895–905, 2010.

[21] J. B. Pinkston, N. Alekseeva, and E. G. Toledo, "Stroke and dementia," *Neurological Research*, vol. 31, no. 8, pp. 824–831, 2009.

[22] A. S. Creighton, E. S. van der Ploeg, and D. W. O'Connor, "A literature review of spaced-retrieval interventions: a direct memory intervention for people with dementia," *International Psychogeriatrics*, vol. 25, no. 11, pp. 1743–1763, 2013.

[23] D. J. Mayer, "Acupuncture: an evidence-based review of the clinical literature," *Annual Review of Medicine*, vol. 51, pp. 49–63, 2000.

[24] J. Park, A. R. White, M. A. James et al., "Acupuncture for subacute stroke rehabilitation: a sham-controlled, subject- and assessor-blind, randomized trial," *Archives of Internal Medicine*, vol. 165, no. 17, pp. 2026–2031, 2005.

[25] H. H. Moffet, "Acupuncture may be ineffective for stroke," *Archives of Internal Medicine*, vol. 166, no. 8, article 930, 2006.

[26] T. Kwok, P. C. Leung, Y. K. Wing et al., "The effectiveness of acupuncture on the sleep quality of elderly with dementia: a within-subjects trial," *Clinical Interventions in Aging*, vol. 8, pp. 923–929, 2013.

[27] W. N. Peng, H. Zhao, Z. S. Liu, and S. Wang, "Acupuncture for vascular dementia," *Cochrane Database of Systematic Reviews*, no. 2, Article ID CD004987, 2007.

[28] C. Paterson and P. Dieppe, "Characteristic and incidental (placebo) effects in complex interventions such as acupuncture," *British Medical Journal*, vol. 330, no. 7501, pp. 1202–1205, 2005.

[29] K. Linde, C. M. Witt, A. Streng et al., "The impact of patient expectations on outcomes in four randomized controlled trials of acupuncture in patients with chronic pain," *Pain*, vol. 128, no. 3, pp. 264–271, 2007.

[30] K. Shah, S. U. Qureshi, M. Johnson, N. Parikh, P. E. Schulz, and M. E. Kunik, "Does use of antihypertensive drugs affect the incidence or progression of dementia? A systematic review," *The American Journal of Geriatric Pharmacotherapy*, vol. 7, no. 5, pp. 250–261, 2009.

[31] E. Martínez-Vila, M. Murie-Fernández, J. G. Pérez-Larraya, and P. Irimia, "Neuroprotection in vascular dementia," *Cerebrovascular Diseases*, vol. 21, no. 2, pp. 106–117, 2006.

[32] C. H. Rojas-Fernandez and P. Moorhouse, "Current concepts in vascular cognitive impairment and pharmacotherapeutic implications," *Annals of Pharmacotherapy*, vol. 43, no. 7-8, pp. 1310–1323, 2009.

[33] J.-M. Park, A.-S. Shin, S.-U. Park, I.-S. Sohn, W.-S. Jung, and S.-K. Moon, "The acute effect of acupuncture on endothelial dysfunction in patients with hypertension: a pilot, randomized, double-blind, placebo-controlled crossover trial," *The Journal of Alternative and Complementary Medicine*, vol. 16, no. 8, pp. 883–888, 2010.

[34] H. MacPherson, "Pragmatic clinical trials," *Complementary Therapies in Medicine*, vol. 12, no. 2-3, pp. 136–140, 2004.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 291680, 9 pages
http://dx.doi.org/10.1155/2015/291680

Research Article

Using Bioinformatics Approach to Explore the Pharmacological Mechanisms of Multiple Ingredients in *Shuang-Huang-Lian*

Bai-xia Zhang,[1] Jian Li,[2] Hao Gu,[3] Qiang Li,[1] Qi Zhang,[1] Tian-jiao Zhang,[2] Yun Wang,[1] and Cheng-ke Cai[1]

[1]*School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100102, China*
[2]*School of Basic Medical Sciences, Beijing University of Chinese Medicine, Beijing 100029, China*
[3]*Institute of Basic Research of Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China*

Correspondence should be addressed to Yun Wang; wangyun@bucm.edu.cn and Cheng-ke Cai; cck98@126.com

Received 5 January 2015; Revised 10 June 2015; Accepted 7 July 2015

Academic Editor: Josiah Poon

Due to the proved clinical efficacy, Shuang-Huang-Lian (SHL) has developed a variety of dosage forms. However, the in-depth research on targets and pharmacological mechanisms of SHL preparations was scarce. In the presented study, the bioinformatics approaches were adopted to integrate relevant data and biological information. As a result, a PPI network was built and the common topological parameters were characterized. The results suggested that the PPI network of SHL exhibited a scale-free property and modular architecture. The drug target network of SHL was structured with 21 functional modules. According to certain modules and pharmacological effects distribution, an antitumor effect and potential drug targets were predicted. A biological network which contained 26 subnetworks was constructed to elucidate the antipneumonia mechanism of SHL. We also extracted the subnetwork to explicitly display the pathway where one effective component acts on the pneumonia related targets. In conclusions, a bioinformatics approach was established for exploring the drug targets, pharmacological activity distribution, effective components of SHL, and its mechanism of antipneumonia. Above all, we identified the effective components and disclosed the mechanism of SHL from the view of system.

1. Introduction

Traditional Chinese Medicine (TCM), one of the main items of complementary and alternative medicine, is a healthcare focused medical system. As the main characteristics, formula is the most important part which has been utilized for treating diseases and promoting the health of humans for thousands of years in TCM practice [1]. As known, herbal formula is the most important part which has been utilized for treating diseases and promoting the health of humans for thousands of years in TCM practice. In the formula, each herb contains many compounds that offer multitarget, multicomponent synergy, and multidimensional pharmacological actions. Taking these concerns, there is a considerable challenge for researchers to disclose the mechanisms of formula using conventional pharmacological methods. Fortunately, with the development of pharmaceutical chemistry, more and more public databases of Traditional Chinese Medicine were built, such as Traditional Chinese Medicine Integrated Database (TCMID), Traditional Chinese Medicine Information Database (TCM-ID), TCMGeneDIT, and Chinese Traditional Medicine Herbs Database [2, 3]. Based on these qualitative databases, some valuable information could be addressed by system biological technology to identify some mechanisms of herbal formula.

Shuang-Huang-Lian (SHL), one of the famous modern formulae prepared from three medicinal herbs including *Flos Lonicerae*, *Radix Scutellariae*, and *Fructus Forsythiae*, mainly has antibacterial, antivirus, and anti-inflammation activities, which is put into clinic for treatment of the diseases including acute respiratory tract infection, bacterial infection, and pneumonia [4]. Currently, SHL has been developed in a variety of dosage forms due to its proved clinical efficacy, for example, SHL capsule, SHL soft capsule, SHL tablet, SHL

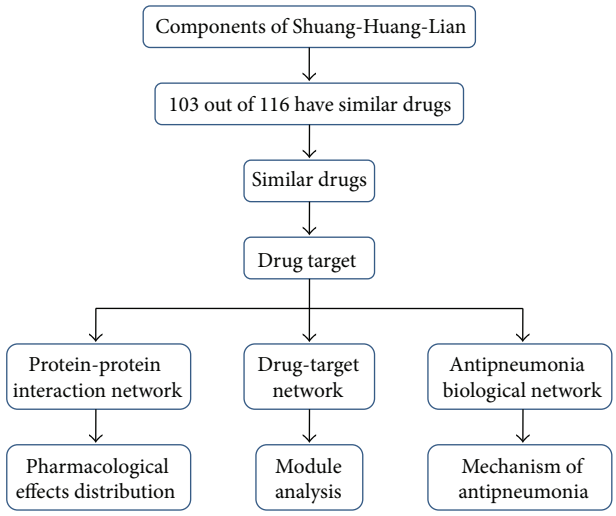


Figure 1: The workflow of SHL network construction and analysis.

oral liquid, SHL mixture, and SHL injection. However, the in-depth researches on holistic and synergetic pharmacological mechanisms of SHL preparations were scarce, and the studies on the targets of multicomponents and pharmacological activity of SHL are necessary for setup.

In this study, a bioinformatics strategy, as shown in Figure 1, was adapted to explore the novel targets and activity of SHL. Hopefully, the approach could arouse a new paradigm for investigating and explaining the roles of herbal formulas.

2. Data Source and Methods

2.1. Data Sources. The components of SHL came from Traditional Chinese Medicine Basic Information Database of State Administration of Traditional Chinese Medicine of People's Republic of China (http://cowork.cintcm.com/engine/windex.jsp) and A Handbook on Analysis of the Active Composition in TCM [5] (Supplementary Table 1 in Supplementary Material available online at http://dx.doi.org/10.1155/2015/291680). The drug targeting proteins data comes from DrugBank database [6]. The context of network comes from REACTOME database [7]; protein-protein interactions (PPI) come from REACTOME database [7] and HPRD database [8]. Pharmacological effects data is collected from DrugBank database. The proteins related to pneumonia are derived from the SciClips database.

2.2. Similar Drugs of the SHL Components Are Predicted on DrugBank Database. Based on two-dimensional structural similarity, we could obtain the similar drugs (Supplementary Table 2) which exist in the DrugBank database by the module of the "ChemQuery Structure Search" in DrugBank (http://www.drugbank.ca/structures//search/small_molecule_drugs/structure). In order to obtain the most credible results, two conditions need to be satisfied: (1) the parameter of similarity threshold was higher than 0.6 and the other parameters' values were default; (2) the most similar drug to each component was retained.

2.3. Network Construction. In order to construct the TCM complex system model, identify effective cluster, and illustrate the mechanisms further, we put forward the directed TCM grammar systems (dTGS) [9, 10] which are based on the theory of Entity Grammar system [11]. dTGS is a tetrad, $G = (V, F, P, S)$, whereas V is the character set representing basic element, F is a finite set of relations for V, V and F were viewed as entity, P is a set of rules to deduce relationships between entities, and S is the starting entity. According to different objective, we can write different reasoning engine. Providing a starting condition (starting entity), dTGS can obtain the result of relationship among entities automatically. With these relationships among entities, we can construct networks. In this paper, we use dTGS as framework to find the relationships between components and other entities. The results, after reasoning and rearranging, were visualized with Cytoscape. Thus, we can construct the drug-target network and the antipneumonia biological network of SHL. The application of dTGS was more flexible compared to traditional network analysis methods. According to different objective, we can define different V, F, P, and S.

Due to the various objectives and data, the V, F, P, and S of the PPI network and antipneumonia biological network should be defined, respectively.

The V, F, P, and S of the PPI network are as follows:

(1) $V = V_1 \cup V_2 \cup V_3 \cup V_4$:

V_1 is the set of components of the corresponding herb, V_2 is the similar drugs of components, V_3 is the set of the targets of the drugs, and V_4 is the set of rest proteins in the background network.

(2) $F = \{\text{herbX(A)}, \text{herbX(A, C)}, \text{drugtarget(C, B)}, \text{link(B, D)}, \text{link(D, E)}\}$:

herbX(A) represents the components existing in the corresponding herb; herbX(A, C) represents the relationship of the components of the corresponding herb and its similar drug; drugtarget(C, B) represents the relationship of similar drug and its target; link(B, D) and link(D, E) represent the relationship of nodes in the background network which was prior knowledge.

(3) $P = P_1 \cup P_2 \cup P_3$:

$P_1 = \{\text{herbX(A, C)}, \text{drugtarget(C, B)}. \Rightarrow \text{net(A, B, 1)}\}$.

$P_2 = \{\text{net(A, B, 1)}, \text{link(B, D)}. \Rightarrow \text{net(A, D, 2)}\}$.

$P_3 = \{\text{net(A, D, } N\text{)}, \text{link(D, E)}, N < 10. \Rightarrow \text{net(A, E, } N + 1\text{)}\}$.

P_1 are used to deduce the relations of chemical components and their targets. P_2 and P_3 are used to deduce the related proteins of components in PPI networks. "link(C, B)" as prior knowledge represents relations of nodes in PPI network. "net(A, B, N)" is the results which are obtained by reasoning engine; N is the cumulative distance from component A to protein E which is less than 10.

(4) $S = S_1 \cup S_2$:

S_1 is the set of entities with structure link(B, D) in background network for deduction. S_2 is the starting point of reasoning with structure herbX(A).

The V, F, P, and S of the antipneumonia biological network is described by the following:

(1) $V = V_1 \cup V_2 \cup V_3 \cup V_4$:

V_1 is the set of targets of components of corresponding herb, V_2 is the set of the pneumonia-related proteins, V_3 is the set of rest proteins in the background network, and V_4 is the set of rest proteins or small molecule metabolites in the background network.

(2) F = {link(A, B, X, Y), in(A), out(B), totalnet(A, B, X, Y), minnet(A, B, X, Y)}:

In link(A, B, X, Y), A, B $\in V_3$, $X \in$ {pos, neg}, $Y \in Z^*$.

In totalnet(A, B, X, Y), A $\in V_1 \cap V_3$, B $\in V_3$, $X \in$ {pos, neg}, $Y \in Z^*$.

In minnet(A, B, X, Y), A $\in V_1 \cap V_3$, B $\in V_2 \cap V_3$, $X \in$ {pos, neg}, $Y \in Z^*$.

In(A) and out(B), A $\in V_1$, B $\in V_2$.

link(A, B, X, Y) defines that A(protein) acts on B(protein) with an effect described in X and through Y reactions so that the distance number is Y. If process happens in one reaction, Y equals 1. In in(A) and out(B), A represents the targets of compounds, while B represents the pneumonia-related proteins. totalnet(A, B, X, Y) represents that A (target protein) affects B (pneumonia-related protein) with an effect described in X by reactions of Y. minet(A, B, X, Y) defines that A (target protein) can affect B (pneumonia-related protein) with an effect described in X by multiple pathway but we just retain the shortest path with Y reactions.

(3) $P = P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5 \cup P_6 \cup P_7 \cup P_8 \cup P_9 \cup P_{10} \cup P_{11}$:

P_1 = {link(A, B, X, Y), in(A) $\Rightarrow$ totalnet(A, B, X, M)}.

P_2 = {totalnet(A, B, pos, D), link(B, C, pos, E) $\Rightarrow$ totalnet(A, C, pos, E + D)}.

P_3 = {totalnet(A, B, pos, D), link(B, C, neg, E) $\Rightarrow$ totalnet(A, C, neg, E + D)}.

P_4 = {totalnet(A, B, neg, D), link(B, C, neg, E) $\Rightarrow$ totalnet(A, C, pos, E + D)}.

P_5 = {totalnet(A, B, neg, D), link(B, C, pos, E) $\Rightarrow$ totalnet(A, C, neg, E + D)}.

P_6 = {#min{D: totalnet(A, B, X, Y)} = M, totalnet(A, B, X, Y), out(B) $\Rightarrow$ minnet(A, B, X, M)}.

P_7 = {distance(C, B, X, Y) $\Rightarrow$ length(Y)}.

P_8 = {link(C, B, D), distance(_, B, _, _) $\Rightarrow$ backward(C, B, D, 1)}.

P_9 = {link(D, C, E), backward(C, B, D, N), $M = N + 1$, $N < Y$, length(Y) $\Rightarrow$ backward(D, C, E, F)}.

P_{10} = {distance(A, _, _, _), backward(A, C, X, Y) $\Rightarrow$ forward(A, C, X, 1)}.

P_{11} = {forward(A, C, X, N), backward(C, D, E, F), $M = N + 1$, $N = Y -$ F, length(Y) $\Rightarrow$ forward(C, D, E, M)}.

P_1 tags the starting point of derivation by in(A) from link(A, B, X, Y), and the tagged link(A, B, X, Y) is named as totalnet(A, B, X, Y). $P_2 \cup P_3 \cup P_4 \cup P_5$ is the set of rules to deduce the eventual effects and distances of target proteins to other downstream proteins in the network. In link(A, B, X, Y), totalnet(A, B, pos, D), and minnet(A, B, X, Y), "X" and "pos" and "Y" and "D" represent the same data type, respectively, just because they located in the same position. P_2 indicates that if the effect of A on B is positive, and the effect of B on C is positive too, then the effect of A on C is positive. Meanwhile, if A affects B through D reactions, and B affects C through E reactions, then the distance of A to C is D plus E. Similar derivations are defined in P_3, P_4, and P_5. They may be used as many times as necessary to the final pneumonia-related protein. P_6 is a rule used to identify the shortest distance from a target protein to a pneumonia-related protein when the paths between them are too complicated to analyze. In P_6, if B is a pneumonia-related protein in totalnet(A, B, X, Y), then #min{D: totalnet(A, B, X, Y)} = M indicates that the shortest distance from A to B is M, which is extracted by minnet(A, B, X, M). In the construction of target network, P_6 will be used as many times as necessary to the target protein pneumonia-related protein pairs. $P_7 \cup P_8 \cup P_9 \cup P_{10} \cup P_{11}$ were used to describe the detailed pathway from A to B with the clear steps of Y. The forward(C, D, E, M) was the final result used to construct the network. After connecting each of the forward(C, D, E, M), we could get the detailed pathway from A to B.

(4) $S = S_1 \cup S_2$:

S_1 is the set of entities with structure link(A, B, X, Y) in background network for deduction. S_2 is the set of starting and end point proteins, described by in(A) and out(B).

3. Results and Discussion

3.1. Topology Analysis of SHL PPI Network. As shown in Figure 2 (clear node label can be seen in Figure 2), the SHL PPI network, there are 1953 nodes and 3112 edges. The network diameter is 12, which means the greatest distance between any pair of vertices is 12. The node degree distribution indicated that the PPI of SHL followed the power law with a degree exponent of 0.969 ($R2 = 0.647$). The Clustering coefficient is 0.322. The connected component is 1.0. Network centralization is 0.3112. And network heterogeneity as 6.017. These common topological parameters suggested that the network exhibited scale-free property and modular architecture.

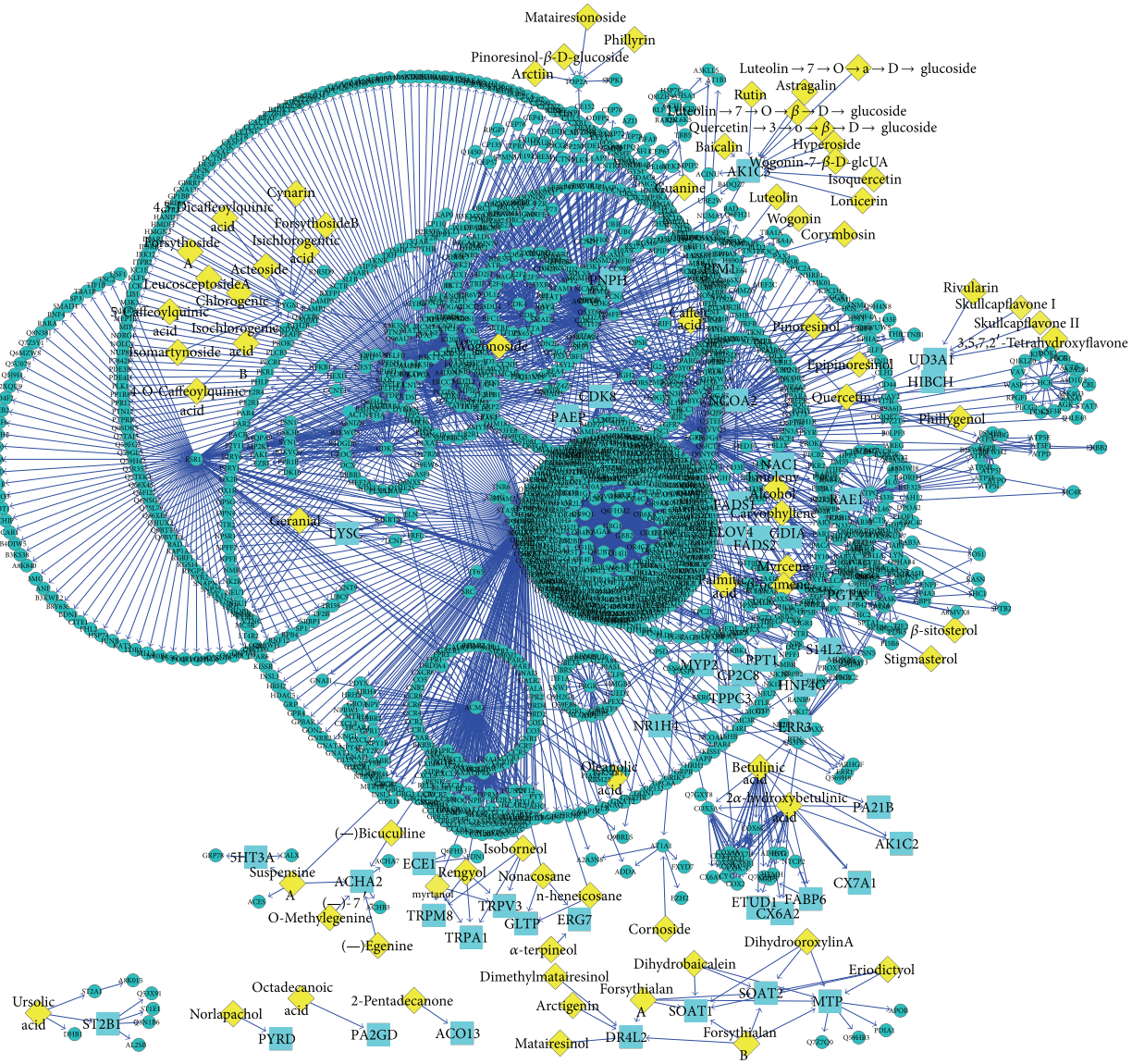


Figure 2: PPI Network of SHL. The components in SHL Network are yellow diamonds. The targets are shown as blue squares. In the PPI network, there are 92 components (Supplementary Table 3) acting on 129 targets which interact with 1787 proteins.

3.2. Drug-Target Network. In order to express the relationship between SHL components more clearly, we constructed a drug-target network of SHL. As shown in Figure 3, we totally labeled the 21 modules. Most of them match the common targets. However, it may be worth nothing that some novel targets are detected, such as thirteen components of the module (6) act on the same targets, PYGM, which play an important role in regulation of cell cycle and cellular macromolecule metabolic process. These biological processes are closely related to cell proliferation and tumorigenesis (anticancer effect). Following the same strategy, it is easy to help us to understand the functional classifications of each SHL component based on the modules in drug-target network.

3.3. Pharmacological Effects Distribution of SHL. In the PPI Network of SHL, we counted pharmacological effects of all the proteins; frequency statistical results are presented in Figure 4. The results indicate that five in top ten pharmacological effects of SHL are linked with antineoplastic agents, protein kinase inhibitors, enzyme inhibitors, growth inhibitors, and phytogenic antineoplastic agents, which suggested that SHL might directed against tumor effect. Furthermore, two pharmacological effects focused on antioxidants and anti-inflammatory agents, which could help tumor suppression. For sure, the novel pharmacological effects of SHL need further empirical data form bench to bedside.

3.4. The Anti-Pneumonia Mechanism of SHL. Using the data derived from the Reactome database as context, we constructed the biological network of the effective components of SHL acting on the pneumonia related targets. Circular node represents disease related targets, diamond node represents effective components of SHL, square node represents proteins

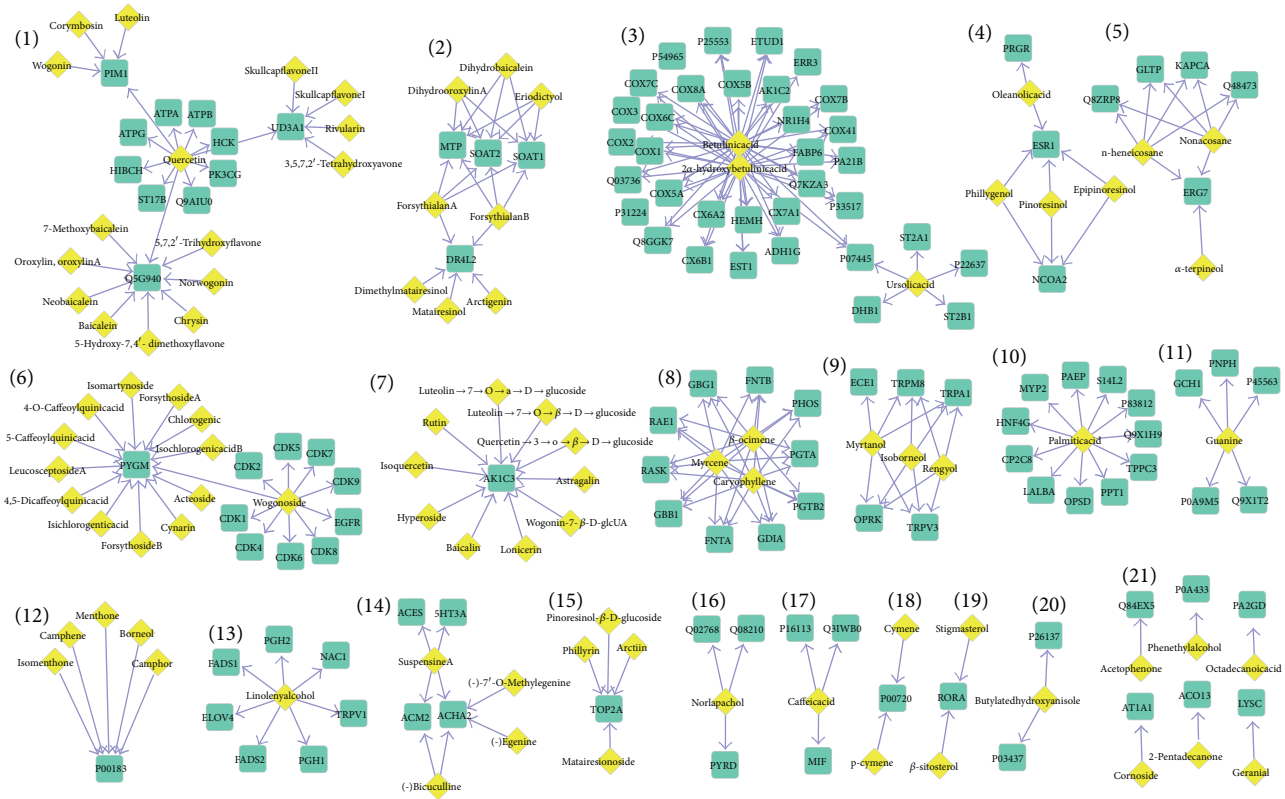


Figure 3: Drug-target Network of SHL. The components are yellow diamonds. The targets are shown as blue squares.

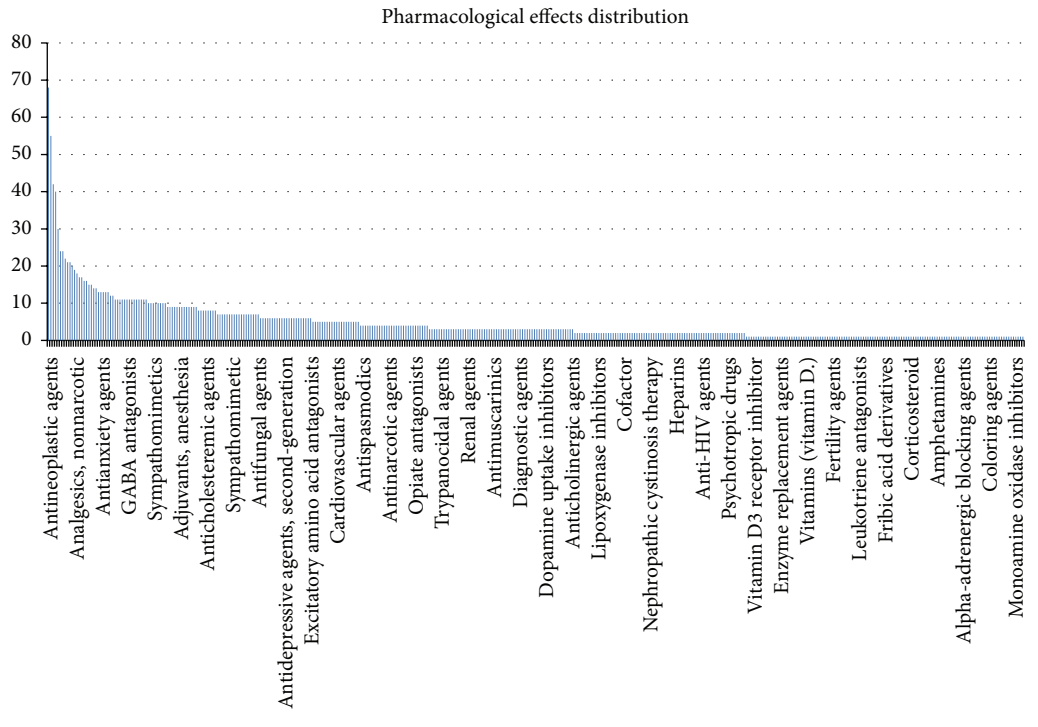


Figure 4: The pharmacological distribution of all 112 components in SHL.

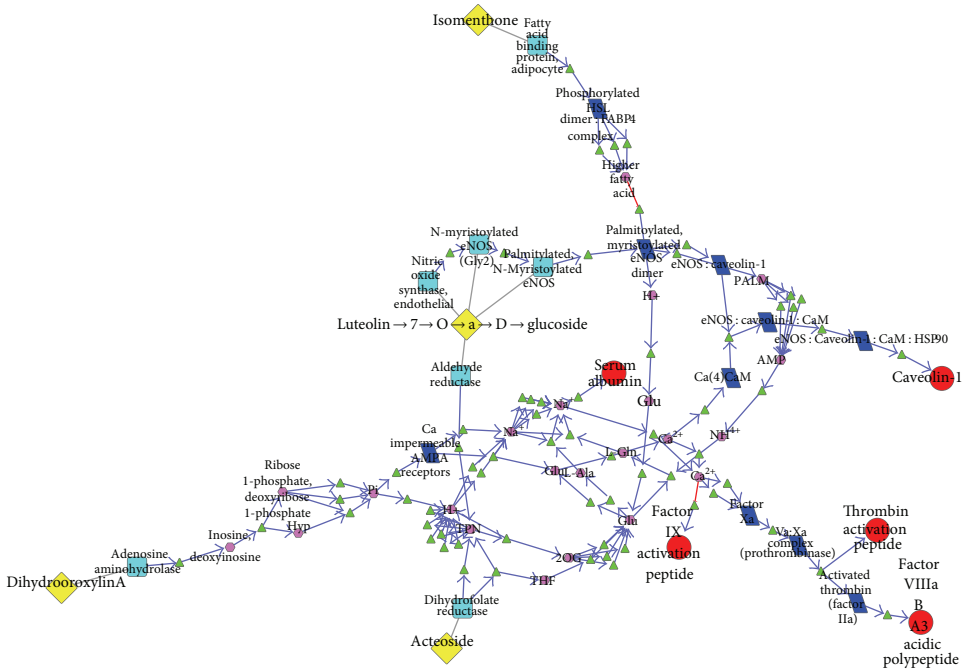


Figure 5: The antipneumonia biological network of SHL.

involved in the pathway, parallelogram node represents complex, hexagon node represents micromolecule, and triangle node represents biological reaction. The edge with triangle arrow represents positive regulation, edge with T-shaped arrow represents negative regulation, and the undirected edge means the direction is uncertain (Figure 5). This biological network which included 4 components, 5 pneumonia related targets, and 26 subnetworks could overall exhibit the effective components and the mechanism of antipneumonia on a molecular level.

This biological network included 26 subnetworks that 4 components that acted on 5 pneumonia related targets. All of these subnetworks intertwined to play the role of antipneumonia collectively and demonstrated the features of biological systems simultaneously, such as robustness, redundancy, crosstalk, and so forth. In order to clearly show the antipneumonia mechanism of SHL effective components, we could extracted 26 subnetworks, respectively. This paper took 2 extracted subnetworks as examples: the biological pathway where luteolin-7-o-α-D-glucoside acted on serum albumin (Figure 6) and the biological pathway where dihydrooroxylin A acted on thrombin activation peptide (Figure 7).

As target of luteolin-7-o-α-D-glucoside, endothelial nitric oxide synthase (eNOS) translocated from Golgi to caveolae after N-myristoylation and palmitoylation. With depalmitoylation of eNOS dimer, it produced PALM. After the reaction of PALM and CoASH, palmitic acid converted to palmitoyl-CoA, accompanyied with energy transformation. After a series of signal transduction of micromolecule (NH^{4+}, L-Gln, Na^{+}, etc.), the content of serum albumin changed. As the biomarker in the early stage, this pathway could elucidate the course of the change. That is, luteolin-7-o-α-D-glucoside acting on eNOS leads to the increase of vasopermeability and serum albumin influxed into the interval of capillaries, accelerated the speed of catabolism, and affected the content of serum albumin at last [12, 13].

Figure 7 shows the initial two steps of blood coagulation: the formation process of thrombin activation peptide and the activation of thrombin. As the target of dihydrooroxylin A, adenosine aminohydrolase was hydrolyzed, dephosphorylated, and oxidated. After some amino acid was taken up and GluR2 transferred, the Ca^{2+} impermeable AMPA receptor and aspartic acid receptor were activated and then extruded of Ca^{2+}. Factor Xa was activated by Ca^{2+}, and Ca^{2+} also could accelerate the combination of factors Xa and Va [14]. Factor Va had no activity itself, but it could enhance activity of factor Xa and accelerated the formation of thrombin [15]. Thrombin could accelerate the blood coagulation and wound healing, thus treating alimentary tract hemorrhage of severe pneumonia. In conclusion, dihydrooroxylin A acting on thrombin plays the role of antipneumonia by accelerating the hemostasis of the alimentary tract.

By summarizing the 26 subnetworks of SHL effective components cluster acting on pneumonia-related targets, we got 3 main pathways where SHL played the role of antipneumonia: (1) regulating the activity of caveolin-1 existing in the signal pathway of inflammatory and endothelial cells affects the response of inflammatory cell to the inflammation, further influencing the process of inflammation; (2) accelerating the blood coagulation and wound healing, thus treating alimentary tract hemorrhage of severe pneumonia; (3) affecting the content of serum albumin, promoting the repair of lung tissue, and enhancing the immune function.

Biological network shows 4 components could act on 5 targets within 10 biological reactions. Other components in SHL may also act on pneumonia-related targets, but

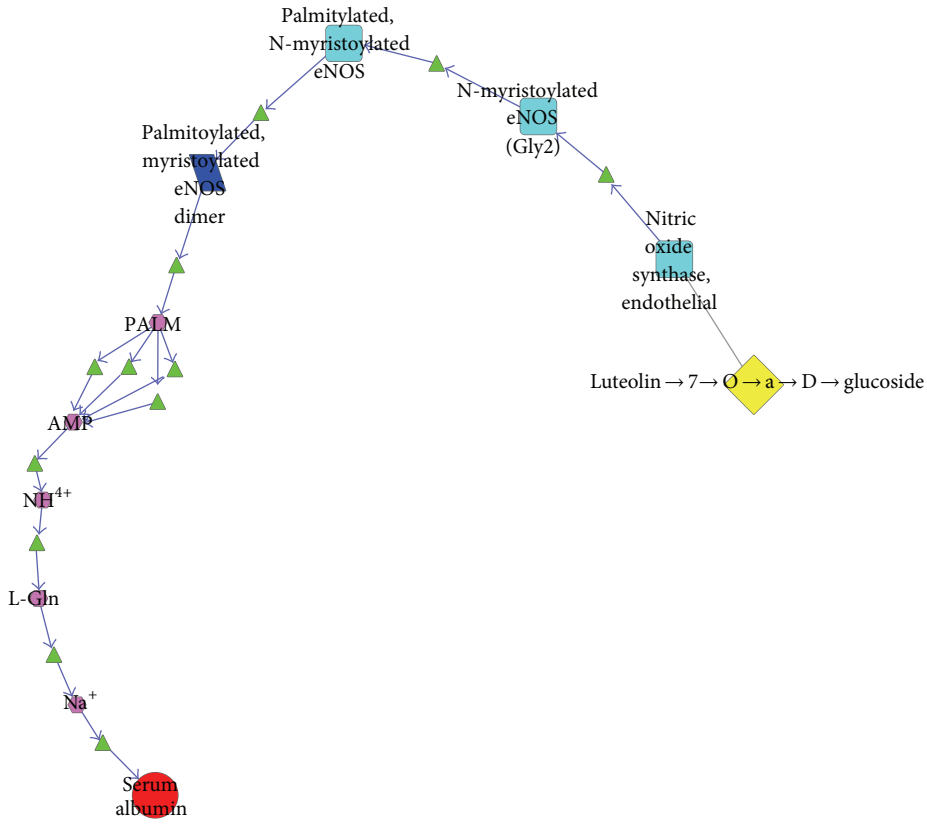


Figure 6: The antipneumonia biological pathway of luteolin-7-o-α-D-glucoside.

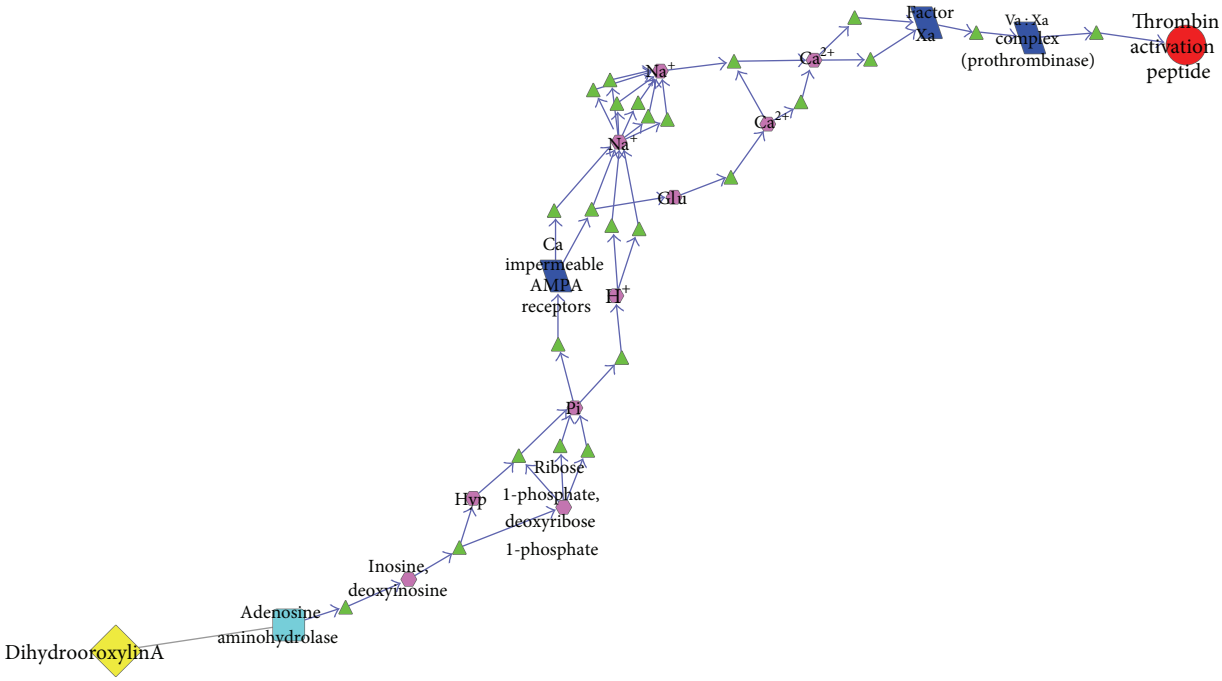


Figure 7: The antipneumonia biological pathway of dihydrooroxylin A.

the number of biological reactions will be more. When we defined the reactions as 15, biological network contained 5 components and 9 pneumonia-related targets. If the reaction was 20, 43 components and 13 targets were included in the biological network. The 4 components, by contrast, were more quickly and efficiently acting on pneumonia-related targets. Compared with the components in SHL which studied frequently, such as chlorogenic acid, forsythin, and baicalin, the content of four components within the network was lower, but they may play the same role of pneumonia treatment.

4. Conclusion

Based on the public databases of TCM, DrugBank database, and directed TCM grammar systems, we have built a PPI network, drug target network, and antipneumonia biological network of SHL. By the modules analysis, we predicted that SHL has the potential to be an antitumor candidate. This result may provide a novel clue for further experimental studies of SHL. In future, the predicted novel pharmacological effects of SHL should be further validated from bench to bedside. The antipneumonia biological network systematically explained the reason why SHL could play the role of antipneumonia and identified its effective components. All these are helpful to guide the quality control and further study of SHL.

Studies have already demonstrated that SHL had the pharmacological actions of antimicrobial, antiviral, antipyretic, anti-inflammatory, antioxidant, antiarrhythmic, and enhanced immunity [16–19], but they only focused on individual component, pharmacological action, or pathway. The component they studied could be extracted and was with a high content. Compared with these studies, our study could elucidate the mechanism holistically and on the molecular level. Meanwhile, enclosing the synergistic effect of various components and the cross of pathways, we not only addressed that luteolin-7-α-D-glucoside [20], dihydrooroxylin A [21, 22], and acteoside [23] played an anti-inflammatory role but also predicted that isomenthone could act on pneumonia-related targets through biological network of SHL, thus playing the role of antipneumonia.

The thought of this study accords with the network pharmacology, but the method we adopted is different from other related works [24, 25]. The whole process of other related works is carried out manually and the batch processing is impossible. With the characteristic of flexibility, dTGS could solve these problems effectively. Carrying out the TCM research under the guidance of network pharmacology is more quick and systematic than traditional methods. Meanwhile, dTGS provides a novel strategy for the study of TCM and other complex systems.

Because the research results were based on the inference of data included in the existing database or literature, and because we do not take into account the quantity of components and the specific interactional environment, this method still has some limitations. With further study of the complex systems, such as TCM formula and related disease, the data we used will be more complete and the results we got will be more precise and integrity.

Conflict of Interests

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

Authors' Contribution

Bai-xia Zhang, Jian Li, and Hao Gu contributed equally to this work.

Acknowledgments

This work was supported by National Science and Technology Major Projects for "major new drugs innovation and development" (Grant no. 2011ZX09201-201-15), the National Natural Science Foundation of China (Grant nos. 81173568 and 81373985), and Program for New Century Excellent Talents in University (Grant no. NCET-11-0605). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

References

[1] S. Li, Z. Q. Zhang, L. J. Wu, X. G. Zhang, Y. D. Li, and Y. Y. Wang, "Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network," *IET Systems Biology*, vol. 1, no. 1, pp. 51–60, 2007.

[2] X. Chen, H. Zhou, Y. B. Liu et al., "Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation," *British Journal of Pharmacology*, vol. 149, no. 8, pp. 1092–1103, 2006.

[3] Z. L. Ji, H. Zhou, J. F. Wang, L. Y. Han, C. J. Zheng, and Y. Z. Chen, "Traditional Chinese medicine information database," *Journal of Ethnopharmacology*, vol. 103, no. 3, 501 pages, 2006.

[4] J. Ye, X. Song, Z. Liu et al., "Development of an LC-MS method for determination of three active constituents of Shuang-huang-lian injection in rat plasma and its application to the drug interaction study of Shuang-huang-lian freeze-dried powder combined with levofloxacin injection," *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, vol. 898, pp. 130–135, 2012.

[5] H. Liu, Y. X. Qiao, B. Liu, and J. J. Zhou, "Transformation of traditional Chinese medicine database into data warehouse," *Chinese Pharmaceutical Journal*, vol. 9, pp. 645–648, 2006.

[6] V. Law, C. Knox, Y. Djoumbou et al., "DrugBank 4.0: shedding new light on drug metabolism," *Nucleic Acids Research*, vol. 42, no. 1, pp. D1091–D1097, 2014.

[7] D. Croft, A. F. Mundo, R. Haw et al., "The Reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 42, no. 1, pp. D472–D477, 2014.

[8] T. S. Keshava Prasad, R. Goel, K. Kandasamy et al., "Human protein reference database—2009 update," *Nucleic Acids Research*, vol. 37, no. 1, pp. D767–D772, 2009.

[9] L. Ji, R. L. Ren, H. Gu et al., "dTGS: method for effective components identification from traditional Chinese medicine formula and mechanism analysis," *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, Article ID 840427, 9 pages, 2013.

[10] J. Yan, Y. Wang, S.-J. Luo, and Y.-J. Qiao, "TCM grammar systems: an approach to aid the interpretation of the molecular interactions in Chinese herbal medicine," *Journal of Ethnopharmacology*, vol. 137, no. 1, pp. 77–84, 2011.

[11] W. Yun, "Entity grammar systems: a grammatical tool for studying the hierarchal structures of biological systems," *Bulletin of Mathematical Biology*, vol. 66, no. 3, pp. 447–471, 2004.

[12] B. Ruot, I. Papet, F. Béchereau et al., "Increased albumin plasma efflux contributes to hypoalbuminemia only during early phase of sepsis in rats," *The American Journal of Physiology—Regulatory Integrative and Comparative Physiology*, vol. 284, no. 3, pp. R707–R713, 2003.

[13] J. U. Hedlund, L.-O. Hansson, and A. B. Ortqvist, "Hypoalbuminemia in hospitalized patients with community-acquired pneumonia," *Archives of Internal Medicine*, vol. 155, no. 13, pp. 1438–1442, 1995.

[14] E. Persson, P. J. Hogg, and J. Stenflo, "Effects of Ca^{2+} binding on the protease module of factor Xa and its interaction with factor Va: evidence for two Gla-independent Ca^{2+}-binding sites in factor Xa," *The Journal of Biological Chemistry*, vol. 268, no. 30, pp. 22531–22539, 1993.

[15] E. A. Norstrøm, S. Tran, M. Steen, and B. Dahlbäck, "Effects of factor Xa and protein S on the individual activated protein C-mediated cleavages of coagulation factor Va," *The Journal of Biological Chemistry*, vol. 281, no. 42, pp. 31486–31494, 2006.

[16] W. Zhou, X. X. Zhu, A. L. Yin et al., "Effect of various absorption enhancers based on tight junctions on the intestinal absorption of forsythoside A in Shuang-Huang-Lian, application to its antivirus activity," *Pharmacognosy Magazine*, vol. 10, no. 37, pp. 9–17, 2014.

[17] X. Gao, M. Guo, Q. Li et al., "Plasma metabolomic profiling to reveal antipyretic mechanism of Shuang-Huang-Lian injection on yeast-induced pyrexia rats," *PLoS ONE*, vol. 9, no. 6, Article ID e100017, 2014.

[18] R. Gao, Y. N. Lin, G. Liang, B. Yu, and Y. Gao, "Comparative pharmacokinetic study of chlorogenic acid after oral administration of lonicerae japonicae flos and shuang-huang-lian in normal and febrile rats," *Phytotherapy Research*, vol. 28, no. 1, pp. 144–147, 2014.

[19] Y. Gao, L. Fang, R. Cai et al., "Shuang-Huang-Lian exerts anti-inflammatory and anti-oxidative activities in lipopolysaccharide-stimulated murine alveolar macrophages," *Phytomedicine*, vol. 21, no. 4, pp. 461–469, 2014.

[20] C. M. Park and Y.-S. Song, "Luteolin and luteolin-7-O-glucoside inhibit lipopolysaccharide-induced inflammatory responses through modulation of NF-κB/Ap-1/PI3K-AKT signaling cascades in RAW 264.7 cells," *Nutrition Research and Practice*, vol. 7, no. 6, pp. 423–429, 2013.

[21] H. Lee, Y. O. Kim, H. Kim et al., "Flavonoid wogonin from medicinal herb is neuroprotective by inhibiting inflammatory activation of microglia," *The FASEB Journal*, vol. 17, no. 13, pp. 1943–1944, 2003.

[22] Y.-C. Chen, L.-L. Yang, and T. J.-F. Lee, "Oroxylin A inhibition of lipopolysaccharide-induced iNOS and COX-2 gene expression via suppression of nuclear factor-κB activation," *Biochemical Pharmacology*, vol. 59, no. 11, pp. 1445–1457, 2000.

[23] J. He, X.-P. Hu, Y. Zeng et al., "Advanced research on acteoside for chemistry and bioactivities," *Journal of Asian Natural Products Research*, vol. 13, no. 5, pp. 449–464, 2011.

[24] L. An and F. Feng, "Network pharmacology-based antioxidant effect study of Zhi-Zi-da-huang decoction for alcoholic liver disease," *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, Article ID 492470, 6 pages, 2015.

[25] F. Zhao, L. Guochun, Y. Yang, L. Shi, L. Xu, and L. Yin, "A network pharmacology approach to determine active ingredients and rationality of herb combinations of Modified-Simiaowan for treatment of gout," *Journal of Ethnopharmacology*, vol. 168, pp. 1–16, 2015.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 946793, 9 pages
http://dx.doi.org/10.1155/2015/946793

Research Article

Detecting Disease in Radiographs with Intuitive Confidence

Stefan Jaeger

U.S. National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA

Correspondence should be addressed to Stefan Jaeger; stefan.jaeger@nih.gov

Received 23 January 2015; Accepted 23 March 2015

Academic Editor: Zhaohui Liang

This paper argues in favor of a specific type of confidence for use in computer-aided diagnosis and disease classification, namely, sine/cosine values of angles represented by points on the unit circle. The paper shows how this confidence is motivated by Chinese medicine and how sine/cosine values are directly related with the two forces Yin and Yang. The angle for which sine and cosine are equal (45°) represents the state of equilibrium between Yin and Yang, which is a state of nonduality that indicates neither normality nor abnormality in terms of disease classification. The paper claims that the proposed confidence is intuitive and can be readily understood by physicians. The paper underpins this thesis with theoretical results in neural signal processing, stating that a sine/cosine relationship between the actual input signal and the perceived (learned) input is key to neural learning processes. As a practical example, the paper shows how to use the proposed confidence values to highlight manifestations of tuberculosis in frontal chest X-rays.

1. Introduction

With increasing performance of automated disease detection, computer-aided diagnosis (CAD) is becoming a serious alternative to the established diagnostic workflow [1]. CAD can lead to better diagnostics by providing physicians with critical information extracted from many relevant cases through machine learning techniques. However, the communication between man and machine should be intuitive so that a physician can readily use the machine output for diagnostics. This paper describes a method for generating confidence values for local manifestations of tuberculosis detected automatically in a frontal chest X-ray (CXR). The proposed confidence values are motivated by Chinese medicine and neural signal processing. The higher the confidence the more likely it is that the detected abnormal region is indeed abnormal. The confidence values correspond to sine or cosine and can thus be represented by a point on the unit circle. A confidence of sin(45°) describes a region that the machine considers neither normal nor abnormal. This confidence therefore describes a state of equilibrium where Yin and Yang are in perfect balance, a state of nonduality. Note that this is in contrast to the conventional understanding in traditional Chinese medicine, where the equilibrium between Yin and Yang denotes a healthy state. However, in the author's opinion this is incorrect because the state of equilibrium between Yin and Yang is the state of nonduality, but assigning attributes, such as "healthy state," clearly introduces duality. Therefore, the equilibrium of Yin and Yang should be associated with neither normality nor abnormality. The proposed confidence values meet this requirement. The state of equilibrium corresponds to a state of uncertainty, in which it is entirely unclear whether a lung region is normal or abnormal. This could be because there is no indication for one or the other, or there is supporting evidence in equal shares for both normality and abnormality.

A second motivation of the confidence values proposed here is that they are designed to be intuitively accessible by a radiologist, in the sense that they can be easily learned without much calibration on the part of the radiologist. The proposed confidence values should represent similar numerical values a human expert would provide if asked to quantify his confidence in a specific abnormality. This would guarantee that the opinion of the machine is put on an equal footing with the opinion of the radiologist. The radiologist can then integrate his own confidence with the machine confidence and reach the final verdict about whether a specific lung region should be considered normal

or abnormal. The paper will resort to principles of neural signal processing to argue that the proposed confidence is intuitive because it forms the basic of synaptic learning.

The paper is structured as follows. In Section 2, the formal definition of Yin and Yang will be presented and the connection between the proposed confidence and the forces of Yin and Yang will be discussed in more detail. This chapter is largely motivated by earlier work in [2, 3]. In Section 3, the proposed confidence values will be used to represent manifestations of tuberculosis in frontal chest X-rays as a saliency map. Finally, a conclusion will summarize the main results.

2. Materials and Methods

Section 2.1 summarizes briefly the basic principles and processes of neural signal transduction that we can observe at chemical synapses. Section 2.2 explains the information-theoretic model that is used in this paper to formalize the information flow occurring at a synapse, similar to [3], and shows why the proposed model makes sense given what is known today about neural signal processing. Following this introduction of the theoretical model, Section 2.3 shows a direct connection to Yin and Yang. Section 2.4 then describes how synaptic learning is explained in the proposed information-theoretic model. Finally, Section 2.5 shows how these results motivate the proposed confidence values.

2.1. Neural Signal Transduction. This section presents a brief overview of the basic signal transduction principles, as they are needed for understanding the remainder of this paper. First, neurons and synapses are explained and then the Hodgkin and Huxley model describing electrical signal processing of nerve cells is presented. Most of the information in this subsection is taken from [3].

2.1.1. Neurons and Synapses. The human nervous system is composed of nerve cells, so-called neurons, which can communicate with each other through synapses. A synapse is a membrane-to-membrane junction that allows either chemical or electrical signal transmission. In the case of chemical synapses, which will be in the focus here, signals are transmitted via neurotransmitters that can bridge the synaptic cleft, a small gap between the membranes of two nerve cells. As an illustration, the diagram in Figure 1 shows two communicating neurons. A neuron can send a signal to another neuron through its axon, which is a protrusion with potentially thousands of synapses and which can extend to other neurons in distant parts of the body. A neuron can receive the signal via its soma or its dendrites that conduct the received signal to the cell body (see Figure 1). In both cases, the signal needs to pass a synapse that transmits the signal by molecular means, via neurotransmitters, through the synaptic cleft, from the presynaptic terminal to the postsynaptic terminal. The small volume of the synaptic cleft allows neurotransmitter concentration to increase and decrease rapidly. Prior to any signal transmission, the neurotransmitters are enclosed in small spheres, synaptic vesicles,

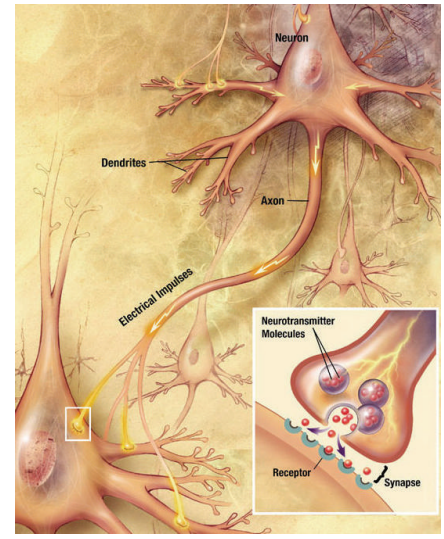


Figure 1: A signal propagating down an axon to the cell body and dendrites of the next cell (Source: NIA/NIH).

at the presynaptic terminal. On the other side, the postsynaptic terminal provides receptors for neurotransmitters traveling through the synaptic cleft. The lower right corner of Figure 1 shows a close-up of a synapse. The adult human brain contains between 10^{14} and 5×10^{14} of these synapses. Synapses, and the way they transmit information, are crucial to the biological computations that underlie perception and thought. The common understanding is that synapses, and changes in their behavior, are responsible for memorization and human learning. To get insight into these processes, it is essential to study the molecular processes underlying signal transmission.

Signal transmission at a chemical synapse is a multistep process (see Figure 2). The transmission is triggered by an electrochemical excitation (action potential) at the presynaptic terminal. The excitation causes calcium channels to open, allowing calcium ions to flow into the presynaptic terminal. The increased concentration of calcium ions in the presynaptic terminal causes the vesicles to release their neurotransmitters into the synaptic cleft. Some of these neurotransmitters bind to the receptors of the postsynaptic terminal, which opens ion channels in the postsynaptic membrane, allowing ions to flow into or out of the postsynaptic cell. This changes the transmembrane potential, leading to an excitation or inhibition of the postsynaptic cell. In this way, the action potential from the presynaptic terminal has created a postsynaptic potential by molecular means. Eventually, the docked neurotransmitters will break away from the postsynaptic receptors. Some of them will be reabsorbed by the presynaptic cell to initiate another transmission cycle.

2.1.2. Model by Hodgkin and Huxley. In 1952, in a seminal paper, Hodgkin and Huxley proposed a set of equations explaining the electrical characteristics of nerve cells and their underlying ionic mechanisms [4]. Their entry point is the sodium conductance at the cell membrane of a nerve cell.

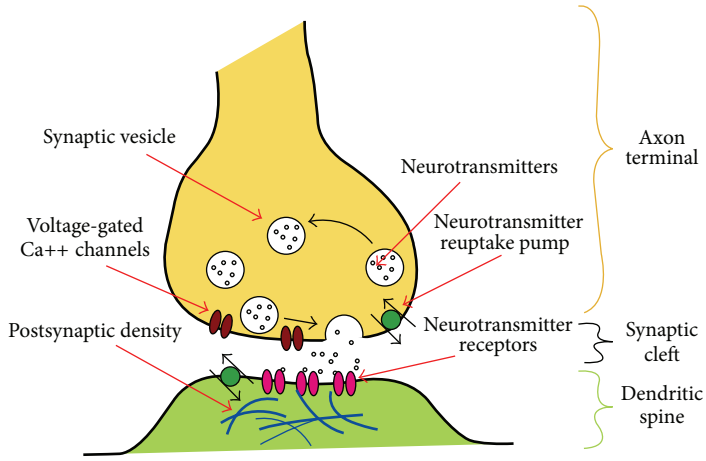


Figure 2: Signal transmission at a chemical synapse [21] (Source: Wikipedia, Surachit, Nrets).

Similar to calcium ions, sodium ions are largely responsible for generating action potentials in nerve cells. A nerve cell membrane has voltage-gated ion channels that are shut when the membrane is close to the resting potential. Once the membrane potential increases to a critical value, these ion channels open and allow sodium ions to pass the cell membrane and travel into the cell. The influx of sodium ions increases the membrane potential even more, causing more ion channels to open and thus allowing more sodium ions to move into the cell. This reinforcing process stops once the membrane potential has reversed and the nerve cell has reached its action potential. After reaching the action potential, the sodium channels close rapidly, preventing any more sodium ions from entering the cell. The sodium ions are then transported out of the nerve cell and the cell returns to its resting potential. Understanding the temporal change of the sodium concentration is therefore important for understanding the generation and transportation of action potentials.

In their mathematical model, Hodgkin and Huxley assume that the sodium conductance is proportional to the number of specific molecules on the inside of the membrane but that the conductance is independent of the number of molecules on the outside [4]. According to Boltzmann's principle the proportion P_i of the molecules on the inside of the membrane is related to the proportion P_o on the outside by

$$\frac{P_i}{P_o} = \exp\left[\frac{(w + zeE)}{kT}\right], \tag{1}$$

where E is the potential difference between the outside and the inside of the membrane, w is the work required to move a molecule from the inside to the outside of the membrane when $E = 0$, e is the absolute value of the electronic charge, z is the valency of the molecule (i.e., the number of positive electronic charges on it), k is Boltzmann's constant, and T is the absolute temperature [4]. With $P_i + P_o = 1$, the expression for P_i becomes

$$P_i = \frac{1}{1 + \exp\left(-\left(w + zeE\right)/kT\right)}. \tag{2}$$

The concentration of the molecules on the inside of the membrane thus follows a sigmoid function, which will become important later in this paper.

2.2. Information-Theoretic Model. The formal derivation of the theoretical information processing model proposed here begins with a closer look at the calcium ion concentration in the synaptic cleft close to the postsynaptic terminal. Let this concentration be p_i, as opposed to the outer concentration in the presynaptic terminal p_o. Then, let us assume that the strength S of the stimulus arriving at the presynaptic terminal is determined by the ratio of outer to inner calcium ion concentration p_o/p_i, or

$$S = \frac{1 - p_i}{p_i}. \tag{3}$$

For example, for an inactive synapse, the concentration of calcium ions in the synaptic cleft will be one, that is, $p_i = 1$, as there will be no calcium ions in the presynaptic terminal. Consequently, according to (3), the signal strength is zero. On the other hand, for an active synapse, calcium ions can freely flow into the presynaptic terminal until a concentration equilibrium is reached between calcium ions in the presynaptic terminal and calcium ions in the synaptic cleft, which means $p_i = p_o = 0.5$. In this case, according to (3), the signal strength S is maximum; that is, $S = 1$.

Now, let us assume that the postsynaptic terminal performs a linear learning function, taking the strength of the input signal as input. Describing synaptic learning by a linear function is not uncommon. In fact, there are reasons to believe that this reflects the biological reality, and there have been many approaches in machine learning that model synaptic learning with linear functions [5–8]. The characteristic feature of the linear learning function presented here is

that it operates on the information content of the input signal rather than on the signal itself. To do so, it uses the standard dual logarithm to measure information, as investigated by Shannon in his seminal paper [9]. The following equation describes this linear relationship [3]:

$$I = -m \cdot \log_2 \left(\frac{1 - p_i}{p_i} \right) + c, \tag{4}$$

where I is the information learned by the linear model at the postsynaptic terminal. The two parameters that affect learning here are the slope m and the offset c of the linear model.

One of the main motivations of using this model is the form the equation of p_i assumes when we resolve (4) for p_i:

$$p_i = \frac{1}{1 + \exp\left(-\left((I - c) \cdot \ln(2)\right)/m\right)}. \tag{5}$$

Note that the dual logarithm has been converted into the natural logarithm in (5). This produces the same type of sigmoid function that was used in Section 2.1.2 to describe the molecule concentration inside of the cell membrane (see also (2)). Therefore, the linear information-theoretic model described here is in accordance with biology and the way concentrations are measured at membrane transitions. Section 2.4 will delve deeper into learning, but before continuing with learning, let us have a closer look at the formalization of Yin-Yang and how it relates to the linear information-theoretic model.

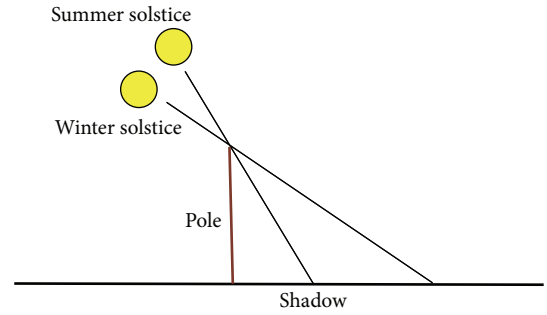


Figure 3: Yin-Yang daylight model [3].

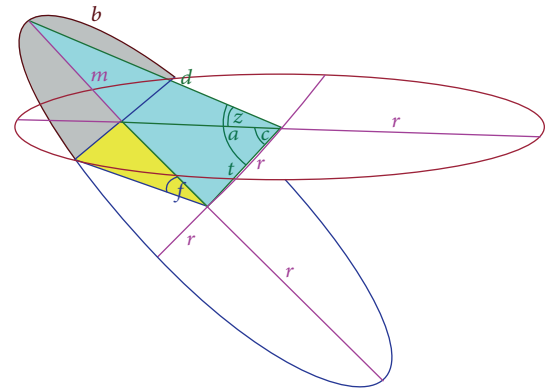


Figure 4: Solar circle for the summer solstice at 45° in the northern hemisphere (H. Glarner [10]).

2.3. Yin-Yang. Duality is not an informal concept with little meaning outside the philosophical realm. On the contrary, this section will show that the high-level concept of Yin and Yang has a well-defined mathematical expression. In fact, Yin and Yang can be formalized with the mathematical, information-theoretic model introduced in Section 2.2, as a linear function of information. To do so, this section shows how the classic symbol of duality, namely, the Yin-Yang symbol, can be rendered using the linear information model. Note that the rendering described in the following is an improvement to the work presented in [3] in that the median distance between the "physical" Yin-Yang symbol and the rendered "information-theoretic" symbol is smaller.

According to the results in [3], the Yin-Yang symbol depicts the length of a pole's shadow when measured at the same time each day throughout the year, as symbolized in Figure 3. Plotting the number of daylight hours for the first half of the year and the number of hours of darkness for the second half of the year in a circular polar plot then produces the Yin-Yang symbol. In the following, the mathematics used to measure the number of daylight hours for each day of the year will be different from the one used in [3]. In particular, the model presented by Glarner in [10] will be used here. This model is clearer from a mathematical point of view in that it does not consider the light refraction in the atmosphere of the earth. The next paragraph follows the description in [10].

The actual day of the year and the latitude of the observer both influence the length of the day. The perceived way of the sun around the planet can be viewed at as the boundary circle of the planet's disc. However, this constellation, in which the sun apparently circles along the disc's boundary, applies only at equinoxes and only at the North Pole. The further away the observer is from the North Pole (towards the equator), the more the surrounding circle is tilted along the west-east axis, until it is completely upright (perpendicular to the planet's disc) at the equator. Furthermore, there is also a shift of the circle away from the disc, along the obliquity of the ecliptic (connecting the centers of the two circles at an angle of 23.439°). This shift can be "upwards" (max. distance at the summer solstice) or "downwards" (max. distance at the winter solstice), depending on the actual latitude. Figure 4 shows the tilted and shifted solar circle for the winter solstice at 45° North. It is only the part b out of the whole circle in which the sun in visible. When the sun is running along the blue part of the circle in Figure 4, it is night for the observer. The way to computing the number of daylight hours is now to calculate the exposed part b in relation to the whole circle [10]. The equations necessary to do so require three input parameters, namely, *Axis*, *Lat*, and *Day*:

(i) *Axis*. This is the obliquity of the ecliptic, which is the angle between the rotation axis of the earth and its orbital plane. The obliquity of the earth is about 23.4°. It can be considered a constant for the purpose of this paper because its values change only slowly over a period of thousands of years.

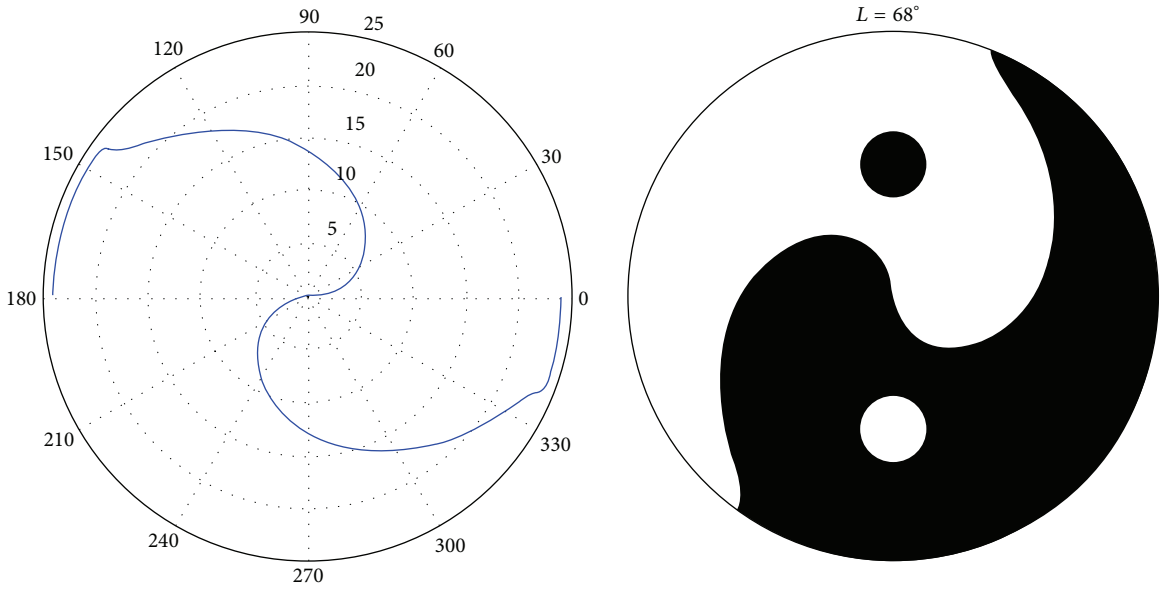


Figure 5: Yin-Yang symbol generated with the daylight model for $L = 68^{\circ}$ [3].

(ii) *Lat*. The latitude of the observer is in degrees. For example, for an observer at the equator, *Lat* is 0°. The latitude will increase for observers further north until it reaches 90° for an observer standing at the North Pole.

(iii) *Day*. This specifies the day of the year. *Day* runs from 0 to 364 for the first year, with 0.25 added from 365 for every completed year. Note that the day of year does not start with the astronomically quite arbitrary January 1st but with the day of the winter solstice in the first year of a four-year cycle [10].

Using these input parameters, computing the exposed fraction b of the sun's circle is a two-step process. First, the following intermediate result needs to be calculated:

$$m = 1 - \tan(Lat)\tan\left(Axis \cdot \cos\left(c \cdot Day\right)\right), \quad (6)$$

where $c \approx 0.0172$ is a constant. Note that the argument of the cos function is in radians, whereas the arguments of the tan functions are in degrees. The fraction b can then be computed as follows:

$$b = \frac{\arccos(1 - m)}{180}. \quad (7)$$

To get the number of hours the sun shines at the given *Day* and at the given Latitude *Lat*, b needs to be multiplied by 24. For a detailed derivation of these equations, readers are referred to [10].

Based on this computation of b, a linear regression function can be used to approximate the daily sunshine hours, as shown in [3]. For example, this produces the following approximation for one branch of the Yin-Yang symbol:

$$\Theta(p) = -3.208 \cdot \log_2(p) + 3.112. \quad (8)$$

The median error for this branch of the Yin-Yang symbol is 0.08 h, which is less than half of the error reported in [3]. This result confirms again that the Yin-Yang symbol describes a linear information-theoretic function, as presented in Section 2.2. The left-hand side of Figure 5 shows the daylight/nighttime hours plotted into a polar plot for Latitude $L = 68^{\circ}$. From the polar plot, the Yin-Yang symbol can be generated by rotating the plot by 90° and filling one area black and the other area white. The well-known dots of the Yin-Yang symbol are plotted halfway between the center of the circle and the circle's perimeter.

2.4. Learning. This section will have a closer look at learning. In particular, the linear learning equation in Section 2.2, (4), will be in the focus here. This equation provides two parameters that can be tuned for learning purposes: the slope m and the offset c. For the sake of simplicity, let us assume that the offset is constant and equals zero; that is, $c = 0$. The slope m then remains as the main parameter a synapse can learn. Furthermore, let us assume that the calcium ion concentration in the synaptic cleft close to the postsynaptic terminal, that is, p_i, is the input that needs to be learned. This makes sense because p_i is directly affected by the input stimulus and the calcium ion concentration p_i can be considered as teaching input to the postsynaptic terminal, where learning takes place. The main learning task of a synapse then involves adjusting the slope m of the linear learning function until it matches the concentration p_i; that is, $m = p_i$. For the completed learning task, (4) can be written as follows:

$$I = -p_i \cdot \log_2\left(\frac{1 - p_i}{p_i}\right). \quad (9)$$

If we now require that the learned concentration p_i, or slope parameter m, is equal to the input stimulus $(1 - p_i)/p_i$, then the following requirement needs to be satisfied [3]:

$$p_i = \frac{1 - p_i}{p_i},$$
$$p_i = \frac{\sqrt{5} - 1}{2} \quad \text{or} \quad p_i = \frac{-\sqrt{5} - 1}{2}, \tag{10}$$
$$p_i \approx 0.618 \quad \text{or} \quad p_i \approx -1.618.$$

This means that the learned concentration equals the input stimulus when the strength of the input stimulus matches the (reciprocal of) the Golden Ratio [11, 12]. As mentioned in [3], the results in the recent literature seem to indicate that the Golden Ratio plays a role in neural signal processing [13, 14]. This is another corroboration of the validity of the learning theory proposed here.

2.5. Dual Computation. The learning scheme represented by (9) describes the synaptic input-output relation after learning. However, the definition of input and output is arbitrary. In fact, the dual computation for which input and output change places is equally meaningful. To formulate a similar learning equation for the dual computation, (9) needs to be converted so that it applies in an antagonistic, symmetric way to the dual computation. This can be accomplished by transforming (9) into a symmetric form, under the assumption that the input signal (or stimulus) and the actually learned signal must be identical. Beginning with the linear learning equation, that is, Equation (9), the following transformations provide the desired symmetric form:

$$\begin{aligned} I &= -p_i \cdot \log_2\left(\frac{1 - p_i}{p_i}\right) \\ &= -{p_i}^2 \cdot \log_2(1 - p_i) \\ &= -p_i \cdot \log_2\left(1 - {p_i}^2\right) \\ &= -p_i \cdot \log_2\left(\sqrt{1 - p_i^2}^2\right). \end{aligned} \tag{11}$$

This can be written as follows:

$$\frac{I}{2} = -p_i \cdot \log_2\left(\sqrt{1 - p_i^2}\right), \tag{12}$$

which shows the symmetric relationship between the input signal and the learned output concentration. Note that I is multiplied by a scalar (1/2) in (12). This linear operation, however, only affects the scale of the learned information I. Because this affects I universally, it does not influence decision making. According to (12), for an input signal $\sqrt{1 - p_i^2}$, the learned output is p_i. Input and output thus define a point on the unit circle. The input can be considered the sine and the output the cosine of a point on the unit circle, as illustrated in Figure 6. All possible input-output combinations, or perceptual states, are points on the unit circle. Therefore, this relationship represents the desired symmetry between input and output. This symmetry allows to measure the input signal as p_i and the learned output signal as $\sqrt{1 - p_i^2}$. For example, if the input signal is one, that is, the argument of the logarithm in (12) is one, then the learned output signal and uncertainty $I/2$ will be zero. However, because the dual computation behaves antagonistic and exchanges input and output, the output uncertainty of the dual computation will be infinite. Conversely, if the input signal is zero, then the learned output signal $\sqrt{1 - p_i^2}$ is 1 and the corresponding output uncertainty $I/2$ is infinite. This means that the output uncertainty of the dual computation is zero. According to these results, the equilibrium state, in which both computations produce the same output uncertainty, is the state with the minimum overall uncertainty for both computations. Geometrically, this state is represented by a point on the unit circle for which both the sine and the cosine are $\sqrt{1/2}$. Therefore, sine and cosine can be considered Yin-Yang counterparts, and the output uncertainty I (or energy) is the Yin or Yang force, depending on the computation.

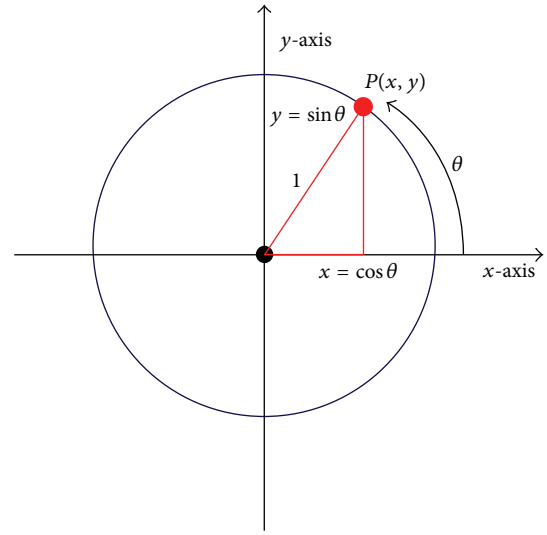


Figure 6: Perception points on the unit circle.

3. Results and Discussion

This section presents an application of the theoretical results derived above. For a lung screening application in which lung regions of chest X-rays are scanned for manifestations of tuberculosis, the machine confidence in abnormal regions will be graphically displayed.

3.1. Tuberculosis. Tuberculosis (TB) is the second leading cause of death from an infectious disease worldwide, after HIV, with a mortality rate of over 1.2 million people in 2010 [15]. With about one-third of the world's population having latent TB, and an estimated nine million new cases occurring every year, TB is a major global health problem. TB is an infectious disease caused by the bacillus Mycobacterium

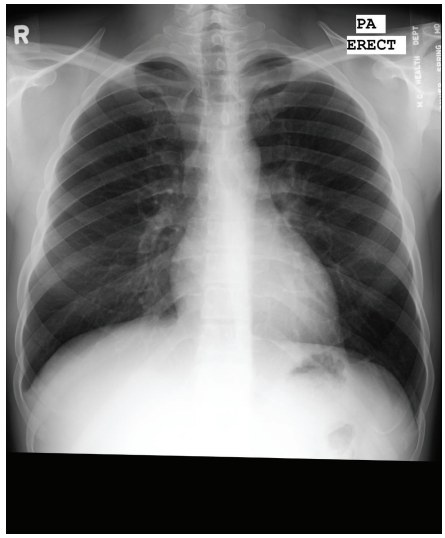


Figure 7: Normal chest X-ray (CXR).

tuberculosis, which typically affects the lungs. It spreads through the air when people with active TB cough, sneeze, or otherwise expel infectious bacteria. TB is most prevalent in sub-Saharan Africa and Southeast Asia, where widespread poverty and malnutrition reduce resistance to the disease. Moreover, opportunistic infections in immunocompromised HIV/AIDS patients have exacerbated the problem. The increasing appearance of multidrug resistant TB has further created an urgent need for a cost-effective screening technology to monitor progress during treatment. A posteroanterior radiograph (X-ray) of a patient's chest is a mandatory part of every evaluation for TB [16, 17]. Therefore, a reliable screening system for TB detection using radiographs would be a critical step towards more powerful TB diagnostics. An automated approach for detecting TB manifestations in chest X-rays (CXRs) would allow cost-effective mass screening of large populations that could not be managed manually [1].

3.2. Computing Confidence. As a step towards a fully automated system for TB screening in CXRs, and as an application example of the confidence values proposed here, this paper presents first experiments with a method for detecting manifestations of TB. The data being used is from Shenzhen No. 3 People's Hospital, China [18]. The CXRs were captured within a one-month period, mostly in September 2012, as part of the daily routine at Shenzhen No. 3 People's Hospital, using a Philips DR Digital Diagnost system. The data contains 342 abnormal images with manifestations of TB. For each image, a radiologist labeled the abnormal regions. Altogether 1671 regions have been annotated by two radiologists using the Firefly labeling tool [19], covering 18 different abnormalities, such as infiltrates, nodules, or effusions. For example, Figure 7 shows a normal chest X-ray, and Figure 8 shows a few samples of annotated abnormal lung patches. Each of the abnormal patches is represented by a set of histogram features that describe textures and shapes within the patch. In particular, the following features are used to describe the patch, each quantized into 32 bins: intensity, gradient magnitude, shape and curvature based on the Hessian eigenvalues, histogram of gradients, and local binary patterns [1]. All histograms are normalized and concatenated into a long 192-dimensional feature vector. To determine whether a region in a CXR is normal or shows a specific abnormality, the distance of the abnormal patches to the lung region can be computed by comparing the patch features with the features of the region. For the results presented here, the following histogram distance function is used to compute the distance between two normalized histograms A and B, where A_i and B_i denote the ith histogram bin, respectively, with $\sum A_i = 1$ and $\sum B_i = 1$:

$$D(A, B) = \frac{1}{2}\sum_{i=1}^{N} \left|A_i - B_i\right|. \quad (13)$$

This is done for each feature, with $N = 32$, and the average distance among all features is computed as the distance between patch features and region features. To compute the confidence value and obtain the desired trigonometric relationship, the similarity between patch and lung region is computed as follows:

$$S(A, B) = \sqrt{1 - D(A, B)^2}. \quad (14)$$

$S(A, B)$ is the confidence provided to the radiologist, indicating the machine confidence in the abnormality of the investigated lung region, given the presented patch.

3.3. Saliency Maps. By moving a known abnormal patch over a new input image and computing the similarity of the abnormal patch to the local lung region at each location, as described above, the entire input image can be screened for abnormal regions similar to the patch. When the confidence is recorded for each location, and different confidence values are displayed with different colors, a so-called saliency map can be generated. The saliency map highlights lung regions for which the machine is confident in their abnormality. For example, regions with high confidence can be marked in red, which provides direct feedback to the radiologist. The radiologist will be able to understand intuitively the different grades of confidence provided by the machine because the confidence is based on the sine/cosine relationship discussed above. Figure 9 shows three examples of saliency maps computed for the Chinese X-rays. The first saliency map on the left-hand side shows an abnormality in the right lung (note that left/right are interchanged when describing lungs). The red color signifies that the machine is very confident that this region is indeed abnormal. The abnormality indicated in the left lung is less reddish, which shows that the machine is less confident that this is indeed an abnormal region. The red lung boundary is the result of an automated lung segmentation method [20]. The radiologist can ignore any machine confidence outside this region, if displayed at all. A similar case is shown in the middle saliency of Figure 9. In this example, the machine has confidence in the entire right lung being abnormal. Finally, in the third example, the machine has identified a relatively small region of abnormality in the upper right lung, with a relatively high confidence.

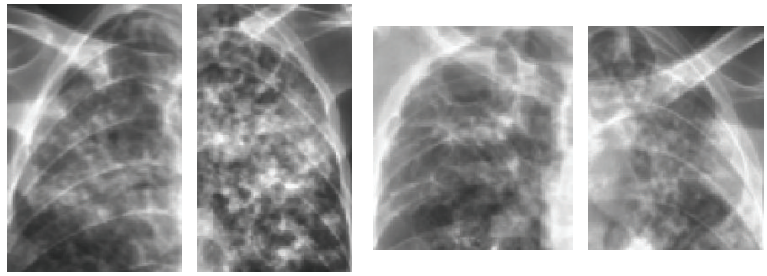


Figure 8: Abnormal lung patches identified by a radiologist.

Figure 9: Saliency maps for X-rays from Shenzhen No. 3 People's Hospital.

In these examples, the similarity function in (14) has been used to compute the confidence in the similarity of a lung region to a previously seen abnormal pattern. Alternatively, any classifier that outputs confidence values can be used for this task, such as the support vector machine used in [1] for discriminating between normal and abnormal lungs. Typically, confidence values are only considered when they exceed a threshold, which defines the operating point of the classifier and optimizes the classifier's sensitivity and specificity for a given cost function. For example, only dark red regions in the saliency maps in Figure 9 could be considered to reduce the false positive rate.

Note that the proposed method for representing confidence values is a postprocessing method. It can therefore be used in combination with any method that provides graded confidence in the similarity or dissimilarity of lung regions. In the lung screening application shown in this paper, the proposed method takes a dissimilarity measure of two lung regions and maps it to a similarity measure, according to the sine/cosine relationship. This would also work in the opposite direction, that is, mapping a similarity measure to a dissimilarity measure. The overall idea is that this sine/cosine relationship between similarity and dissimilarity is more intuitive, or natural, for human observers.

4. Conclusions

The paper proposes confidence values that can improve human-machine interaction for computer-aided diagnosis. The confidence values are motivated by neural signal processing and are intuitive in the senses that they are compatible with neural learning processes. Therefore, the operator of a computer-aided diagnostic system should be able to intuitively grasp the machine confidence and integrate it with his or her own confidence to reach a final diagnostic decision. The paper has revealed a direct connection to Chinese medicine and the dual concept of Yin-Yang, via a mathematical formalization of Yin and Yang. Furthermore, the paper shows that different learning states of a synapse, for which the input signal corresponds with the learned signal, can be represented by points on the unit circle. For the golden ratio, the learned signal corresponds with the actual signal. The state of equilibrium between Yin and Yang is the state for which sine and cosine are identical, which is the case at 45°. This state of equilibrium signifies neither normality nor abnormality and is thus a state of nonduality. As a practical example, the paper shows how the proposed confidence values can be used to highlight manifestations of tuberculosis in chest X-rays. In particular, the paper computes saliency maps where colors represent the magnitude of confidence values, indicating the confidence of the machine in the abnormality of a region in the chest X-ray. According to the theory set forth in the paper, the color intensities should be intuitive and the dynamic range of the colors (confidence values) should be similar to the representation a radiologist would use. As future work, the information-theoretic model presented in this paper could also help explain the efficacy of acupuncture in a more formal

framework. For example, the paper explains Yin and Yang as energies of two dual computations performed at a synapse. Furthermore, the paper provides a well-defined mathematical definition of the state of equilibrium between Yin and Yang, which minimizes the overall energy for a synapse. A high energy in the input or output of a synapse could indicate an abnormal state, such as an inflammation for example. Acupuncture could bring such an abnormal state back into the equilibrium state, where information can flow freely, and where there is no excessive heat or cold at the input or output of a synapse.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). The author would like to thank L. Folio and J. Siegelman for labeling abnormal lung regions.

References

[1] S. Jaeger, A. Karargyris, S. Candemir et al., "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, 2014.

[2] S. Jaeger, "A geomedical approach to chinese medicine: the origin of the Yin-Yang symbol," in *Recent Advances in Theories and Practice of Chinese Medicine*, H. Kuang, Ed., pp. 29–44, InTech, Rijeka, Croatia, 2012.

[3] S. Jaeger, "The neurological principle: how traditional Chinese medicine unifies body and mind," *International Journal of Functional Informatics and Personalised Medicine*, vol. 4, no. 2, pp. 84–102, 2013.

[4] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952.

[5] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1996.

[6] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, 1990.

[7] W. S. McCulloch and W. A. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.

[8] M. Minsky and S. Papert, *Perceptron*, MIT Press, 1970.

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[10] H. Glarner, Length of Day and Twilight, 2015, http://gandraxa.com/length_of_day.xml.

[11] M. Livio, *The Golden Ratio*, Random House, 2002.

[12] H. E. Huntley, *The Divine Proportion*, Dover Publications, 1970.

[13] H. Weiss and V. Weiss, "The golden mean as clock cycle of brain waves," *Chaos, Solitons & Fractals*, vol. 18, no. 4, pp. 643–652, 2003.

[14] B. Pletzer, H. Kerschbaum, and W. Klimesch, "When frequencies never synchronize: the golden mean and the resting EEG," *Brain Research*, vol. 1335, pp. 91–102, 2010.

[15] WHO, *Global Tuberculosis Report*, World Health Organization, 2012.

[16] C. C. Leung, "Reexamining the role of radiography in tuberculosis case finding," *The International Journal of Tuberculosis and Lung Disease*, vol. 15, no. 10, p. 1279, 2011.

[17] L. R. Folio, *Chest Imaging: An Algorithmic Approach to Learning*, Springer, Berlin, Germany, 2012.

[18] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.

[19] D. Beard, *Firefly- web-based interactive tool for the visualization and validation of image processing algorithms [M.S. thesis]*, University of Missouri, Columbia, Mo, USA, 2009.

[20] S. Candemir, S. Jaeger, K. Palaniappan et al., "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2014.

[21] R. M. Julien, *A Primer of Drug Action: A Comprehensive Guide to the Actions, Uses, and Side Effects of Psychoactive Drugs*, Worth Publishers, New York, NY, USA, 2005.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 148948, 7 pages
http://dx.doi.org/10.1155/2015/148948

Research Article

Patterns Exploration on Patterns of Empirical Herbal Formula of Chinese Medicine by Association Rules

Li Huang,[1] Jiamin Yuan,[1] Zhimin Yang,[1] Fuping Xu,[1] and Chunhua Huang[2]

[1] *The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, 111 Dade Road, Guangzhou 510120, China*
[2] *Neurology Department, Jiangxi Provincial Hospital of Chinese Medicine, Nanchang 330006, China*

Correspondence should be addressed to Zhimin Yang; yangyo@vip.tom.com

Received 21 November 2014; Revised 28 January 2015; Accepted 29 January 2015

Academic Editor: Zhaohui Liang

Background. In this study, we use association rules to explore the latent rules and patterns of prescribing and adjusting the ingredients of herbal decoctions based on empirical herbal formula of Chinese Medicine (CM). *Materials and Methods*. The consideration and development of CM prescriptions based on the knowledge of CM doctors are analyzed. The study contained three stages. The first stage is to identify the chief symptoms to a specific empirical herbal formula, which can serve as the key indication for herb addition and cancellation. The second stage is to conduct a case study on the empirical CM herbal formula for insomnia. Doctors will add extra ingredients or cancel some of them by CM syndrome diagnosis. The last stage of the study is to divide the observed cases into the effective group and ineffective group based on the assessed clinical effect by doctors. The patterns during the diagnosis and treatment are selected by the applied algorithm and the relations between clinical symptoms or indications and herb choosing principles will be selected by the association rules algorithm. *Results.* Totally 40 patients were observed in this study: 28 patients were considered effective after treatment and the remaining 12 were ineffective. 206 patterns related to clinical indications of Chinese Medicine were checked and screened with each observed case. In the analysis of the effective group, we used the algorithm of association rules to select combinations between 28 herbal adjustment strategies of the empirical herbal formula and the 190 patterns of individual clinical manifestations. During this stage, 11 common patterns were eliminated and 5 major symptoms for insomnia remained. 12 association rules were identified which included 5 herbal adjustment strategies. *Conclusion.* The association rules method is an effective algorithm to explore the latent relations between clinical indications and herbal adjustment strategies for the study on empirical herbal formulas.

1. Background

Empirical herbal formula is a relatively stable combination of herbs under the theoretical framework of Chinese Medicine. The development of empirical herbal formula undergoes long-term history of clinical practice and it is on the basis of the summarized experience of a variety of senior Traditional Chinese Medicine (TCM) doctors. Nowadays, the development of TCM clinical guidelines also stresses the use of empirical herbal formulas which are considered as valuable heritage of TCM. In order to promote the clinical use of these formulas, it is needed to interpret them with clear and accurate descriptions with rigorous scientific language. However, the previous research on empirical herbal formulas of TCM is subjective in the fact that they are simply summaries by the TCM doctors or their apprentices. This kind of subjective study methods causes two major demerits: at first, the adjustment of herbal components in the empirical formula is guided by the experience and knowledge of the doctors. Therefore, the subjective summary for the potential knowledge and implicit rules is not sufficient and complete [1]. And furthermore, it usually needs longer time to implement this kind of research. On the other hand, the results of the summarization by people with different knowledge background and preference are usually inconsistent; thus the study reliability and stability cannot be confirmed. The application of empirical herbal formulas can be improved and promoted to other doctors only after they are to be studied

and explored by objective and reliable methods. And then we will be able to inherit the valuable clinical experience of senior TCM doctors.

In order to carry out this task, we imitated a study on the empirical herbal formula si-ni and gui-gang-long-mu Tang by Professor Zhimin Yang, which is a herbal decoction originated from classic TCM prescriptions to treat insomnia. The association rules based on machine learning are used to explore the patterns of herbal combination strategy in real practice of Professor Zhimin Yang. The method of association rules is a well-known informatics approach to explore the latent correlation among a variety of factors [2]. The technology of informatics is being introduced to TCM, and a variety of studies were implemented previously. For example, Zhang et al. introduced a machine learning model to assess the efficacy of acupuncture by establishing a high quality training set [3] with the prediction accuracy over 80%. Liang et al. introduce a k nearest neighbor (kNN) algorithm to analyze different efficacy parameters of acupuncture for neck pain, and the research indicates high reliability for analyzing multidimensional outcomes [4].

In this study, an algorithm based on association rules is developed to explore the latent correlation among herbal components in empirical formulas and clinical indication and symptoms detected by examination. It is a good study example and attempt to use machine learning to explore the herb prescribing patterns of TCM doctors.

2. Materials and Methods

2.1. Ethics Statement. The research protocol of this study is reviewed and approved by the Ethics Committee of Guangdong Provincial Hospital of Chinese Medicine. All research activities comply with the principles of Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013). The personal information of all included patients is concealed against any exposure to the public. No information or clinical photographs relating to individual patients are included in this paper.

2.2. Study Rationale. Because of the theoretical and related clinical practice discrepancy of Western medicine and TCM, the TCM doctors actually undergo a different thinking procedure in their clinical visits by patients. Therefore, before we begin to develop the herbal prescriptions, we need to understand the development procedure of the formula or the mechanism of selecting empirical formula for a specific problem. An example of the procedure is illustrated in Figure 1.

As shown in Figure 1, TCM doctors view all organic systems of the human body as a whole. A TCM syndrome diagnostic system will be applied under the framework of TCM knowledge. In TCM syndrome diagnosis, the identification of a specific TCM syndrome includes the main symptoms which refer to the common characteristics of the disease and the key clinical symptoms for the empirical formula. The secondary symptoms refer to the nontypical or individual symptoms for the disorders, which will provide important reference to herbal adjustments. Recipe means the empirical formula and herbal combination adjustment for the empirical basic formula. Main symptoms are the main index for choosing the empirical formula. Then the secondary symptoms are the complementary index for choosing the herbal adjustment strategy for the basic empirical formula. The objective of this procedure is to combine the two ways (i.e., empirical formula + main symptoms and herbal adjustment + secondary symptoms) to develop the final individualized herbal prescription for the patient. And in order to explore this procedure, we separately study Way 1 and Way 2 by the algorithm of association rules.

2.3. Analytic Procedure. The analytic procedure for development patterns of the empirical herbal formula by TCM doctors is illustrated in Figure 2.

As indicated in Figure 2, the consideration patterns and procedure of a TCM doctor to develop an empirical formula are treated as a black box with implicit knowledge. The study is to visualize these patterns and procedure by the association rules method. The first process of choosing a basic empirical formula is to identify the main symptoms, in which the key indications for the empirical formula are to be detected with clear links to choosing. In clinical practice, TCM doctors make inevitable mistakes of inaccurately recognizing the indications and then choosing a wrong empirical formula. When the patients who are allocated to the ineffective group report ineffective outcome, they do have the indications of empirical formula. According to the comparison of symptoms between the invalid group and the effective group after the test, we analyzed the general indications of empirical formula and interpreted the relations between indications with association rules. And in the study of Way 2 (see Figure 1), the study objective is to find the relation between secondary symptoms and the herbal adjustment strategy based on the empirical herbal formula. According to the analytic result of Way 1, we cancelled the general indications among numerous symptoms. Then we found the individual symptoms for the patients. An association rule is built between the detected symptoms and herbal components as herbal adjustment strategy by the algorithm of association rules.

2.4. Inclusion and Exclusion Criteria for Patients

2.4.1. Inclusion Criteria. The patients will be included in the study if they satisfy the following criteria:

(1) the patient who is visited and treated by the empirical herbal formula of Professor Zhimin Yang for insomnia;

(2) the patient who is diagnosed with insomnia in accordance with diagnosis of insomnia by the Chinese Mental Disorder Diagnosis III (CCMD-3) [5] and guidelines for clinical trials of new drugs of Traditional Chinese Medicine for treatment of insomnia [6];

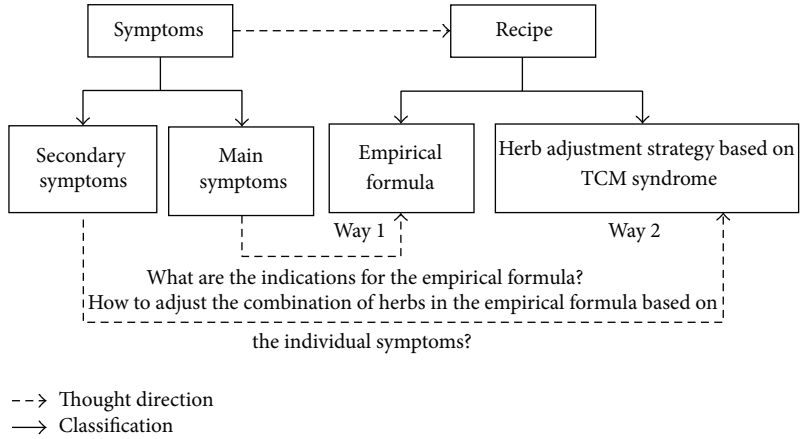


Figure 1: Procedure of developing and selecting an empirical formula.

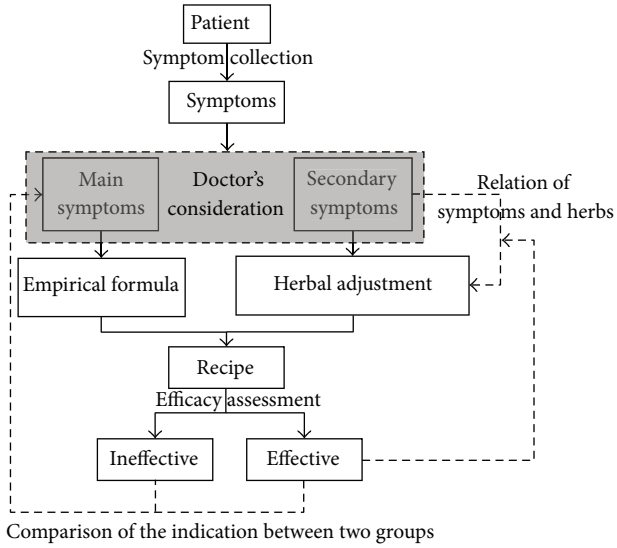


Figure 2: Flowchart of the study on pattern analysis of empirical herbal formula.

(3) the patient whose score is more than 7 by Pittsburgh Sleep Quality Index (PSQI);

(4) the patient whose age ranges from 18 to 75;

(5) the patient who is willing to participate and sign an informed consent document.

2.4.2. Exclusion Criteria

(1) Pregnancy or lactation women or those incapable of taking effective contraceptive measures in the study.

(2) Patients incapable of understanding or communicating.

(3) Patients who are participating in other clinical studies.

(4) Patients with mental retardation, alcohol dependence, substance abuse, and suicide.

(5) Patients with severe cardiovascular, cerebrovascular, lung, liver, kidney, and endocrine systems disease.

(6) Patients with serious sleep syndrome.

2.5. Data Source and Study Period. All included patients are from the Neurology Clinic of Professor Zhimin Yang in Guangdong Provincial Hospital of Chinese Medicine. The study period is one year from January 2010 to January 2011.

The observed outcomes include demographic characteristics, medical history based on the four diagnostic methods of TCM, and assessed outcomes by the Pittsburgh Sleep Quality Index (PSQI). The conditions of patients are to be evaluated before the treatment and one month after the treatment.

2.6. Efficacy Evaluation and Analytic Strategy. The judgment standard of efficacy is as follows: PSQI reduced rate = [(PSQI of before-treatment − PSQI of post-treatment 1 month)/PSQI of before-treatment] × 100%.

It will be considered as totally recovered if the PSQI reduced rate > 75%, as excellently effective if the PSQI reduced rate ranges from 51% to 75%, as effective if the PSQI reduced rate ranges from 25% to 50%, and as ineffective if the PSQI reduced rate < 25%. The totally recovered, excellently effective, and effective cases are all allocated to the effective group and the ineffective cases make up the ineffective group.

The data were input by Epidata and analyzed by SPSS18.0 and Clementine12.0. The test significant level is $\alpha = 0.05$ with two-sided Confidence interval.

2.7. Statistics. Record the data by Epidata. Analyze the data by SPSS18.0 chi-square and Clementine12.0. Association rules use GRI algorithm. Statistical test uses double-tailed test and α set 0.05.

3. Results

To the analytic procedure above, first, we should insure the effect of treatment, dividing the samples into an effective

Table 1: Treatment efficacy.

Efficacy judgment	Case	%
Ineffective	12	30.00
Effective	13	32.50
Excellently effective	14	35.00
Totally recovered	1	2.50
Total	40	100.00

group and ineffective group. Second, compare symptoms between effective group and ineffective group and initially find the main symptoms. Third, use the association rule to determine the relationship between the main initial symptoms. Fourth, remove the main symptom and analyse the association of remaining symptoms and drug by association rule. Present the results as follows.

3.1. Demographic Characteristics. 40 cases of confirmed insomnia patients from the Neurology Clinic of Professor Zhimin Yang are collected. All of them were treated by Professor Zhimin Yang with the empirical prescription of si-ni and gui-gang-long-mu Tang. The average age of the patients is 45.03 ± 12.17 years. 27.5% of them are male and 72.5% of them are female.

The longest medical history of insomnia ranges from 1 month to 240 months with the median of 120 months. 62.5% of the patients have midrange insomnia and 35.0% of them have serious insomnia.

3.2. Treatment Efficacy. The treatment efficacy is judged by PSQI reduction rate. The outcomes are classified into four levels: totally recovered, excellently effective, effective, and ineffective. The previous three levels (i.e., totally recovered, excellently effective, and effective) are allocated to the effective group and the last level is allocated to the ineffective group. We finally had 28 patients in the effective group and 12 patients in the ineffective group. The total effective rate was 70% (see Table 1).

3.3. Clinical Indication Filtering. 206 features were investigated on each case. We filtered 11 features from the effective group and the invalid group by chi-square test ($P < 0.05$). Eight symptomatic patterns (i.e., fear of cold, fatigue, irascibility, palpitations, chilliness, ice cold limbs, thirst, and hidrosis) show positive correlation with the curative effect. Three symptomatic features (anxiety, constipation, and a red tongue tip) show negative correlation with the curative effect (see Table 2).

3.4. Relations among Clinical Indications. In the analysis of the clinical indications of the ineffective group, 10 patterns were identified with relations (frequency of "thirst" was zero) by the a priori algorithm based on association rules (support 10%, confidence 80%). Finally 4 rules were extracted from the ineffective group (see Table 3).

In the analysis of the clinical indications of the effective group, 11 patterns were identified with relations by the a priori algorithm (support 60%, confidence 80%). Finally 15 rules were extracted from the effective group (see Table 4).

3.5. The Herbal Adjustment Strategy. In the analysis of the herbal adjustment strategy for the empirical formula, all prescriptions were adjusted from the original formula based on the patients' condition. 26 kinds of herbal medicine were involved in the adjustment. For individual symptoms, we gathered 206 features in total and 11 common symptoms were excluded and 5 main symptoms of insomnia were cancelled during the analysis. 190 individual symptoms remained. In the analysis of the effective group, the GRI algorithm of association rules was used, in which the minimum support rate is 10%, the minimum confidence level is 90%, the amount of preitems is at most 3, and the rule is only to display the real. We are to analyze the association rules between items, which included 28 strategies of herbal adjustment, and the preitems which included 190 symptoms. 12 association rules are extracted, which include five kinds of herbal adjustment strategies (see Table 5).

4. Conclusion

4.1. Application of Association Rules to Empirical Formula. The algorithm based on association rules is capable of identifying the relations between sets of items in a large database. In the recent studies [7, 8], the association rules algorithms were applied to the empirical herbal arrangement of TCM. They were mainly used in drugs concerted application of retrospective medical records. However, it is seldom applied to the analysis of relations of clinical indication and empirical herbal prescriptions. Relevant studies on this area have not yet been reported.

In our study, we introduced a database for prospective observed cases in order to evaluate the efficacy of the herbal prescription after the patients were treated by the empirical herbal formula. Thus the study methodology is quite different from the retrospective analysis of medical records in the past. We believe this new study method can render a reliable conclusion for the efficacy of TCM empirical formula.

4.2. Indications of the Empirical Formula. The symptoms of anxiety, palpitations, and a red tongue tip show negative correlation with the efficacy of TCM empirical formula by univariate analysis. And the rules from the ineffective group had the symptom of palpitations. However, the rules from the effective group do not have the palpitations symptom. Therefore, the comparison among anxiety, a red tongue tip, and palpitations symptoms is important. And palpitations are an inhibition of the empirical prescription application.

In TCM, the rules extracted from the ineffective group (i.e., chilly → afraid of cold) and the effective groups (i.e., chilly → fatigue, ice cold limbs → afraid of cold) are regarded as the indication for the Yang or Qi asthenia syndrome. These rules exist both in the ineffective group and in the effective group. Therefore, the Yang or Qi asthenia syndrome is a necessary condition for the application of empirical herbal prescription. However, not all the patients who had Yang or

Table 2: Related clinical indications.

Symptom	Group	Negative (%)	Positive (%)	χ^2	P
Fear of cold	Ineffective	8 (66.7)	4 (33.3)	5.08	0.04
	Effective	8 (28.6)	20 (71.4)		
Fatigue	Ineffective	4 (33.3)	8 (66.7)	4.52	0.03
	Effective	2 (7.1)	26 (92.9)		
Anxiety	Ineffective	8 (66.7)	4 (33.3)	4.52	0.03
	Effective	26 (92.9)	2 (7.1)		
Irascibility	Ineffective	9 (75.0)	3 (25.0)	12.06	0.01
	Effective	5 (17.9)	23 (82.1)		
Palpitations	Ineffective	10 (83.3)	2 (16.7)	13.41	0.01
	Effective	6 (21.4)	22 (78.6)		
Chilliness	Ineffective	10 (83.3)	2 (16.7)	7.62	0.01
	Effective	10 (35.7)	18 (64.3)		
Ice cold limbs	Ineffective	11 (91.7)	1 (8.3)	9.31	0.01
	Effective	11 (39.3)	17 (60.7)		
Thirst	Ineffective	12 (100.0)	0 (0.0)	7.35	0.01
	Effective	16 (57.1)	12 (42.9)		
Hidrosis	Ineffective	11 (91.7)	1 (8.3)	6.22	0.03
	Effective	14 (50.0)	14 (50.0)		
Constipation	Ineffective	9 (75.0)	3 (25.0)	7.59	0.02
	Effective	28 (100.0)	0 (0.0)		
A red tongue tip	Ineffective	9 (75.0)	3 (25.0)	7.57	0.02
	Effective	28 (100.0)	0 (0.0)		

Table 3: Extracted rules from the ineffective group.

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Fatigue	Palpitations	16.67	100.00	1.50
Fear of cold	Chilliness	16.67	100.00	3.00
Fatigue	Constipation and fear of cold	16.67	100.00	1.50
Fear of cold	Constipation and fatigue	16.67	100.00	3.00

Table 4: Extracted rules from the effective group.

Consequent	Antecedent	Support (%)	Confidence (%)	Lift
Palpitations	Irascibility	82.14	86.96	1.11
Palpitations	Irascibility and fatigue	75.00	85.71	1.09
Fear of cold	Ice cold limbs	60.71	82.35	1.15
Fatigue	Ice cold limbs	60.71	94.12	1.01
Fatigue	Chilliness	64.29	100.00	1.08
Fatigue	Fear of cold	71.43	95.00	1.02
Fatigue	Palpitations	78.57	90.91	0.98
Fatigue	Irascibility	82.14	91.30	0.98
Fatigue	Palpitations and irascibility	71.43	90.00	0.97
Irascibility	Ice cold limbs	60.71	88.24	1.07
Irascibility	Chilliness	64.29	83.33	1.01
Irascibility	Palpitations	78.57	90.91	1.11
Irascibility	Fatigue	92.86	80.77	0.98
Irascibility	Chilliness and fatigue	64.29	83.33	1.01
Irascibility	Palpitations and fatigue	71.43	90.00	1.10

Table 5: Result of the association rules for relations between herbal adjustment strategy and clinical patterns in the effective group.

Preitems	Postitems	Support rate (%)	Confidence level (%)	Lift
Thirst and loose stools	White atractylodes rhizome	10.71	100.00	2.80
Thirst and loose stools	*Poria cocos*	10.71	100.00	2.33
Dizziness	*Poria cocos*	10.71	100.00	2.33
Dizziness and sense of suppression in the chest	*Poria cocos* with Rhizoma Pinellinae Praeparata	10.71	100.00	2.80
Greasy fur of tongue and short breath	*Poria cocos* with Rhizoma Pinellinae Praeparata	10.71	100.00	2.80
Greasy fur of tongue and pale complexion and short breath	*Poria cocos* with Rhizoma Pinellinae Praeparata	10.71	100.00	2.80
Postmenopausal women	Fresh dogwood with *Fructus mume*	17.86	100.00	1.56
Irritability	Fresh dogwood with *Fructus mume*	10.71	100.00	1.56
Sore throat	Fresh dogwood with *Fructus mume*	10.71	100.00	1.56
Tinnitus	Equivalent Guizhi with Aglaophotis radix herbaceu	10.71	100.00	2.00
Chest pain	Equivalent Guizhi with Aglaophotis radix herbaceu	10.71	100.00	2.00
Forgetfulness and breast pain	Equivalent Guizhi with Aglaophotis radix herbaceu	10.71	100.00	2.00

Qi asthenia syndrome had effective outcomes to the empirical prescription. Therefore, the Yang or Qi asthenia syndrome is not a sufficient condition for the empirical prescription application.

Among the extracted rules from the effective group, 10 rules include the irascible symptom. These rules merge red the Yang or Qi asthenia syndrome but only exist in the effective group. Therefore, based on the Yang or Qi asthenia syndrome, we believe the irascible symptom is a sufficient condition of the empirical prescription application.

4.3. Herbal Adjustment Strategy for the Empirical Formula. In accordance with the TCM theory, the application of the empirical formula should be based on the main clinical symptoms or patterns of the patients. The herbal adjustment strategies for the empirical herbal formula depend on the symptoms of the individual patient. If we analyze all the collected clinical characteristics, it is not possible to discriminate whether the changes of herbal ingredients are the cause of temporary application ideas of empirical formula or a fix herbal adjustment strategy. Therefore, in this study we set 190 clinical features of disease excluding the indications of empirical formula and insomnia main symptoms as preitems. In that case, we can distinguish application ideas of the empirical formula and the drug of addition and subtraction. Because of rarely using some herbs, we only excavated 5 in a total of 28 kinds of addition and subtraction herbs.

4.4. Feedback by the Original Expert. Professor Yang gave her feedback to the above data analysis results. She considered that the indications of empirical prescription corresponded to the application thinking, and the 10th and 11th rules in addition and subtraction and clinical features are not her original application ideas of drugs of addition and subtraction. However, as a complement, they are important to further study in clinical practice. The remaining rules are consistent with her application ideas of herbal adjustment strategy.

5. Discussions

In summary, when compared with the conventional method based on statistics, the machine learning method of association rules has specific merits in analyzing the indications of empirical herbal prescription of TCM. And it is also good at exploring the relations among herbal adjustment for empirical formula. This method provides an effective and reliable approach for the study on TCM empirical herbal prescriptions in the future.

Conflict of Interests

The authors declare that they have no conflict of interests.

Authors' Contribution

All authors read and approved the final paper. Li Huang is responsible for the entire work and developed the study project. Jiamin Yuan gave advice on the analytic algorithm and drafted and finished the paper. Zhimin Yang is the coordinator of the study. Fuping Xu and Chunhua Huang collected clinical data for the study.

Acknowledgments

The work is supported by the National Basic Research Program of China (973 Program: 2011CB505404). And it is also funded by the Sleep Laboratory of Guangdong Provincial Finance Bureau (2012KT1420) and the Administration of Traditional Chinese Medicine in Guangdong Province (20131217).

References

[1] Y. P. Fan, W. Mao, Y. B. Lu et al., "To investigate the function of inheritance work of clinical studio in TCM from the perspective

of tacit knowledge management," *Journal of Administration of Traditional Chinese Medicine*, vol. 17, no. 3, p. 193, 2009.

[2] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, 1993.

[3] G. Zhang, Z. Liang, J. Yin, W. Fu, and G.-Z. Li, "A similarity based learning framework for interim analysis of outcome prediction of acupuncture for neck pain," *International Journal of Data Mining and Bioinformatics*, vol. 8, no. 4, pp. 381–395, 2013.

[4] Z. H. Liang, G. Zhang, G. Z. Li, and W. B. Fu, "An algorithm for acupuncture clinical assessment based on multi-view KNN," *Journal of Computational Information Systems*, vol. 8, no. 21, pp. 9105–9112, 2012.

[5] Psychiatry Branch of the Chinese Medical Association, *Chinese Classification and Diagnostic Criteria of Mental DisordersIII (CCMD-3)*, Shandong Science and Technology Press, Jinan, China, 2001.

[6] Ministry of Health of the People's Republic of China, *Guidance Principle of Clinical Study on New Drug of TCM*, vol. 1, Ministry of Health, Beijing, China, 1993.

[7] H. Cheng, N. N. Xiong, Y. J. Jiang, M. H. Wu, and J. Zhou, "Chinese medicine treatment of primary insomnia clinical trial design features," *Liaoning Traditional Chinese Medicine*, vol. 33, no. 12, pp. 1550–1551, 2006.

[8] M. Chen and S. H. Hang, "Disease association rules in traditional Chinese medicine diagnosis of the syndrome," *Journal of Chinese Medicine*, vol. 4, no. 5, pp. 14–16, 2004.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 125736, 9 pages
http://dx.doi.org/10.1155/2015/125736

Research Article

Syndrome Differentiation Analysis on Mars500 Data of Traditional Chinese Medicine

Yong-Zhi Li,[1] Guo-Zheng Li,[2] Jian-Yi Gao,[1] Zhi-Feng Zhang,[3] Quan-Chun Fan,[1] Jia-Tuo Xu,[3] Gui-E Bai,[1] Kai-Xian Chen,[3] Hong-Zhi Shi,[1] Sheng Sun,[4] Yu Liu,[1] Feng-Feng Shao,[4] Tao Mi,[1] Xin-Hong Jia,[5] Shuang Zhao,[1] Jia-Chang Chen,[4] Jun-Lian Liu,[1] Yu-Meng Guo,[4] and Li Ping Tu[3]

[1]*China Astronaut Research and Training Center, Beijing 100094, China*
[2]*Data Center of Traditional Chinese Medicine, China Academy of Chinese Medicine Science, Beijing 100700, China*
[3]*Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[4]*Department of Control Science and Engineering, Tongji University, Shanghai 201804, China*
[5]*Shanghai Daosheng Medical Technology Co. Ltd., Shanghai 201203, China*

Correspondence should be addressed to Guo-Zheng Li; drgzli@gmail.com

Received 10 October 2014; Accepted 18 December 2014

Academic Editor: Zhaohui Liang

Mars500 study was a psychological and physiological isolation experiment conducted by Russia, the European Space Agency, and China, in preparation for an unspecified future manned spaceflight to the planet Mars. Its intention was to yield valuable psychological and medical data on the effects of the planned long-term deep space mission. In this paper, we present data mining methods to mine medical data collected from the crew consisting of six spaceman volunteers. The synthesis of the four diagnostic methods of TCM, inspection, listening, inquiry, and palpation, is used in our syndrome differentiation. We adopt statistics method to describe the syndrome factor regular pattern of spaceman volunteers. Hybrid optimization based multilabel (HOML) is used as feature selection method and multilabel k-nearest neighbors (ML-KNN) is applied. According to the syndrome factor statistical result, we find that qi deficiency is a base syndrome pattern throughout the entire experiment process and, at the same time, there are different associated syndromes such as liver depression, spleen deficiency, dampness stagnancy, and yin deficiency, due to differences of individual situation. With feature selection, we screen out ten key factors which are essential to syndrome differentiation in TCM. The average precision of multilabel classification model reaches 80%.

1. Introduction

With the development of the three-phase strategy, our manned space programme entered a new manned space station construction stage. How spaceman adapts to longtime isolation environment and overcomes the challenges from the aspects of body, mind, and spirit became a burning question in the area of manned space [1].

Mars500 mission was a psychology and physiology isolation experiment conducted by Russia, the European Space Agency, and China, in preparation for an unspecified future manned spaceflight to the planet Mars. A total of 640 experiment days were scheduled between 2007 and 2011, divided into three stages of differing length. During each stage, the crew of volunteers lived and worked in a mockup spacecraft. Communication with outside world was limited, and it was conducted with a realistic time delay of up to 25 minutes, to simulate the real-life communications lag between Mars and Earth. The final stage of the experiment was intended to simulate a 520-day manned mission. The mission was intended to yield valuable psychological and medical data on the effects of the planned long-term deep space mission. The experiment permitted the study of the technical challenges, work capability of crew, and management of long-distance spaceflight. Communications lag, autonomy, resource rationing, health, conditions of isolation,

and hermetically closed, confined environment are the main peculiarities of the Martian flight.

As a complete medical system, TCM plays an indispensable role in medical care in China. Different from the reductionism thinking mode of western medicine, TCM is based on the holistic and systematic ideas. TCM practices are believed to be effective by many patients and scientists, sometimes offering palliative efficiency, while the practices of western medicine fail or are unable to provide treatment. We have reason to believe that TCM can play an important role in health security mission in long-term space flight.

In Mars500 mission, inspection, inquiry, and palpation of TCM were applied to study the state of human life activities in longtime isolation environment and to interpret the features and change rules. In this research, digital instrument was used to collect TCM diagnostic information of the spaceman volunteers. The scale of syndrome and symptom was designed to quantize the degree syndrome and symptom. Then, we got the digital and normalized information which was used to find the relationship between the symptoms and syndromes.

In our research, we apply statistics method to describe the syndrome factor regular pattern and find that qi deficiency is a base syndrome pattern throughout the entire experiment process. At the same time, there are different associated syndromes such as liver depression, spleen deficiency, dampness stagnancy, and yin deficiency, due to the differences of individual situation. Then, we search the objective and inherent relationship between the symptoms and syndromes.

In clinical practice, the relationship between symptoms and syndromes can be seen as multilabel classification problem in which many symptoms may present various syndromes. Many researches have been down by using multilabel learning in biomedical feild [2–6]. In our work, hybrid optimization based multilabel (HOML) [7] is used to select related features, and multilabel k-nearest neighbors (ML-KNN) [8] is applied as the multilabel classifier. In our model, ten important symptoms for syndrome differentiation are selected and they are all from inspection which includes complexion and tongue diagnosis. Then, we analyze the characteristics of complexion and tongue picture, finding that the changes of complexion and tongue picture are consistent with changes of syndromes.

The remaining of the paper is organized as follows: in Section 2, we introduce the data collection and preprocessing methods, the feature selection HOML, and the ML-KNN; we give the results and discussions of our research in Section 3; then, we make a conclusion.

2. Methods

Data collection, preprocessing, and data features TCM interpretation and software analysis were made before we got the dataset. The details are as follows.

2.1. Data Collection. According to the scheme of the TCM research of human body in Mars500 longtime isolation environment, DS01-T and auxiliary diagnosis system were used to collect TCM data from six spaceman volunteers every two weeks from June 3, 2010, to November 4, 2011. Inquiry, inspection (complexion and tongue picture), and palpation data of the spaceman volunteers are collected. This work is sponsored by China Astronaut Research and Training Center and all investigators signed the informed consent.

2.2. Data Preprocessing. The collected data were preprocessed and the ones meeting the requirement were stored in the database.

2.2.1. Inquiry Data Preprocessing. Inquiry data in the scale of the inquiry were selected and united as the clinical terms by the panel of the TCM. The invalid data caused by mistakes of eyes or writing were eliminated. For example, the choice should be "before meals," but the spaceman volunteers selected "after meals" instead. The same case may also happen to "daytime" and "night." These mistakes were caused by writing obviously and were eliminated directly. There were also some logical conflicting mistakes. For example, the "bulimia" and "loss of appetite" may be selected at the same time. These mistakes were modified by eliminating one of the options according to the analysis of TCM experts.

2.2.2. Inspection Data Preprocessing. The invalid data in the inspection data caused by accident were removed, such as blur pictures caused by the failure of the camera focusing or the shake of volunteers' tongues.

2.2.3. Palpation Data Preprocessing. Palpation data were rectified by the panel of the TCM experts and the invalid ones caused by incorrect installation of sensor or sudden shake of volunteers were removed. The mistaken data contained the pulse information which could not be recognized by software and the experts or the signal that results from main peak of the pulse was less than 10 mmHg.

2.3. TCM Interpretation and Software Analysis of Data Features. Interpretation of the TCM experts and analysis of the software were introduced to the interpretation and analysis of data features. The details can be as follows.

2.3.1. TCM Experts Interpretation of Data Features. The panel of TCM consisted of three chief physicians whose clinical experience was over 20 years. Three experts worked alone at first and then compared their results. Results would be obtained as final ones when their results were consistent. Otherwise, the final results would be made by the panel of TCM discussion with the other three TCM experts:

(i) interpretation of inspection data: interpretation of inspection data was generated from analysis of tongue and facial pictures; then, the qualitative description and possible medical significance of tongue and facial features were given;

(ii) interpretation of palpation data: palpation pictures of spaceman volunteers were analyzed to generate the interpretation of palpation data; the information of volunteers' palpation pictures, such as pulse position, pulse rate, pulse power, rhythm, and pulse shape was

analyzed combined with common pulse condition model in the former database; the qualitative interpretation and the possible medical significance of the pulse condition features were given and used as one basis of the syndrome differentiation;

(iii) interpretation of inquiry data: the descriptions of part of symptom in the syndrome and symptom scale were translated into standard terminology; the symptom was regarded as main symptom or general symptom by TCM experts according to the frequency and degree of the symptom and the clinical experience; then, inquiry results were interpreted to analyze volunteers' health condition;

(iv) interpretation of syndrome: analysis of syndrome was based on the information fusion of the inquiry, complexion, and tongue picture and pulse condition.

2.3.2. Software Analysis of the Data Features

(i) Interpretation of tongue picture and complexion features: interpretation of tongue picture and complexion features generated from analysis software is listed in Table 1. Results from analysis software should be considered with the ones from TCM experts.

(ii) Interpretation of pulse condition features: pulse pictures were obtained from palpation data. Figure 1 presents the basic structure of pulse picture. Relationship between amplitude and phase of pulse wave was analyzed by software using time-domain analyzing method. The analysis content contained the recognition of height of wave and gorge, the corresponding value, and the area of the pulse picture. Notations in Figure 1 can be interpreted as follows:

h_1: amplitude of the main wave,
h_3: front wave amplitude of dicrotic pulse,
h_4: amplitude of dicrotic notch,
h_5: amplitude of dicrotic pulse,
t_1: acute ejection period value,
t_4: systole value,
t_5: diastole value,
t: pulsation period,
w: one-third of h_1.

Then, the features of pulse condition were interpreted by using frequency-domain analysis and time-frequency analysis, based on principles of hemodynamic methods.

2.4. Dataset Description. Through above preprocessing, we get a data set with 222 cases in which each case has 389 features and 11 labels. The inspection data contains 245 features, the palpation data contains 30 features, and the inquiry data have 114 features.

2.5. Feature Selection. In TCM diagnosis, a patient may be associated with more than one symptom, and its computer-aided diagnosis is a typical application in the domain of multilabel learning of high-dimensional data. It is common that a great deal of symptoms can occur in TCM diagnosis, which affects the modeling of diagnostic algorithm. Feature selection entails choosing the smallest feature subset of relevant symptoms and maximizing the generalization performance of the model. In this work, HOML is used to analyze feature selection for multilabel TCM data. HOML combines the relatively strong global optimization ability of simulated annealing algorithm (SA) [9], genetic algorithm (GA) [10], and the strong local optimization capability of greedy algorithm [11]. The following is the details of HOML, which organizes a search in three stages.

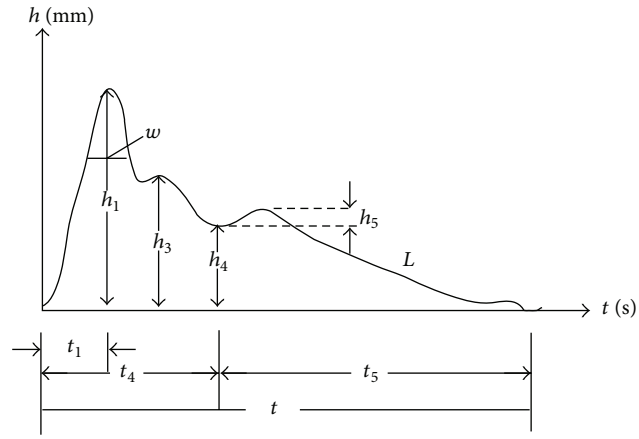


Figure 1: Basic structure of pulse picture.

Stage 1. A simulated annealing (SA) is employed to guide the global search in a solution space. SA would accept every solution if the temperature is very high, which then yields a near random search through the search space. As the temperature becomes close to zero, only improvements are accepted. The SA is run for approximately 50% of the total time available.

Stage 2. A GA is employed to perform optimization. The GA population is set at 100. The initial population consists of the best solutions detected by SA. The crossover operator enables the good solutions to exchange information, and the mutation operator in GA introduces new genes into the population and retains genetic diversity. The GA runs for about 30% of total time spent by HOML to find the optimal feature subset solution.

Stage 3. A hill-climbing feature selection algorithm is applied. The greedy algorithm performs a local search on the k-best solutions on the k-best (k represents the dimensionality of feature) solutions given by two global optimization algorithms (SA and GA).

2.6. Multilabel Classifier. In our study, the multilabel k-nearest neighbour (ML-KNN) algorithm is used to analyse syndromes models. KNN is an algorithm whose idea is to search for the nearest point in training dataset [12]. In KNN algorithm, an instance is regarded as a point. And the label of

Table 1: Qualitative and quantitative features of inspection.

Parameter	Qualitative features	Quantitative features
Tongue color	Dark, pale red, red, crimson, pale purple, dark purple	RGB, HSV mean, feature color proportion
	Red tongue tip	Tongue tip RGB, HSV mean color difference between tongue tip and the rest
Fur color	White, yellow and white, yellow, and gray black	RGB, HSV mean, proportion of feature color
Texture of fur	Thick fur, thin fur	Density degree
	Curdy fur, greasy fur	Density degree, distribution
	Less moss	Coverage ratio
	No moss	Coverage ratio
	Peeling fur	Defect area
Tongue shape	Fat, thin	Degree of circularity, length-width ratio, with degree
	Fast insertion	Number and color of tongue tip's circle dot
	Crack	Proportion of crack area
	Indentation	Size of indentation area
	Petechia	Number and color of tongue tip's circle dot
	Ecchymosis	Ecchymosis position, ecchymosis RGB, HSV mean
Complexion	Blue, red, yellow, white, black, normal	RGB, HSV mean, feature color proportion
Gloss	Glossy, few gloss, no gloss	RGB, HSV mean, feature color proportion
Lip color	Dark, red, dark red, purple	RGB, HSV mean, feature color proportion

a test instance is probably similar to that of several nearest points. Based on this theory, the algorithm of KNN is to search for k train instances nearest to the test instance, then according to their labels, to predict the label of the test instance. Compared with other mining algorithms, the advantage of KNN lies in simpler training process, better efficiency, and forecast accuracy.

In the multilabel data, just simple splitting may result in data loss because of the relationship between each label. At this condition, multilabel KNN would be a better choice to solve this problem. ML-KNN is the first multilabel lazy learning algorithm, which is derived from the popular k-nearest neighbor (KNN) algorithm. The basic idea of ML-KNN is to adapt k-nearest neighbor techniques to deal with multilabel data, where maximum a posteriori (MAP) rule is applied to make prediction with the labeling information embodied in the neighbors [13]. In a word, the labels of each instance are judged by its nearest neighbors. Brief introduction of this algorithm is shown as follows.

Step 1. The conditional probability distribution between each instance and its associated label set would be calculated at first.

Step 2. Calculate the distance between each test instance and the training instances; then find k-nearest instances for each test instance.

Step 3. For each test instance, its forecast results would be acquired according to the labels of k-nearest training instances and the conditional probability associated to each label.

Step 4. Evaluate the forecast results according to multilabel evaluation criteria.

2.7. Experimental Design and Evaluation. In our experiment, 5-fold cross-validation is utilized to test the accuracy of the classification. We firstly build three classification models with four types of diagnostic fusion data, inspection data, and palpation data, respectively, by using ML-KNN. Then, we apply HOML to the model which obtains the best performance.

Let X denote the domain of instances and let $Y = \{1, 2, \ldots, Q\}$ be the finite set of labels. The multilabel classification problem can be formulated as follows. Given a training set $T = \{(x_1, Y_1), (x_2, Y_2), \ldots, (x_m, Y_m)\}$ $(x_i \in X, Y_i \in Y)$, drawn from an unknown distribution D, the goal of the learning system is to output a multilabel classifier $h : X \to 2^Y$ which optimizes some predefined criteria. The learning system will tend to output larger values for labels in Y_i than those which are not in Y_i according to a real-valued function of the form $X \times Y \to R$. For example, if $y_1 \in Y_i$ and $y_2 \notin Y_i$, then $f(x_i, Y_1) > f(x_2, Y_2)$.

A ranking function $\text{rank}_f(\cdot, \cdot)$, which can be the transformed form of the real-valued function $f(\cdot, \cdot)$, maps the outputs of $f(x_i, y)$ for any $y \in \mathbb{Y}$ to $\{1, 2, \ldots, Q\}$. For example, for the $f(x_i, y_1) > f(x_2, y_2)$, there will be $\text{rank}_f(x_i, y_1) < \text{rank}_f(x_i, y_2)$. Then, the multilabel classifier $h(\cdot)$ can be represented as $h(x_i) = y \mid f(x_i, y) > t(x_i)$, $y \in Y$, in which $t(\cdot)$ is a threshold function.

It is worth noting that in multilabel learning paradigm, various evaluation criteria have been proposed to measure the performance of a multilabel learning system. Given a test set $S = \{(x_1, Y_1), (x_2, Y_2), \ldots, (x_p, Y_p)\}$, the following multilabel evaluation metrics are used in this paper [8].

(1) Hamming loss is defined as

$$\text{hloss}_S(h) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{Q} \left| h(x_i) \, \Delta Y_i \right|, \quad (1)$$

where Δ stands for the symmetric difference between two sets. Note that when $|Y_i| = 1$, for all instances, a multilabel system is in fact a multiclass single-label one and the hamming loss is $2/Q$ times the usual classification error. Hamming loss is used to evaluate how many times an instance-label pair is misclassified. The smaller the value of $\text{hloss}_S(h)$, the better the performance.

(2) One-error is defined as

$$\text{one-error}_S(f) = \frac{1}{p}\sum_{i=1}^{p}\left[\left[\left[\left[\arg\max_{y\in Y} f(x_i, y)\right] \notin Y_i\right]\right]\right], \quad (2)$$

where for any predicate π, $[\pi]$ equals 1 if π holds and 0, otherwise. Note that, for single-label classification problems, the one-error is identical to ordinary classification error. One-error is used to evaluate how many times the top-ranked label is not in the set of proper labels of the instance. The smaller the value of $\text{one-error}_S(f)$, the better the performance.

(3) Ranking loss is defined as

$$\text{rloss}_S(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|\left|\overline{Y_i}\right|}\left|\left\{(y_1, y_2) f(x_i, y_1) \le f(x_i, y_2), (y_1, y_2) \in Y_i \times \overline{Y_i}\right\}\right|, \quad (3)$$

where $\overline{Y}$ denotes the complementary set of Y in y. Ranking loss is used to evaluate the average fraction of label pairs that are reversely ordered for the instance. The smaller the value of $\text{rloss}_S(f)$, the better the performance.

(4) Average precision is defined as

$$\text{avgprec}_S(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|} \times \sum_{y\in Y_i}\frac{\left|y' \mid \text{rank}_f(x_i, y') \le \text{rank}_f(x_i, y), y' \in Y_i\right|}{\text{rank}_f(x_i, y)} \quad (4)$$

and is used to evaluate the average fraction of labels ranked above a particular label $y \in Y$ which actually are in Y. The bigger the value of $\text{avgprec}_S(f)$, the better the performance.

3. Results and Discussion

3.1. Syndrome Factor Statistical Result. In this section, statistics method is applied to describe the syndrome factor regular pattern and the result is shown in Figure 2. From Figure 2, we find that qi deficiency is a base syndrome pattern throughout the entire experiment process and, at the same time, there are different associated syndromes such as liver depression, spleen deficiency, dampness stagnancy, and yin deficiency, due to differences of individual situation.

3.2. Results by Using Multilabel Learning Methods. Results of ML-KNN without HOML are shown in Table 2 and results of ML-KNN with HOML are shown in Table 3. Comparing Tables 2 and 3, we find that ML-KNN with HOML obtains better performance than that without HOML which means that feature selection plays an important role in our model. Feature selection results of our model are shown in Table 4. As shown in Table 4, we can see that the ten important features selected are all in complexion and tongue diagnosis. Then, we analyze the characteristics of complexion and tongue picture in the following.

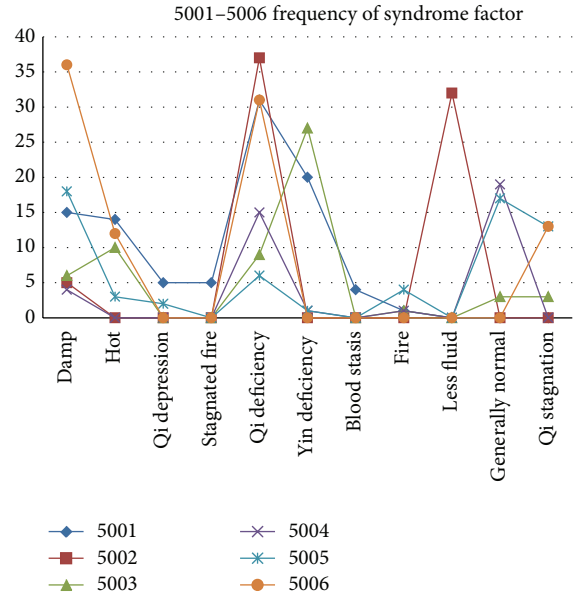


Figure 2: Frequency of syndrome factor.

Table 2: Classification results without HOML.

Dataset	Average precision	Ranking loss	One error	Hamming loss
Four diagnostic fusion data	0.78	0.10	0.26	0.12
Inspection data	0.77	0.10	0.31	0.12
Palpation data	0.69	0.15	0.36	0.14

Table 3: Classification results with HOML.

Dataset	Average precision	Ranking loss	One error	Hamming loss
Four types of diagnostic fusion data	0.80	0.09	0.25	0.12

3.3. Characteristics of Complexion and Tongue Picture

3.3.1. Change Characteristics of Tongue Picture Objective Indicators. After extraction of characteristic value of tongue picture, respectively, we calculate sample for tongue body and each part of coating on the tongue on Lab average value. The facial overall is the average value for each part and we select smooth quarter (3 months). Using them, we map the time trends figure. Because of the original material, we deal

Table 4: Feature selection results by using HOML.

Dataset	Feature selection results
Four types of diagnostic fusion data	RGB_B_face, Lab_A_tongue_coating, HSV_S_face HSV_H_face, RGB_G_face, Lab_B_face, Lab_L_face RGB_R_face, HSV_V_face, Lab_B_tongue_coating

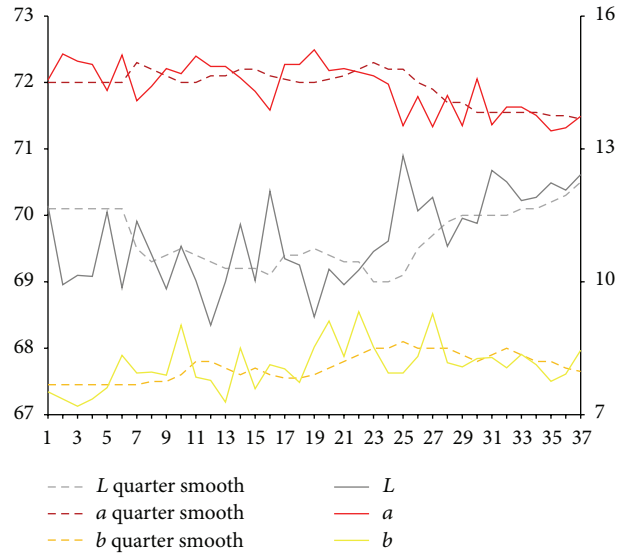


Figure 3: Lab value changing trend of tongue body.

with the lack of four tongue images using nearest neighbor interpolation method in which it is the average value of before and after neighbors.

Results are shown in Figures 3 and 4. From Figures 3 and 4, it can be seen that, compared with the initial state, the brightness L values of tongue body and coating on the tongue first show a trend of decrease and later show a substantial increase. In the late (beginning from 25th), the brightness L values increase significantly. Early and midterm body of the tongue show a weakly rising, and, in the late (beginning from 25th), a values show a slight reduction. In the late (beginning from 20th), b value of coating on the tongue continues to rise significantly.

3.3.2. Change Characteristics of Complexion Objective Indicators. After the extraction of characteristic value of complexion image, respectively, we calculate sample for each part of facial overall on Lab average value. The facial overall is the average value for each part and we select smooth quarter (3 months). Using them, we map the time trends figure. Because of the original material, we deal with the lack of three complexion images using nearest neighbor interpolation method in which it is the average value of before and after neighbors.

Lab value changing trend of facial overall is shown in Figure 5. From Figure 5, it can be seen that, compared with the initial state, the brightness L value of facial overall first shows a trend of continued increase and a value shows a trend of continued decrease.

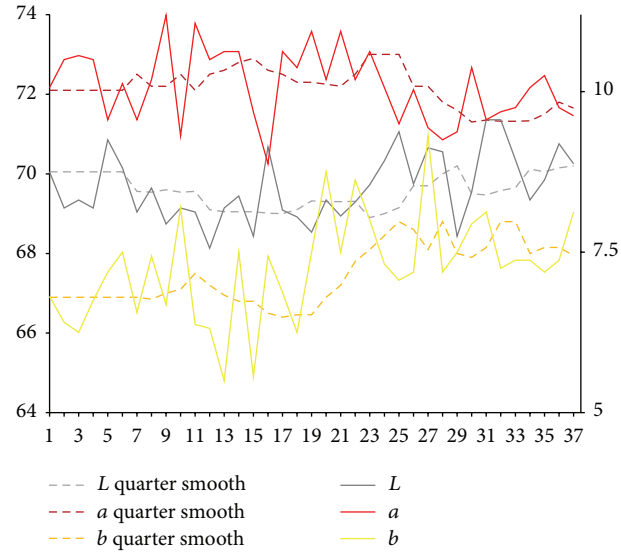


Figure 4: Lab value changing trend of coating on the tongue.

From Figures 6, 7, 8, 9, 10, and 11, it can be seen that compared with the initial state, the brightness L values of forehead, nose, left cheek, right cheek, and lip show a trend of increase, in which forehead and nose are more obvious. Brightness L value of underjaw shows a cyclical change. Values of forehead, nose, left cheek, right cheek, lip, and underjaw show a slight reduction in different degree, in which nose is the most obvious.

3.3.3. Analysis of Complexion and Tongue Picture Characteristics. From the above results, it can be seen that, after spaceman volunteers get into airtight cabin, changes of tongue picture and complexion are consistent with changes of syndromes which are shown in Figure 12. Brightness L values of tongue body and coating on the tongue firstly show a trend of decrease, then increase, and begin to increase significantly since the 50th week. At early and middle stage of getting into cabin, a value of tongue body increases weakly, and, since the 50th week, a value tends to slightly decrease. Since the 40th week of getting into cabin, b value of coating on the tongue continues to increase significantly. Compared with the early stage of getting into cabin, brightness L value of facial overall continues to increase, and a value continues to decrease. Brightness L values of forehead, nose, left cheek, right cheek, and lip show an increasing trend, among which forehead and nose are more obvious and lip increases more obvious at late stage (since 50th). Brightness L value of underjaw changes periodically. Values of forehead, nose, left cheek, right cheek, and lip show decreasing trend in different degrees, and nose is most obvious, which indicates brightness L values and a values of tongue picture; complexion of spaceman volunteers has different degrees of volatility; prompting volunteers are in the adaptation state, and all syndromes are existent. After a period of time of getting into cabin especially since the 50th week, brightness L values of tongue picture and complexion gradually increase, while a values gradually decrease, which presents deficiency syndrome, and it is consistent with the

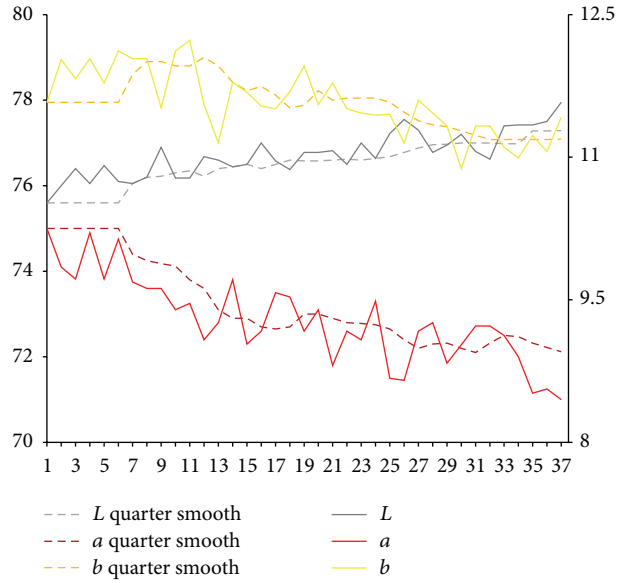


Figure 5: Lab value changing trend of facial overall.

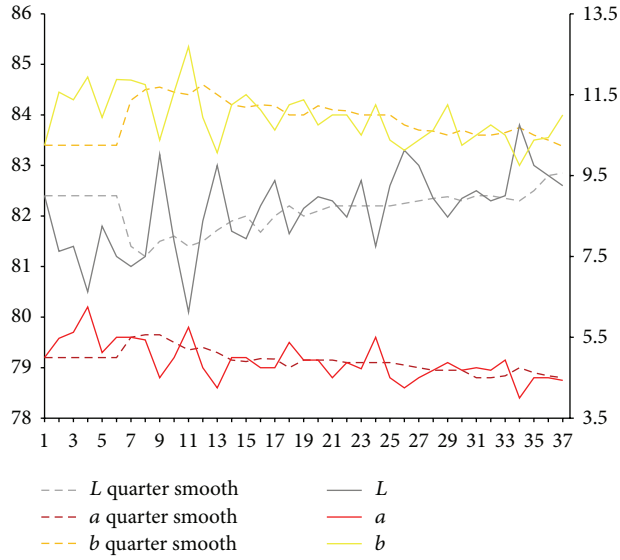


Figure 6: Lab value changing trend of forehead.

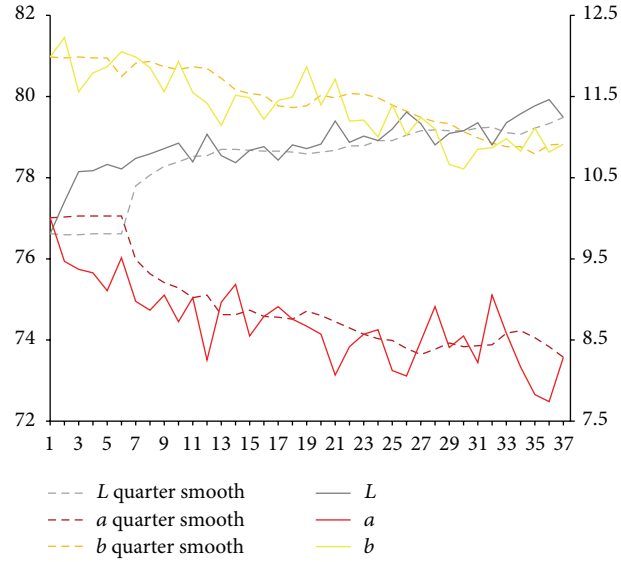


Figure 7: Lab value changing trend of nose.

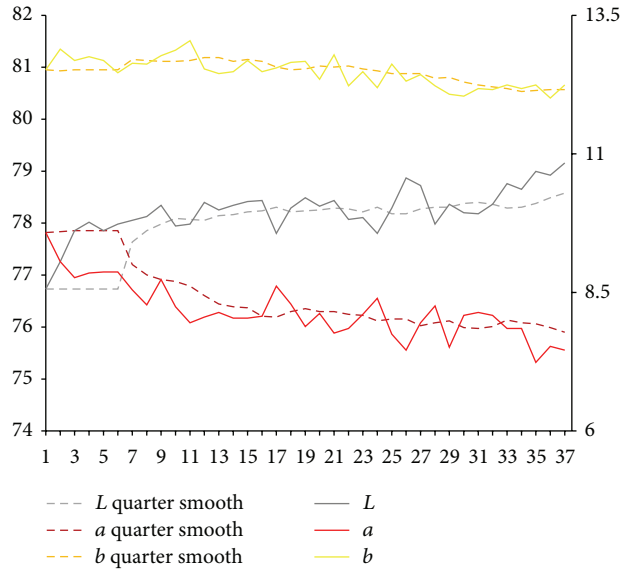


Figure 8: Lab value changing trend of left cheek.

changes of syndromes. Besides, in the isolation environment, body cannot be exposed to sunlight for long time, and there will be changes in color values of complexion, manifested as increasing brightness and decreasing red luminosity.

3.4. Analysis of Feature Selection Results of Our Model. In the mining process, we found that complexion was obvious in feature selection among the single index of complexion, pulse, and tongue. One reason is that changes of deficiency of qi and blood are before the change of pulse. Because of deficiency of qi and blood, complexion is easy to show up. The contractility, resistance, and tonicity are no response from pulse, because the deficiency of vital energy is weak and body is in good physical quality. There are no problems about cardiac systolic function and the appearance of peripheral resistance. On the characterization, mainly, color and pulse condition are not significantly affected. Red color is the reaction of tongue body and the thin white coating is the reaction of the coating on the tongue.

Qi deficiency is the main syndrome of spaceman volunteers in isolation environment. Tongue and complexion are the most sensitive among four diagnostic methods in qi deficiency.

Operation and system error also have a certain influence to the contribution of pulse. People can maybe do this, but machine has the certain difficulty.

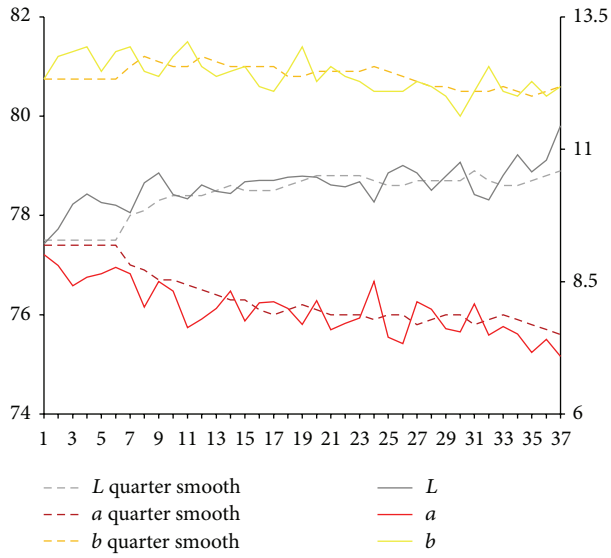


Figure 9: Lab value changing trend of right cheek.

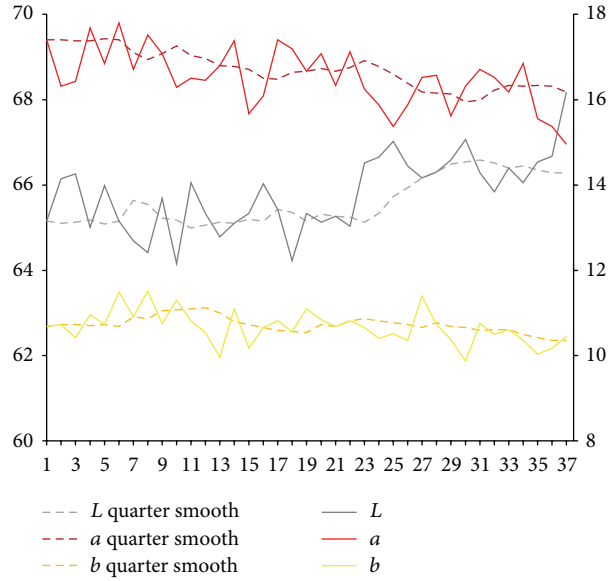


Figure 10: Lab value changing trend of lip.

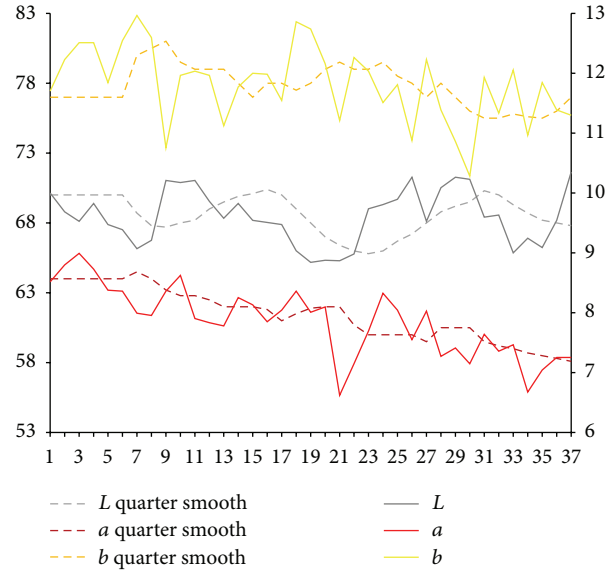


Figure 11: Lab value changing trend of underjaw.

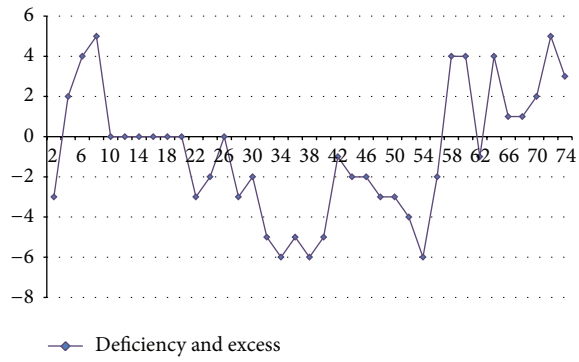


Figure 12: Variation of syndromes.

4. Conclusions

In this paper, statistics method is adopted to describe the syndrome factor regular pattern, finding that qi deficiency is a base syndrome pattern throughout the entire experiment process. While there are different associated symptoms such as liver depression, spleen deficiency, dampness stagnancy, and yin deficiency, due to differences of individual situation, machine learning methods are applied to mine the relationship between symptoms and syndromes. In our work, HOML is used to selected related symptoms and ML-KNN is used as the multilabel classifier. Compared with the model without HOML, the model with HOML obtains better performance. Through feature selection, ten key symptoms are selected for syndrome differentiation. Then, we give a detailed discussion for the feature selection results. At the same time, the average precision of multilabel classification model reaches 80%.

In this research, our syndrome differentiation results reveal base syndrome features and evolvement rule for human body in longtime isolation environment, which lays the foundation for further research. In the next work, we would do much research on how to improve the classification accuracy and, with higher classification accuracy, the multilabel classification model can aid decision making for syndrome differentiation.

Conflict of Interests

The authors declare that they have no competing interests.

Author's Contribution

All authors read and approved the final paper. Yong-Zhi Li and Kai-Xian Chen conceived the project and gave valuable advice to write this paper. Guo-Zheng Li, Zhi-Feng Zhang,

Jia-Tuo Xu, Sheng Sun, Feng-Feng Shao, Jia-Chang Chen, Yu-Meng Guo, and Li Ping Tu performed the computational experiments. Sheng Sun, Feng-Feng Shao, and Jia-Chang Chen wrote the paper; Guo-Zheng Li, Hong-Zhi Shi, and Zhi-Feng Zhang revised the paper. Yong-Zhi Li, Jian-Yi Gao, Quan-Chun Fan, Gui-E Bai, Hong-Zhi Shi, Yu Liu, Tao Mi, Xin-Hong Jia, Shuang Zhao, and Jun-Lian Liu explained the data and results from the view of Chinese medicine.

Acknowledgment

This work was supported by the Natural Science Foundation of China under Grants nos. 61273305 and 81173200.

References

[1] Y.-Z. Li, G.-Z. Li, J.-Y. Gao et al., "Mars500 analysis on tcm monitoring data of healthy status collected from mars500 volunteers," *Computer Engineering*, vol. 40, no. 9, pp. 13–18, 2014.

[2] X. Wang and G.-Z. Li, "Multilabel learning via random label selection for protein subcellular multilocations prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 2, pp. 436–446, 2013.

[3] X. Wang, G.-Z. Li, J.-M. Liu, and R.-W. Zhao, "Multi-label learning for protein subcellular location prediction," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM '11)*, pp. 282–285, IEEE, November 2011.

[4] G.-Z. Li, S.-X. Yan, M. You, S. Sun, and A. Ou, "Intelligent zheng classification of hypertension depending on ML-kNN and information fusion," *Evidence-Based Complementary and Alternative Medicine*, vol. 2012, Article ID 837245, 5 pages, 2012.

[5] X. Wang and G. Z. Li, "A multi-label predictor for identifying the subcellular locations of singleplex and multiplex eukaryotic proteins," *PLoS ONE*, vol. 7, no. 5, Article ID e36317, 2012.

[6] 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.

[7] H. Shao, G.-Z. Li, G.-P. Liu, and Y.-Q. Wang, "Symptom selection for multi-label data of inquiry diagnosis in traditional Chinese medicine," *Science China Information Sciences*, vol. 56, no. 5, pp. 1–13, 2013.

[8] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[9] R. Meiri and J. Zahavi, "Using simulated annealing to optimize the feature selection problem in marketing applications," *European Journal of Operational Research*, vol. 171, no. 3, pp. 842–858, 2006.

[10] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature Extraction, Construction and Selection*, pp. 117–136, 1998.

[11] J. Dréo and P. Siarry, "Continuous interacting ant colony algorithm based on dense heterarchy," *Future Generation Computer Systems*, vol. 20, no. 5, pp. 841–856, 2004.

[12] D. W. Aha, "Editorial," *Artificial Intelligence Review*, vol. 11, no. 1–5, pp. 7–10, 1997.

[13] G.-P. Liu, G.-Z. Li, Y.-L. Wang, and Y.-Q. Wang, "Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning," *BMC Complementary and Alternative Medicine*, vol. 10, no. 1, article 37, 2010.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 189357, 6 pages
http://dx.doi.org/10.1155/2015/189357

Research Article

Researches on Mathematical Relationship of Five Elements of Containing Notes and Fibonacci Sequence Modulo 5

Zhaoxue Chen

School of Medical Instrument & Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Correspondence should be addressed to Zhaoxue Chen; chenzhaoxue@163.com

Received 27 September 2014; Revised 8 December 2014; Accepted 10 December 2014

Academic Editor: Zhaohui Liang

Considering the five periods and six qi's theory in TCM almost shares a common basis of stem-branch system with the five elements of containing notes, studying the principle or mathematical structure behind the five elements of containing notes can surely bring a novel view for the five periods and six qi's researches. By analyzing typical mathematical rules included in He tu, Luo shu, and stem-branch theory in TCM as well as the Fibonacci sequence especially widely existent in the biological world, novel researches are performed on mathematical relationship between the five elements of containing notes and the Fibonacci sequence modulo 5. Enlightened by elementary Yin or Yang number grouping principle of He tu, Luo shu, the 12534 and 31542 key number series of Fibonacci sequence modulo 5 are obtained. And three new arrangements about the five elements of containing notes are then introduced, which have shown close relationship with the two obtained key subsequences of the Fibonacci sequence modulo 5. The novel discovery is quite helpful to recover the scientific secret of the five periods and six qi's theory in TCM as well as that of whole traditional Chinese culture system, but more data is needed to elucidate the TCM theory further.

1. Introduction

Combining the 10 celestial stems and 12 earth branches, the Chinese sexagenary cycle can be formed, which is widely used in Chinese ancient culture as the most important means of counting time of years, months, or dates. In TCM (Traditional Chinese Medicine), given that each year is associated with a different combination of stems and branches over a cycle of sixty years, all possible climatic constellations of the year may sequentially occur in a specific way, which can be characterized by the stem-branch combination closely related with the five periods and six qi. In other words, the doctrine of the five periods and six qi explains relationships ancient Chinese observers assumed to exist between climate and a broad range of natural phenomena, including human health and illness. And the concepts of the five periods and of the six qi were introduced to distinguish among and specify climatic characteristics of well-defined time periods. By drawing on notions of a cyclical recurrence of calendric terms and by adopting the doctrines of yin yang and of the five elements, an attempt was made to order what may at first glance appear to be disorder, namely, the occurrence of rain and wind, dryness, cold, and heat in the course of the four seasons and over the years. Knowledge of a distinct regularity uncovered in frequent climatic changes not only permitted an understanding of the generation, growth, maturity, and death of numerous phenomena in nature in general but also, more importantly, enabled man to integrate himself into eternal laws governing all existence. The doctrine of the five periods and six qi is outlined in the Su wen in seven "comprehensive discourses," which comprises about one-third of the entire text of the Su wen. But the origin of the notions is unclear and no parallel literary sources outside the Su wen are known that could be used to date the early development of these thoughts [1]. All the same, there are also some important clues in other fields of ancient Chinese culture which can be used as a reference for researchers of the five periods and six qi in TCM. Therein, theory of the five elements of containing notes originated at least before Qin Dynasty is the most notable one, where each combination of a stem and a branch is also vital for description of yin yang and five elemental characteristics and each attributed to a corresponding element named as the so-called

five elements of containing notes [2, 3]. The na yin wu xing, that is, the element representing the stem-branch of one's birth-year, is often used to judge one's fate by folk fate calculators up to nowadays and in the Book of the Master Who Embraces Simplicity by Ge Hong (who is a famous Taoist priest in Jin Dynasty) it is recorded that the fate corresponding to the five elements of containing notes decided by the stem-branch combination corresponding to one's birth year can be used as a guidance of color selecting of the medicine to be taken; Ge Hong says in volume number 11 of his inner book of the Master Who Embraces Simplicity: "...According to the book of Yu ce ji and the book of Kai ming jing,..., if one's fate is soil, he is not fit to take medicine with cyan color; medicines with red color are not fit for persons of metal fate; white color is not fit for wood fate; yellow color is not fit for water fate and black color is not fit for fire fate. That is exactly because, according to meaning of the five elements, the wood restricts the soil, the soil restricts the water, the water restricts the fire, the fire restricts the metal and the metal restrict the wood...". Although some of Ge Hong's ideas are contentious with a mysterious tendency of a common Taoist, it is true that his Handbook of Prescriptions for Emergencies inspired the modern discovery of artemisinin [4]. No matter in theory of the TCM five periods and six qi or in theory of the five elements of containing notes, the same 60 cyclical stem-branch combinations by years are paid much attention and related judgments or doctrines are formed based on the stem and branch related theory of yin yang and five elements. Considering the five periods and six qi's theory of TCM almost shares a common basis of stem-branch system [1, 2], studying the principle or mathematical structure behind the five elements of containing notes can surely bring a novel view for the five periods and six qi's related researches in TCM.

The Fibonacci sequence, that is, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233,..., is a famous series universally used in various modern disciplines such as computer science, optimizing theory, biological mathematics and physics, number theory and combinatorics, and material science [5–9]. Even there is a special formal periodical (namely, The Fibonacci Quarterly) dedicated to Fibonacci sequence related topics and researches [10]. In particular, having close relationship with the Golden Section Number, it is an interesting mathematical sequence that widely exists in the biological world. For example, there is a peculiar pattern in the flower petals of nearly all the flowers; the number of their petals is one number of the Fibonacci sequence. Small flowers of a sunflower, the heart of chrysanthemum, squama on surface of pinecones, and tumors-like structure of pineapple all have shown similar two near-perfect spirals in two opposite spiral directions, respectively. And the ratio of two spiral numbers has close relationship with the Fibonacci sequence. In the pinecones the ratio is 5 : 8, the pineapple is 8 : 13, Marguerite daisy is 21 : 34, and the sunflower is 34 : 55,...; the series of number couples are all from two adjacent numbers of the Fibonacci series exactly. Any face plate of the Asteraceae family has the same characteristics with the sunflower. In animal cells, hollow cores of microtubules constituted by protein polymer, which form the cell cytoskeleton, help to maintain a certain shape and act as "nervous system" of cells. Typical mammalian cell microtubule is constituted by 13 original fibers, of which 5 are dextrorotation fibers and 8 are laevorotation ones (herein, 5, 8, and 13 are all adjacent Fibonacci numbers). Moreover, people have occasionally found a double-microtubule with an outer layer, and it is constituted by 21 original fibers, which happens to be the next number in Fibonacci series. Moreover, it is well known that molecules of the B-DNA expose a double-helix geometric structure, and the helix length of the double-helix DNA structure is 34 angstroms and its radius is 21 angstroms. 34 and 21 happen to be the two adjacent numbers in the Fibonacci sequence [6].

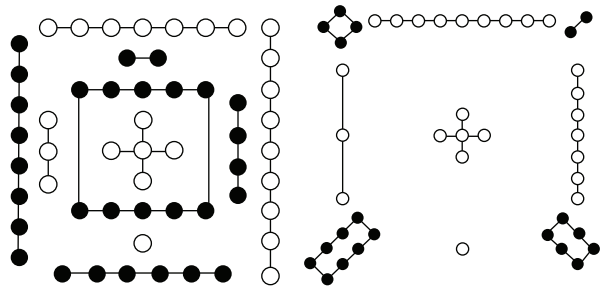
Moreover, the Fibonacci sequence is also found to have close relationship with traditional Chinese culture. Elizabeth Moran and her coauthors have discovered that four of feng shui's (feng shui is another mysterious system in traditional Chinese culture system with almost the same theory basis as TCM) fundamental principles correspond to numbers in the Fibonacci sequence: Taiji (1), yin and yang (2), heaven, earth, and human qi (3), five phases (5), and eight trigrams (8) [11]. And according to some other researchers, the Fibonacci numbers also have close relationship with the most famous two key diagrams of He tu and Luo shu in ancient Chinese culture [12, 13].

In this paper, according to some typical mathematical rules included in He tu, Luo shu in TCM, novel associations are discovered between the five elements of containing notes and the Fibonacci modular sequence, which can set up a mathematical bridge between the five elements of containing notes widely existent in Chinese traditional culture and abundant modern Fibonacci series related researches. The novel discovery is of great value to reduce the mysterious sense of ancient Chinese culture and is quite helpful to uncover the final scientific secret of the five periods and the six qi's theory in TCM.

2. Materials and Methods

2.1. Obtaining of Two Key Subsequences of Fibonacci Sequence Modulo 5. The He tu and Luo shu are two of the most famous diagrams in Chinese traditional culture as well as in TCM system. As shown in Figure 1 [1, 11], according to legend, both He tu and Luo shu initially emerged as groups of black and white dots, where each group of black dots has even dot number and the other groups composed of white dots have odd number of dots, respectively. In particular, either He tu or Luo shu has shown the same proneness to emphasize group number of 5 and the group of dots with number 5 is exactly positioned in the center of He tu and Luo shu which are arranged as the shape of a cross. Further according to the Hong fan section in Shang shu, number one is water; number two is fire; number three is wood; number four is metal; number five is soil [1]. The number of different groups of dots in He tu or Luo shu can thus form a direct association with the five elements' theory in TCM or the five elements of containing notes.

The Fibonacci sequence has a recycled property while performing modular operation by an integer n. While setting the value of n be 5, that is, the center number of He tu or Luo



(a) The He tu diagram (b) The Luo shu diagram

Figure 1: The He tu and Luo shu diagram.

shu, the period is equal to 20 and the corresponding recycled sequence can be listed as below [14]:

1 1 2 3 0 3 3 1 4 0 4 4 3 2 0 2 2 4 1 0

Suppose each index value of numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 in He tu or Luo shu is equal to each number itself; then no matter He tu and Luo shu or 10 stems and 12 branches in TCM can all be partitioned into two groups of Yin and Yang based on the parity of index values. In other words, for 10 stems and 12 branches, those with odd index numbers are attributed to the Yang group and the others with even numbers are grouped as Yin, which just corresponds to the white or black dots group in He tu and Luo shu, respectively. Similarly, according to parities of the index of each number the obtained 20 Fibonacci recycled sequence numbers by modulo 5 above can also be divided into two groups: subsequence Yang and subsequence Yin. The grouping result is listed as below:

subsequence Yang: 1 2 0 3 4 4 3 0 2 1,

subsequence Yin: (1 3) 3 1 0 4 2 2 4 0 1 3,

for analyzing convenience; all the numbers in the two subsequences are rearranged into two 5-tuples and all numbers in subsequence Yin are circularly right-shifted by 2 positions. Apparently the obtained subsequences of Yang and Yin are just composed of series of 1 2 0 3 4 and 3 1 0 4 2 together with their corresponding palindrome, that is, two subsequences both taking 1 2 0 3 4 and 3 1 0 4 2 as the key sequence, respectively. In mathematics, the number 0 is just equal to 5 in the viewpoint of operation modulo 5; therefore, the key number series of 1 2 0 3 4 and 3 1 0 4 2 can also be seen as 1 2 5 3 4 and 3 1 5 4 2.

2.2. New Discovery on Arrangement of the Five Elements of Containing Notes. Based on traditional Chinese culture, each of the containing notes associates a unique symbol as well as an attribute and Hong fan number of five elements with a stem-branch combination and two neighbored stem-branch combination pairs share the same symbol and attribute [2, 3]. The naming of the associated symbols, deriving, and relationship between such association and ancient Chinese music theory are beyond the scope of this paper; those who are interested in it can refer to related literatures directly. All stem-branch combinations and each corresponding symbol as well as the associated attribute and Hong fan number of corresponding five elements are all summarized as in Table 1 based on description in [2, 3].

Actually all the stem-branch combinations are just the permutation and combination between Yang group of stems and branches and Yin group of them. And the thorough building process of associations shown in Table 1 is still a mystery. While studying the rules behind the stem-branch combinations, the author found a new arrangement of the stem-branch combinations based on a simple and determinate transforming operation from the original arranging order in traditional Chinese culture which is demonstrated in Table 1 by the "number" column. The discovery of this new arrangement can help to provide research clues of mathematical structure of the five elements of containing notes. That is, take the JiaXu YiHai pair (11th and 12th stem-branch combinations in Table 1) as the initial pair and take equation (m + 22) mod 60 (herein "mod" represents modular operation) as a recursive rule to obtain the successive one of all recycled pairs; for example, if substituted with m = 11 and 12, the index of the second recycled pair can be obtained; we have

(11 + 22) mod 60 = 33

(12 + 22) mod 60 = 34;

then look up the indexes 33 and 34 in Table 1; the second pair BingShen DingYou can be searched out. Similarly, the index of the third pair can be computed as (33 + 22) mod 60 = 55, (34 + 22) mod 60 = 56,. . ., and finally, the whole periodical new arrangement can all be obtained. And the 5 by 6 array form of the new arrangement is shown in Table 2.

For comparison convenience, substitute each stem-branch pair in the new arrangement shown in Table 2 by its corresponding Hong fan number according to Table 1; it produces a 5 by 6 number array:

2	2	2	4	4	4
5	5	5	1	1	1
3	3	3	2	2	2
4	4	4	5	5	5
1	1	1	3	3	3

Looking up from up to bottom, it is obvious that each column of the number array is one of the recycle-shifting sequence of 12534 just by one position in the left or right

Table 1: Associated stem-branch combinations, symbols, and the attributes and Hong fan numbers of five elements.

Sequence number	Stem-branch combination	Symbol	Attribute and Hong fan number
1	JiaZi		Metal
2	YiChou	The metal in the sea	4
3	BingYin		Fire
4	DingMao	The fire in the stove	2
5	WuChen		Wood
6	JiSi	The wood of a great forest	3
7	GengWu		Soil
8	XinWei	The wayside soil	5
9	RenShen		Metal
10	GuiYou	The metal on the sword-blade	4
11	JiaXu		Fire
12	YiHai	The fire on the hill-top	2
13	BingZi		Water
14	DingChou	The brook water	1
15	WuYin		Soil
16	JiMao	The soil on the city-wall	5
17	GengChen		Metal
18	XinSi	The metal on the white candle	4
19	RenWu		Wood
20	GuiWei	The willow wood	3
21	JiaShen		Water
22	YiYou	The spring water in a well	1
23	BingXu		Soil
24	DingHai	The soil on the roof of a house	5
25	WuZi		Fire
26	JiChou	the thundering fire	2
27	GengYin		Wood
28	XinMao	The wood of the pine or cypress	3
29	RenChen		Water
30	GuiSi	The flowing water	1
31	JiaWu		Metal
32	YiWei	The metal in the sand	4
33	BingShen		Fire
34	DingYou	The fire at the foot of a hill	2
35	WuXu		Wood
36	JiHai	The wood on a plain	3
37	GengZi		Soil
38	XinChou	The soil on the roof of a house	5
39	RenYin		Metal
40	GuiMao	The metal on the paper money	4
41	JiaChen		Fire
42	YiSi	The fire of a lamp under cover	2
43	BingWu		Water
44	DingWei	The water of the heavenly river	1
45	WuShen		Soil
46	JiYou	The soil of the highway station	5
47	GengXu		Metal
48	XinHai	The gold of the hairpin	4
49	RenZi		Wood
50	GuiChou	The wood of the mulberry tree	3

Table 1: Continued.

Sequence number	Stem-branch combination	Symbol	Attribute and Hong fan number
51	JiaYin		Water
52	YiMao	The water of a great stream	1
53	BingChen		Soil
54	DingSi	The soil in the sands	5
55	WuWu		Fire
56	JiWei	The heavenly fire	2
57	GengShen		Wood
58	XinYou	The wood of the pomegranate	3
59	RenXu		Water
60	GuiHai	The water of the sea	1

Table 2: The 5 by 6 array form of the new arrangement.

JiaXu YiHai	BingShen DingYou	WuWu JiWei	GengChen XinSi	RenYin GuiMao	JiaZi YiChou
BingXu DingHai	WuShen JiYou	GengWu XinWei	RenChen GuiSi	JiaYin YiMao	BingZi DingChou
WuXu JiHai	GengShen XinYou	RenWu GuiWei	JiaChen YiSi	BingYin DingMao	WuZi JiChou
GengXu XinHai	RenShen GuiYou	JiaWu YiWei	BingChen DingSi	WuYin JiMao	GengZi XinChou
RenXu GuiHai	JiaShen YiYou	BingWu DingWei	WuChen JiSi	GengYin XinMao	RenZi GuiChou

Table 3: New arrangement based on the 12 branches.

JiaZi YiChou4	BingYin DingMao2	WuChen JiSi3	GengWu XinWei5	RenShen GuiYou4	JiaXu YiHai2
BingZi DingChou1	WuYin JiMao5	GengChen XinSi4	RenWu GuiWei3	JiaShen YiYou1	BingXu DingHai5
WuZi JiChou2	GengYin XinMao3	RenChen GuiSi1	JiaWu YiWei4	BingShen DingYou2	WuXu JiHai3
GengZi XinChou5	RenYin GuiMao4	JiaChen YiSi2	BingWu DingWei1	WuShen JiYou5	GengXu XinHai4
RenZi GuiChou3	JiaYin YiMao1	BingChen DingSi5	WuWu JiWei2	GengShen XinYou3	RenXu GuiHai1

direction, that is, 25341 or 41253, and looking from the end there are exactly three appositions of 31542 31542.

Above discovery is not only a special case. Stem-branch combinations and their corresponding Hong fan number given in Table 1 can also derive two other arrangements as listed in Tables 3 and 4 based on the 12 branches or base number 8 after extending one duplicate of whole 60 elements.

Obviously, in each column of Table 3 the Hong fan numbers are the recycle-shifting sequence of 12534 by a certain number. And in each column of Table 4, by recycling view, the Hong fan numbers all show a similar appositional pattern of three same recycle-shifting sequences of 31542.

3. Results, Discussions, and Conclusions

In this paper, enlightened by elementary Yin or Yang number grouping principle of He tu, Luo shu, and stem-branch theory, the 12534 and 31542 key number series of Fibonacci sequence modulo 5 (the center number of He tu or Luo shu) are obtained. And by simple and determinate transforming operation of original order of the stem-branch combinations widely existent in traditional Chinese culture, a new 5 by 6 arrangement about the five elements of containing notes is firstly introduced, which has shown close relationship with the two obtained key subsequences of Fibonacci sequence modulo 5. As the new derived arrangement is directly acquired from the original indices of stem-branch combinations by fixed mathematical operation, the rule implied in the new arrangement also reflects that of the original arrangement of stems and branches in traditional Chinese culture. Another two derived arrangements about the five elements of containing notes also show close relationship with sequence 12534 or 31542 respectively; therefore it can be deduced that the five elements of containing notes must have underlying profound mathematical relationship with the Fibonacci modular sequence. Actually, besides applications in fate judgments or in ancient Taoist system, the five elements of containing notes associated with 60 stem-branch combinations have also close relationship with traditional Chinese musical system [3]. And for ancient Chinese people, it is very common and natural to describe the musical and calendric laws in the same form based on stems and branches. In their opinions, the climate throughout a whole year is closely related with certain musical notes denoted by a stem-branch combination corresponding to the year as the five periods and six qi in TCM do. Moreover, the Fibonacci sequence widely exists in various modern disciplines especially in the biological world, and as a special traditional medicine system, herbs from the biological world are widely used in TCM. Therefore, although researches in this paper are mainly performed based on arrangements of the five elements of containing notes, considering it almost shares a

Table 4: New arrangement based on base number 8 after extending one duplicate of 60 elements.

Jiazi YiChou4	BingYin DingMao2	WuChen JiSi3	GengWu XinWei5
RenShen GuiYou4	Jiaxu YiHai2	BingZi DingChou1	WuYin JiMao5
GengChen XinSi4	RenWu GuiWei3	JiaShen YiYou1	BingXu DingHai5
WuZi JiChou2	GengYin XinMao3	RenChen GuiSi1	JiaWu YiWei4
BingShen DingYou2	WuXu JiHai3	GengZi XinChou5	RenYin GuiMao4
JiaChen YiSi2	BingWu DingWei1	WuShen JiYou5	GengXu XinHai4
RenZi GuiChou3	JiaYin YiMao1	BingChen DingSi5	WuWu JiWei2
GengShen XinYou3	RenXu GuiHai1	JiaZi YiChou4	BingYin DingMao2
WuChen JiSi3	GengWu XinWei5	RenShen GuiYou4	JiaXu YiHai2
BingZi DingChou1	WuYin JiMao5	GengChen XinSi4	RenWu GuiWei3
JiaShen YiYou1	BingXu DingHai5	WuZi JiChou2	GengYin XinMao3
RenChen GuiSi1	JiaWu YiWei4	BingShen DingYou2	WuXu JiHai3
GengZi XinChou5	RenYin GuiMao4	JiaChen YiSi2	BingWu DingWei1
WuShen JiYou5	GengXu XinHai4	RenZi GuiChou3	JiaYin YiMao1
BingChen DingSi5	WuWu JiWei2	GengShen XinYou3	RenXu GuiHai1

common basis of stem-branch system with the theory of the five periods and six qi in TCM, the mathematical relationship of five elements of containing notes and Fibonacci sequence modulo 5 found in this paper is undoubtedly quite helpful to recover the scientific secret of the five periods and six qi's theory in TCM as well as that of whole traditional Chinese culture system, but more data is still needed to elucidate the TCM related theory in future researches.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Authors' Contributions

Zhaoxue Chen carried out all the studies of this paper and drafted the paper, as well as reading and approving the final paper.

Acknowledgments

This work is under the auspice of Shanghai Municipal Education Commission to Scientific Innovation Research Funds (no. 11YZ116) and BME first-class discipline (B) building project for Shanghai universities.

References

[1] P. Unschuld, *Huang Di Nei Jing Su Wen: Nature, Knowledge, Imagery, in an Ancient Chinese Medical Text*, University of California Press, Berkeley, Calif, USA, 2003.

[2] W.-P. Chao and Z. W. Bang, "The Chinese science of fate-calculation," *Folklore Studies*, vol. 5, pp. 279–315, 1946.

[3] H. Datong, "On '60 jiazi containing notes' study," *Studies in Culture & Art*, no. 4, pp. 64–98, 2009.

[4] Y. Tu, "The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine," *Nature Medicine*, vol. 17, no. 10, pp. 1217–1220, 2011.

[5] C. Zhou, *Fibonacci-Lucas Sequences and Their Applications*, Hunan Science and Technology Publishing House, Changsha, China, 1993.

[6] Z.-X. Chen, Z.-Z. Wang, Y. Sun, and F.-L. Bei, "Discussion on microcosmic derivation of biological golden section phenomena from DNA geometric structure and snow flower generation," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 3, no. 1, pp. 31–35, 2011.

[7] V. Weiss and H. Weiss, "The golden mean as clock cycle of brain waves," *Chaos, Solitons and Fractals*, vol. 18, no. 4, pp. 643–652, 2003.

[8] C. J. Glasby, S. P. Glasby, and F. Pleijel, "Worms by number," *Proceedings of the Royal Society B: Biological Sciences*, vol. 275, no. 1647, pp. 2071–2076, 2008.

[9] C. R. Li, X. N. Zhang, and Z. X. Cao, "Triangular and Fibonacci number patterns driven by stress on core/shell microstructures," *Science*, vol. 309, no. 5736, pp. 909–911, 2005.

[10] http://www.fq.math.ca/.

[11] E. Moran, M. V. Biktashev, V. Biktashev, and J. Yu, *The Complete Idiot's Guide to Feng Shui*, Pearson Education, 2nd edition, 2002.

[12] Z. Gao, *Quest on Secret of Jiugong Diagram*, Hongkong Pegasus Books Limited Company, Hong Kong, 2004.

[13] J. Geng, "Relationship between Loshu and the Fibonacci sequence," *Bulletin of Mathematics*, vol. 47, no. 5, pp. 46–49, 2008.

[14] M. Renault, *Properties of the Fibonacci sequence under various moduli [M. S. thesis]*, Wake Forest University, Winston-Salem, Calif, USA, 1996.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2015, Article ID 820436, 5 pages
http://dx.doi.org/10.1155/2015/820436

Research Article

Standardization of Syndrome Differentiation Defined by Traditional Chinese Medicine in Operative Breast Cancer: A Modified Delphi Study

Qianqian Guo and Qianjun Chen

Department of Mammary Disease, Guangdong Provincial Hospital of Chinese Medicine, Guangdong, China

Correspondence should be addressed to Qianjun Chen; cqj55@163.com

Received 29 July 2014; Revised 3 September 2014; Accepted 25 September 2014

Academic Editor: Zhaohui Liang

Objective. The aim of this study was to establish the standardization of syndrome differentiation of operative breast cancer treated with Traditional Chinese Medicine (TCM) by the modified Delphi method. *Method*. A literature search for standardization of syndrome differentiation of operative breast cancer was conducted and eligible articles were identified in indexed databases from 1982 to 2013. We carried out two rounds of investigation between March and October 2013 and organized 20 experts who focused on TCM or integrative medicine in breast cancer research. Experts' judgments were collected *via* posted questionnaires or e-mail. A final evaluation was carried out after the end of both rounds. *Result*. The response ratio of the 1st round investigation reached 100%, and two experts were excluded due to the uncompleted questionnaire. The 2nd round investigation was completed by 18 experts in the 1st round panel board. In both rounds, the experts agreed that the stage of breast cancer defined by TCM could be divided into the perioperation period, the perichemotherapy period, the periradiotherapy period, and the consolidation period. *Conclusion*. We identified the feasibility and reasonability to establish the standardization of syndrome differentiation of operative breast cancer. According to the suggestions from experts in our Delphi study, we preliminarily established the TCM standard of syndrome differentiation based on different treatment stages of operative breast cancer.

1. Introduction

Globally, breast cancer is the most frequently diagnosed cancer and the second cause of cancer death in women. According to the statistics from the World Health Organization in 2011, breast cancer has become a major factor leading to cancer death in women worldwide and accounted for 14% of global cancer deaths [1]. Meanwhile, based on the statistics from the American Cancer Society estimates, there would be 235,030 new cases diagnosed with breast cancer in 2014, and 40,430 cases might die [2]. Clinical investigation found that although breast cancer incidence ratio was significantly increased around the world in the past few years, the survival period was greatly improved due to the distinguished advancements made by the multidisciplinary treatment [3–6].

Long-term clinical practice showed that TCM could reduce the side effects of chemo-drugs and improve the patient's immunity as well. Regarding the same disease, there are varieties of syndromes related to different climates and geographical factors. Currently, the syndrome differentiation defined by TCM on breast cancer treatment is not unified, and the related description or records in the ancient books and literature are limited, resulting in a severe impact on the standardization of TCM treatment in breast cancer. Therefore, it is necessary to explore and establish standard criteria of syndrome differentiation of TCM on breast cancer. Based on the preliminary research of the TCM syndromes, we applied the modified Delphi method to obtain the consensus on the criteria of syndrome differentiation in breast cancer treatment.

The Delphi method is a consensus methodology attempting to assess the extent of agreement and to resolve disagreement in medical and health service research [7, 8]. Its main characteristics include anonymous feedback and statistics. Application of Delphi method in medical research

originated from the nursing work in 2000. Mitten-Lewis [9] organized 28 educationalists and 43 nursing experts to form a committee aiming at studying the skill and ability obstacles faced by nurses in Australia. Thereafter the Delphi method was gradually applied in various fields of medicine, such as therapeutic strategy evaluation of esophageal cancer, breast cancer, and colorectal cancer [10]. In the field of Chinese medicine, Delphi can be used in the research of the syndrome distribution, the standard for diagnosis disease, and so on. Here, our study aims to establish the standardization of syndrome differentiation of operative breast cancer treated with TCM by using the modified Delphi method.

2. Participants and Methods

2.1. Selection Criteria. Delphi methodology suggests that the number of advisory experts can range from ten to fifty. According to the principle of representation and authority, twenty nationwide experts who majored in TCM and integrative medicine were selected in the first round. Based on the first round results, eighteen experts in the first round were further consulted in the second round.

Twenty clinicians working in different regions were selected for evaluating the syndrome results based on their experiences in breast cancer treatment by TCM. In order to increase the credibility of our results, the factor of expert's geographical distribution was also considered. All participating experts were engaged in breast cancer research for at least ten years, and each expert had published at least ten research articles on breast cancer therapy. Their academic position was associate professor or above.

The leading group was formed by key members of the Chinese Medical Society of Breast Diseases Prevention and Control Cooperative Work Committee, who are responsible for the design of the questionnaire. These four professors were the most respected leaders in research and administrative authorities in China. The mean working experience of the leading group was more than 20 years.

2.2. Literature Review. A literature review was performed to define the initial list of symptoms of breast cancer. The keywords were "breast cancer in TCM" and "different research of breast cancer in TCM." The related articles published between 1982 and 2013 from two databases were selected. Case reports and articles based on subjective experience and treatment were excluded.

2.3. The Delphi Procedure. To obtain a consensus among experts, a Delphi process was used. The method is a way to obtain expert opinion in a systematic manner. Experts are recruited individually and anonymously. The survey is conducted over several rounds, and the results are analyzed and finally reported to the leading group. The process is considered completed when there is a convergence of opinion or when a point of diminishing returns is reached.

The protocol and consensus assessment used in the Delphi study were based on the structured integration of empirical data and experts' judgments [8]. The Delphi model could systematically evaluate the final agreement based on the individual opinion. According to the literature search, a presurvey questionnaire was suggested to be designed to ensure the accuracy and relevance, and the second round questionnaire was generated based on the results from the first round. We carried out two rounds of investigation between March and October 2013.

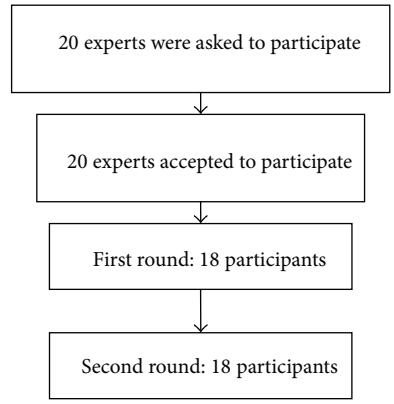


Figure 1: Flow of participants through the Delphi study.

2.4. Data Collection and Analysis. In each round, experts' judgments were collected *via* post or e-mail using the designed questionnaire. Meanwhile, the questions in both rounds were the same and the feedback of the first round results was also released to the experts during the second round investigation. Each questionnaire included two parts: one part described the experts' opinion on the classification of treatment phase and syndrome differentiation and the other part focused on the definition of each therapeutic phase (see Table 3). The level of agreement of each question was defined as "yes" or "no." Participants were expected to give their opinion on each question.

The quantitative analysis was carried out by calculating the supportive ratio for each question. It was considered that the experts reached a consensus on a question whenever the supportive ratio exceeds 50%.

3. Results

3.1. First Round. In the first round, the questionnaires were mailed to twenty breast experts and with a positive coefficient and callback rate of 100%, respectively. However, because two questionnaires were unqualified, the effective rate only reached 80%. All experts are associate professors or professors with work experiences over ten years. Eleven experts were with academic experiences more than twenty years and took up 55% of the whole experts (Figure 1), suggesting that the experts have high representation and authority in this study.

18 experts in the first round reached an agreement that the TCM treatment of the operative breast cancer could be divided into four phases including perioperative period, perichemotherapy period, periradiotherapy period, and consolidation period. With regard to syndrome differentiation, only 10 experts agreed that the "sick undocumented

Table 1: The experts' opinion on the classification of treatment phase and syndrome differentiation in the first round.

Stage	Syndrome manifestations	Number	Percent
The operation period			
Preoperation	Stagnation of liver-qi and phlegm	18	1
	Phlegm and blood stasis syndrome	18	1
	Chongren imbalance	18	1
	Deficiency of healthy energy and sthenia of evil	18	1
	Sick undocumented type	10	0.56
Postoperation	Spleen disharmony syndrome	18	1
	Deficiency of vital energy and blood pattern	18	1
	Qi-Yin deficiency type	16	0.89
Chemotherapy period	Spleen disharmony syndrome	18	1
	Deficiency of vital energy and blood pattern	16	0.89
	Qi-Yin deficiency type	15	0.83
	Deficiency of the liver and kidney	15	0.83
	Deficiency of the spleen and kidney	17	0.94
Radiotherapy period	Deficiency of vital energy and blood pattern	14	0.78
	Qi-Yin deficiency type	17	0.94
	Yin-chun deficiency	17	0.94
	Deficient yin induces vigorous fire	13	0.72
Consolidation period	Deficiency of vital energy and blood pattern	17	0.94
	Qi-Yin deficiency type	17	0.94
	Deficiency of the spleen and kidney	17	0.94
	Chongren imbalance	14	0.78

type" deems necessary in the perioperative period (55.56%). Besides this, the approval ratio for other syndrome differentiations all reached over 70% (Table 1).

For the definition of each therapeutic phase, the results from the first round investigation were listed as follows: the perioperation refers to the period from surgical admission to the beginning of the first cycle of chemotherapy; the perichemotherapy period is from the start of chemotherapy until one week later of the last cycle of chemotherapy; the periradiotherapy period is from the beginning of the radiotherapy to one week later of the endpoint; finally, the consolidation period refers to the five years after the end of radiotherapy and/or chemotherapy. During the presurvey, all experts reached an agreement about the time length definition on the perioperation and consolidation period, but not for the perichemotherapy and periradiotherapy period. There were three different opinions about the time of the perichemotherapy and the periradiotherapy period. The first opinion considered that the peri-chemotherapy/radiotherapy period refers to the start of chemotherapy/radiotherapy until one week later after the endpoint; the second one claimed that the peri-chemotherapy/radiotherapy period refers to the start of chemotherapy/radiotherapy until two weeks later after the endpoint, and the last one recommended that the periods should be from the start of chemotherapy/radiotherapy until one month later after the endpoint (Table 2).

3.2. Second Round. In the second round investigation, eighteen questionnaires were all returned with a recovery ratio of 100%. According to the first round conclusions, breast cancer

Table 2: The experts' opinions on the time length of perichemotherapy/radiotherapy in the first round investigation.

Proposal	Opinion
Extended to two weeks	55.56%
Extended to one week	27.78%
Extended to one month	16.67%

TCM treatment was divided into the perioperation period, the perichemotherapy period, the periradiotherapy period, and the consolidation period. In addition to the fact that the "sick undocumented type" was revealed with only 55.56% approval ratio, the rest of other syndrome differentiations were approved with more than 80.00% support from experts (Table 2). Therefore, we finally decided to kick out the syndrome "sick undocumented type" from the diagnostic criteria.

With regard to the time length of each TCM therapeutic phase, the second round results were listed in Table 4. Based on the results, we finally decided to determine the period of perichemotherapy and periradiation therapy lasting to two weeks after the end of treatment.

4. Discussion

In the past two decades, Delphi study has not only become a widely applied prediction model but also emerged as an important means of assessment and decision-making strategy [11, 12]. Since the method is based on scientific data collection

Table 3: The experts' opinion on the classification of treatment phase and syndrome differentiation in the second round.

Stage	Syndrome manifestations	Number	Percent
The operation period			
Preoperation	Stagnation of liver-qi and phlegm	7	1
	Phlegm and blood stasis syndrome	18	1
	Chongren imbalance	18	1
	Deficiency of healthy energy and sthenia of evil	17	0.94
	Sick undocumented type	10	0.56
Postoperation	Spleen disharmony syndrome	17	0.94
	Deficiency of vital energy and blood pattern	18	1
	Qi-Yin deficiency type	15	0.83
	Spleen disharmony syndrome	18	1
Chemotherapy period	Deficiency of vital energy and blood pattern	17	0.94
	Qi-Yin deficiency type	16	0.89
	Deficiency of the liver and kidney	17	0.94
	Deficiency of the spleen and kidney	18	1
Radiotherapy period	Deficiency of vital energy and blood pattern	15	0.83
	Qi-Yin deficiency type	18	1
	Yin-chun deficiency	18	1
	Deficient yin induces vigorous fire	15	0.83
Consolidation period	Deficiency of vital energy and blood pattern	18	1
	Qi-Yin deficiency type	18	1
	Deficiency of the spleen and kidney	18	1
	Chongren imbalance	15	0.83

Table 4: The experts' opinions on the time length of perichemotherapy/radiotherapy in the second round investigation.

Proposal	Opinion
Extended to two weeks	72.22%
Extended to one week	22.22%
Extended to one month	5.56%

and analysis, it can be used as a powerful tool to evaluate the clinical experience of TCM experts.

The advantages of Delphi depend on the open questionnaire, which can bring brainstorm and mutual discussion, while its disadvantages include the relatively long time, curt feedback, and the timeliness. The classic Delphi method was generally divided into four rounds, but the modified Delphi method cannot include the consultation round. As long as the opinion of experts reached a consensus, the investigation can be terminated and therefore avoid the long time for consultation. During the consultation process, it is better not to adopt "a piece of paper law" in the first round in order to refrain from disperse answers. In this study, according to the modified Delphi method, a predesigned questionnaire was designed based on the preliminary literature search results, which is equivalent to the completion of the first round consultation and effectively shortens the consultation period. Meanwhile, the questionnaire design should have as much detail as possible and leave enough space for expert review.

On the other hand, the selection of experts is also a critical factor determining the final successful results [13]. The number of experts was generally recommended between 10 and 50 people and must be from the same academic field. The study relies on the Surgery Branch of Chinese Medical Association Professional Committee of Breast Disease Platform and has established good cooperative relations in the process of preliminary enquiry. Regarding distributing the questionnaire, we contacted experts by a combination of post, e-mail, and telephone to explain the questionnaire and listen to their suggestions or recommendations. In fact, our result demonstrated that above communications effectively enhance the enthusiasm of experts and made up the flaws of tight time and curt answers.

In previous studies involving the use of Delphi method for the syndrome differentiation defined by TCM in breast cancer, we found that the inconsistence of experts in each round investigation would greatly impact the final statistical results. In order to avoid such mistake, our study guaranteed that the experts in both rounds had high consistency by selecting eighteen experts from the first round investigation. Meanwhile, we followed the principle of the Delphi method to ensure that the consultation and the leading group consisted of different people in order to achieve the high authenticity and accuracy of the results.

Previous research suggested that the perichemotherapy/radiotherapy refers to the period from the start until one week later after the end of chemotherapy or radiotherapy. However, our study indicated that experts tended to define the period extending two weeks later after the end of treatment. Meanwhile, the experts pointed out that the side effects of chemotherapy/radiotherapy could be generally relieved

in two weeks after the end of treatment in the presurvey; therefore they proposed to define the above-mentioned time length and the suggestion was also supported by most experts in consultation. What is more, in clinical setting we also recommended that patients start to receive endocrine therapy two weeks later after the end of chemo/radiotherapy, since the side effects will be usually relieved at that time point. Based on clinical observation and experts' discussion, the perichemotherapy/radiotherapy was finally defined two weeks later after the end of treatment.

At present, the treatment of breast cancer in western medicine includes surgery, chemotherapy, radiotherapy, endocrine therapy, and molecular targeted therapy and has developed rapidly, but the side effects of western medicine severely influenced patients' quality of life and prognosis. Therefore, TCM plays an important role in relieving the side effects brought by chemotherapy or radiotherapy. The aim of the syndrome differentiation defined by Traditional Chinese Medicine in operative breast cancer is to treat the patients on different Zheng composition and to get a better quality of life.

The establishment of each syndrome differentiation involved in this study was based on systematic literature review and experts' discussion. However, some of the methods and conclusions may be premature and require further research to consummate the detailed syndrome classification in order to achieve higher applied significance in clinical practice.

Conflict of Interests

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

Acknowledgments

The authors would like to thank all members in the working group for their contribution to this study and to express their heartfelt gratitude to their mentor and their research team.

References

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[2] R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014," *CA: A Cancer Journal for Clinicians*, vol. 64, no. 1, pp. 9–29, 2014.

[3] Australian Institute of Health and Welfare and National Breast Cancer Centre, *Breast Cancer in Australia: An Overvie*, vol. 34, Australian Institute of Health and Welfare (AIHW), Canberra, Australia, 2006.

[4] A. Micheli, P. Baili, M. Quinn et al., "Life expectancy and cancer survival in the EUROCARE-3 cancer registry areas," *Annals of Oncology*, vol. 14, supplement 5, pp. v28–v40, 2003.

[5] M. Sant, T. Aareleid, F. Berrino et al., "EUROCARE-3: survival of cancer patients diagnosed 1990-94—results and commentary," *Annals of Oncology*, vol. 14, supplement 5, pp. v61–v118, 2003.

[6] M. P. Coleman, G. Gatta, A. Verdecchia et al., "EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century," *Annals of Oncology*, vol. 14, supplement 5, pp. v128–v149, 2003.

[7] H. Linstone and M. Turoff, *The Delphi Method: Techniques and Applications*, Institute of Technology, Newark, NJ, USA, 2002.

[8] J. Jones and D. Hunter, "Consensus methods for medical and health services research," *The British Medical Journal*, vol. 311, no. 7001, pp. 376–380, 1995.

[9] D. Pelletier, C. Duffield, A. Adams, S. Nagy, J. Crisp, and S. Mitten-Lewis, "Australian nurse educators identify gaps in expert practice," *Journal of Continuing Education in Nursing*, vol. 31, no. 5, pp. 224–231, 2000.

[10] W.-C. Chie, C.-S. Huang, J.-H. Chen, and K.-J. Chang, "Utility assessment for different clinical phases of breast cancer in Taiwan," *Journal of the Formosan Medical Association*, vol. 99, no. 9, pp. 677–683, 2000.

[11] A. Moussa and C. Bridges-Webb, "Quality of care in general practice. A delphi study of indicators and methods," *Australian Family Physician*, vol. 23, no. 3, pp. 465–473, 1994.

[12] M. Gallagher, C. Bradshaw, and H. Nattress, "Policy priorities in diabetes care: a Delphi study," *Quality and Safety in Health Care*, vol. 5, no. 1, pp. 3–8, 1996.

[13] B. Brown, "Delphi process: a methodology using for the elicitation of opinions of experts," *The Rand Corporation*, vol. 14, no. 9, p. 3925, 1987.