Review Article

Applications of Data Mining Methods in the Integrative Medical Studies of Coronary Heart Disease: Progress and Prospect

Yan Feng,[1] Yixin Wang,[1] Fang Guo,[1] and Hao Xu[2]

[1] *Department of General Practice, Anzhen Hospital, Capital Medical University, Beijing 100029, China*
[2] *Department of Cardiology, Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing 100091, China*

Correspondence should be addressed to Yixin Wang; wangyixin6417@sina.com and Hao Xu; xuhaotcm@hotmail.com

A large amount of studies show that real-world study has strong external validity than the traditional randomized controlled trials and can evaluate the effect of interventions in a real clinical setting, which open up a new path for researches of integrative medicine in coronary heart disease. However, clinical data of integrative medicine in coronary heart disease are large in amount and complex in data types, making exploring the appropriate methodology a hot topic. Data mining techniques are to analyze and dig out useful information and knowledge from the mass data to guide people's practices. The present review provides insights for the main features of data mining and their applications of integrative medical studies in coronary heart disease, aiming to analyze the progress and prospect in this field.

1. Introduction

Coronary heart disease (CHD) is a serious threat to human health, especially for the elderly. Integrative medicine (IM) specialists have accumulated a large number of data in the clinical practice of CHD, which contain important information about diseases, syndromes, syndrome diagnosis and thinking skills, prescription medication, treatment, prognosis and evolution syndrome, and other aspects of development trends. How to do our clinical researches relying on these objective, dynamically updated massive clinical data of IM for CHD, is the primary challenge for us [1–4].

At present, on current clinical research methods, due to the strict limitations in the included crowd and medication conditions for randomized controlled trials (RCTs), the studies have high internal validity but poor in external making difficulty for the findings in promoting practical application. On basis of practical international RCTs, real-world study (RWS) concepts and methods gradually rise, which is to reflect the real world as a whole through the "real world sample." It is to choose interventions according to the actual condition and willingness of the patient and evaluate the effects of interventions with more comprehensive coverage of the crowd using statistical methods such as propensity score to control confounding bias. Thus, RWS has strong external validity than the traditional RCTs and can evaluate the effect of interventions in a real clinical setting. Therefore, the results are much closer to clinical practice. Integrative interventions emphasize individualized treatment, focusing on holistic, complex, and multiple effects in the evaluation of clinical efficacy. RWS undoubtedly opens up a new path for researches of integrative medical in CHD.

However, clinical data of IM in CHD are large in amount and complex in data types. All are multivalued and multitypes data, the attribute and label of each record have one or more options, and the clinical research data also have more confounding factors, making exploring the appropriate methodology become a hot topic.

Data mining is an interdisciplinary research field, which combines the latest research achievements such as statistics, data warehousing, information retrieval, machine learning, artificial intelligence, high performance computing, and data visualization. Data mining techniques are to analyze and dig out data useful information and knowledge from the mass data to guide people's practices, which is changing the use

patterns of data with a new concept. Making the data mining techniques becomes a new researching way of RWS of IM for CHD.

To analyze clinical syndrome diagnosis and prescription experience of CHD in the real world with data mining methods, it cannot only find clinical rules and improve clinical diagnostic accuracy of CHD for IM physicians, but also get a deep understanding of IM academic thinking and grasp disease treatment rule. Therefore, using data mining methods for IM study in CHD will greatly improve the level of clinical diagnosis and treatment of IM study in CHD [5, 6] and has broad application prospects.

The main features of data mining include correlation analysis, classification and prediction, cluster analysis, and evolution analysis. In the field of exploring these features, RWS of IM in CHD has made considerable progress.

2. Correlation Analysis

Correlation analysis is to find interesting links between the association forms as X-Y items from large data, which can be interpreted as the probability that if X occurs, then Y also appears. Correlation analysis methods commonly include methods as association rule and complex network analysis. Data mining correlation analysis is widely used in etiology, clinical diagnosis, drug compatibility, and so on for IM clinical researches.

2.1. Association Rule

2.1.1. Principle. It is a description for relationship between one thing and other associated or interdepended things [7–11], focusing on characterization [12] and the degree of association between the objects in the database [13]. Among the association rules mining, Apriori algorithm is the most basic, famous, and influential one. The core idea of the Apriori algorithm is a recursive method based on the frequency set theory, whose purpose is to dig out association rules with support and confidence no less than the minimum support threshold (min_sup) and minimum confidence threshold (min_conf) from the database.

2.1.2. Characteristics. Interesting links hidden in large data can be showed by association rules or frequent item sets.

2.1.3. Application Examples in CHD. A study found the most common syndromes of CHD after statistical analysis, which are mostly phlegm and blood stasis, water and blood stasis, and obstruction of coronary circulation syndromes; secondly common syndromes were syndrome of deficiency complicated with excessiveness, such as qi deficiency and blood stasis and yin deficiency with yang hyperactivity syndromes. The results reflect the thought of "two bu," "three tong" in the treatment of CHD [14].

The association rule mining method was applied to a small sample study of CHD clinical data, using contingency table of definite probability method as a measure of association rules to find relationships between variables in small samples and get all two rules [15]. After a number of rules

pruning based on the confidence, the results can more fully reveal the implications of the data information than that obtained by logistic regression method [16]. In the analysis of the law of prescription for famous TCM doctors with association rule, some basic recipes such as Huoxuetongmai agents, Shengmaisan, for treatment of CHD were also found [17]. There is theoretical and practical significance in how quickly and effectively mining association rules for some rational use from a massive database [18, 19].

2.2. Complex Network

2.2.1. Principle. For the multiple elements of complex systems, it can be interconnected and form nodes due to certain inherent potential relationships. Most of the nodes have only a few connections, but some nodes have a lot of connections with other nodes. Complex system is a network system hosted by a few hubs, having a large number of functional groups composed with connecting hubs, which may reflect some or all of its overall characteristics [20–23]. Complex network is a method of data mining to dig out implicit, previously unknown, and potentially valuable to the decision-making relationships, patterns, and trends from large data [24].

2.2.2. Characteristics. This method can find potential links between two or more elements of complex systems, showing the relationship between picture elements.

2.2.3. Application Examples in CHD. Using cross-sectional survey, the study collected the clinical information of 3018 hospitalized CHD patients through individualized information acquisition platform of CHD. The relationships among syndrome, therapeutic treatment, and Chinese herbs were excavated by means of complex networks based on theory of correspondence between prescription and syndrome.

It found that the fundamental syndrome factors were blood stasis, qi deficiency, phlegm-turbid, yin deficiency, yang deficiency, qi stagnation, and blood deficiency. The therapeutic treatment mainly included activating blood circulation, clearing heat, invigorating qi, resolving turbid and phlegm, nourishing yin, warming yang qi, and dispersing obstruction. These methods constituted an association with major syndrome factors. The major syndrome factors constituted an association with the following Chinese herbal medicines: Huang qi (Radix Astragali Mongolici), Chen pi (Pericarpium Citri Reticulatae), Di Huang (Radix Rehmanniae), Chuanxiong (Rhizoma Chuanxiong), Baizhu (Rhizoma Atractylodis Macrocephalae), Taoren (Semen Persicae), Fuling (Poria), Gan cao (Radix Glycyrrhizae), Ban Xia (Pinellia Ternate), Ze Xie (Rhizoma Alismatis), Chi Shao (Radix Paeoniae Rubra), Dang Gui (Radix Angelicae Sinensis), Danshen (Radix Salviae Miltiorrhizae), Zhi Qiao (Fructus Aurantii Submaturus), Gui Zhi (Ramulus Cinnamomi), and Mai Dong (Radix Ophiopogonis Japonici). The efficacies of Chinese herbal medicines associated with syndrome factors mainly include alleviating pain, resolving turbid and phlegm, clearing heat, activating blood circulation, invigorating qi, cooling blood, promoting urination, resolving stagnation, removing toxic material, nourishing

blood, regulating qi, quieting spirit, invigorating spleen, regulating menstruation, promoting defecation, moistening dryness, and resolving stasis.

The therapeutic methods for CHD are based on consistency in theory, method, formula, and medicines. The application of therapeutic methods for clearing heat and removing toxical material was compared relatively more with other methods, so it is necessary to separate heat as the complement blood stasis, phlegm, and qi stagnation syndromes to more fully reflect evolution and characteristics of CHD syndromes [25].

3. Classification and Prediction

Classification and prediction are two forms of data analysis. Classification is to analyze training data set to identify the typical characteristics of data in the same class based on the characteristics of data and use them to classify new data. The key of classification is to export functions or models for classification. There is a long time to study the issue and put forward many methods used to derive the classification models, including decision tree classification and neural network. Classification is for predicting discrete categories of data objects; when the prediction data objects are not a class but continuous value, it is often called prediction.

The data mining functions of classification and prediction is widely used in medical diagnostics, disease risk prediction and other fields for clinical practice of IM.

3.1. Decision Tree

3.1.1. Principle. It is a data classification process through a series of rules [26]. Determined by a series of "if then" logic (branching) relationships, this method inference a set of classification rules from a set of no order and no rules examples, and express the distribution probabilities of all possible outcomes 9 with a tree chart as the decision tree, so as to achieve the purpose of predicting accurately or correct classification [27].

Decision tree is being used more and more in clinical studies, especially in the clinical diagnosis [28]. In IM researches, the tree model is mainly applied in standardization of syndromes characteristics and diagnosis [29], setting up medical model [30], influencing factors of syndrome changes, and evaluation on the efficacy of IM researches [31].

3.1.2. Characteristics. This method combines the disorder existing data together and build relationships connected layer upon layer to classify and predict the targets or outcomes.

3.1.3. Application Examples in CHD. Decision tree pattern could identify phlegm-blood stasis syndrome of unstable angina patients clearly and more intuitively, and it also could self-extract recognition rules. It had advantage in the data mining of syndrome-clinical physicochemical index corresponding pattern [32–35]. Besides that, tree structure models were built to summarize the correspondence between qi deficiency syndrome and physicochemical index based on t test, nonparametric analysis, and Spearman correlation analysis. The study found that the accuracy identification rate of tree structure model of qi deficiency syndrome with six core indexes, such as EF and P-R interval, was 77.78%. Decision tree model can identify qi deficiency syndrome of CHD patients with diabetes clearly and more intuitively. Decision tree model is a promising method in data mining of qi deficiency syndrome and index association patterns [36].

3.2. Artificial Neural Network

3.2.1. Principle. Artificial neural network, also known as connection machine model, is produced on interdisciplinary researches in modern neurology, biology, psychology, and so forth. As a computing system developed on the basis of simulation of human brain tissue, it reflects the fundamental process of processing external things by biological neural system. It is a network system composed of a large number of interconnected processing with basic characteristics of the biological nervous system. It is reflects the several brain functions to a certain extent, a kind of simulation to biological system, having the abilities of nonlinear mapping and learning, adaptability, fault tolerance, and associative storage.

3.2.2. Characteristics. The independent variables may be continuous in the model application or may be discrete, regardless of whether the normality of variables and independence between variables and other conditions are satisfied. It can identify the complex nonlinear relationships between the variables; especially when using conventional methods cannot achieve the purpose of statistical analysis or ineffective, this model can often receive good results.

3.2.3. Application Examples in CHD. Based on the clinical epidemiology investigation in coronary heart disease, the study constructed an artificial neural network model of the CHD of Chinese medicine syndromes on the basis of neural network toolbox, and then it tests the performance of this model using a retrospective of inspection and prospective testing method. The diagnostic accuracy of rate is 90.5% in 496 cases of already collected of retrospective examination showed, and specific syndrome types of discrimination and sample number accuracy were positively correlated. The diagnostic accuracy of rate is 91.36% in new collection of 132 cases of the prospective examination showed. It is thought that the artificial neural network can better the forensics of syndrome because it can explore the internal rules of TCM syndrome. And there is a good prospect using artificial neural network in the study of standardization of traditional Chinese medical syndrome [37].

4. Cluster Analysis

Cluster analysis is to divide physical or abstract objects into several groups or classes based on "greatest intragroup similarity, smallest intergroup similarity" principle. Clustering is unsupervised learning, and the input set is a set of nonpredefined classes records without any classification. Good clustering method ensures that intragroup similarity is very high and intergroup similarity is very low. Cluster

analysis function of data mining is widely used in data integration, analysis of clinical features, and other aspects in the field of IM for CHD.

4.1. Clustering Analysis

4.1.1. Principle. It is a kind of mathematical statistics for the study of "like attracts like." Cluster analysis can put some of the observed objects to be classified according to certain characteristics, and there is a wide range of applications in biology and medical classification problems [38].

4.1.2. Characteristics. It can do some basic classification for hybrid data to make it well-organized and easy to find data characteristics.

4.1.3. Application Examples in CHD. With cluster method to analyze the database, we summarize the syndromes features of four periods: (1) early onset period: qi open-minded syndrome and qi and yin deficiency syndrome, (2) paroxysm period: qi stagnation and phlegm obstruction syndrome, yang malaise heart syndrome, cold coagulation heart vessel syndrome, and blood stasis yang syndrome, (3) remission period: liver and spleen no coordination syndrome, yang deficiency of heart and kidney syndrome, and cardiopulmonary qi deficiency syndrome, and (4) recovery period: heart qi deficiency syndrome, yang deficiency and qi stagnation syndrome, and qi and yin deficiency syndrome. That enriches the content of CHD syndromes [39].

The K-center clustering method was adopted for the analysis of clinical data and syndrome information of 154 cases of prethrombosis state. The results showed that traditional clinical syndrome differentiation presented 12 syndrome patterns, and they were blood stasis syndrome, qi deficiency syndrome, damp turbidity syndrome, yin deficiency syndrome, yang deficiency syndrome, phlegm turbidity syndrome, damp-heat (toxicity) syndrome, qi stagnation syndrome, blood deficiency syndrome, phlegm heat syndrome, and cold accumulation syndrome. Among the 12 patterns, the blood stasis syndrome and qi deficiency syndrome were more commonly seen than other syndromes, accounting for 49.1%, and cold accumulation syndrome was most rarely seen. Syndrome clustering analysis results presented 4 syndrome patterns, yang deficiency and blood stasis accounted for 60.4%, syndrome of phlegm damp aggregated with heat and qi stagnation accounted for 20.1%, qi and yin deficiency syndrome accounted for 13.0%, and cold accumulation syndrome accounted for 6.5%. Yang deficiency and blood stasis syndrome was the most common type. Cluster analysis is thought to be helpful for the research of Chinese medical syndrome and can provide reliable basis for syndrome differentiation, which will lay the foundation for IM treatment and efficacy evaluation [40].

4.2. Shannon Entropy Mutual Information

4.2.1. Principle. It is such a complex system divided manner with probability entropy as premise to extract feature combinations with maximum information by calculating a correlation coefficient for each variable and the other variables.

4.2.2. Characteristics. This method is not entirely dependent on the frequency, but also on the two variables appearing or not simultaneously to characterize the correlation between the two aspects. This method is used not only to deal with linear data, but also to deal with nonlinear data.

4.2.3. Application Examples in CHD. The data mining technology based on the Shannon entropy mutual information was used to analyze the complicated correlations of the statistical distribution of CHD syndromes associated physical and chemical indexes. It was found that 7 in 13 syndrome factors including qi deficiency, blood stasis, turbid phlegm, yin deficiency, cold coagulation, yang deficiency, and qi stagnation which have been most researched were involved in about 134 physical and chemical indexes mentioned in the literature. It obtained the ranked top 10 physical and chemical indexes of each syndrome factors, after analysis calculation. The study suggested there were perplexing relations with Chinese medical syndromes and physical and chemical indexes, which can be revealed by the data mining technology based on the Shannon entropy mutual information [41].

5. Evolutionary Analysis

Evolutionary analysis is to describe and model following the changing laws or trends of time-varying objects, including data analysis as time series, sequence, or cycle matching pattern. The evolution of data mining analysis capabilities in the field of IM is widely used to predict clinical outcomes and evaluate the efficacy of clinical programs.

Technologies used in the studies of CHD by IM include Bayesian network, support vector machine, Markov model, and random walk model.

5.1. Bayesian Network

5.1.1. Principle. Based on Bayes' theory, *Bayesian Network* has solid foundation of statistical theory and strong integration capabilities for the sample information and prior knowledge. As one of the important data mining algorithms for classification [42], *Bayesian Network* is an ideal reasoning method of uncertainty researches, which can give the probability for samples belonging to a specific class and also minimize rates of error and risk in the reasoning process.

If the likelihood of event results cannot be predicted, then *Bayesian Network* is the only way to quantify that the probability is to obtain the occurrence of the event. Bayesian classification is a typical statistical classification method. The common practice is to establish a link of the prior probability and posterior probability of the event and then to determine the category to the largest posterior probability sample through the judgment of posterior probability.

5.1.2. Characteristics. The occurrence probability of final outcome is calculated by existing data.

5.1.3. Application Examples in CHD. Based on Bayesian network, a paper studies the construction of Chinese medical

clinical diagnosis model for CHD, and gain information algorithm is used to choose the fields used for the two models. The experimental data are selected from the electronic patient information database, and the experimental results show that Bayesian networks have better classification capability in Chinese medical clinical diagnosis mode [43].

With syndrome factors and combination law as the observing points, another study used Bayesian network to do some qualitative and quantitative researches for syndrome factors and dig CHD syndrome from the prominent Chinese medical doctors database and achieved good results. Besides the above, it also used in blood stasis syndrome differentiation and setting up CHD diagnosis model [44–46].

5.2. Support Vector Machine (SVM)

5.2.1. Principle. As a monitoring statistical learning method based on the structural risk minimization principle, *SVM* can get a globally optimal solution without the need to seek prior probability. It can use a certain preselected nonlinear mapping to make the input vectors mapped to a high dimensional feature space and in the high dimensional feature space to construct the optimal separating hyperplane, finally, to do classification making use of the hyperplane.

5.2.2. Characteristics. It is a statistical learning model under recognized high-dimensional and small sample sizes and is generally applicable in classification and regression studies. Its trained models have the characteristics of global optimum, and as long as the parameters $\{C, \sigma\}$ are the same, the results will remain stable and consistent training [47, 48].

5.2.3. Application Examples in CHD. A study set up a database of diagnosing and treating CHD and was on the basis of 115 typical medical records from prominent Chinese medical doctors. The syndrome factors and relevant studies were analyzed and conducted by using SVM. It found that there were mainly 8 syndrome factors draw, including blood stasis, turbid phlegm, qi deficiency, yang insufficiency, yin deficiency, inner heat blood deficiency, and qi stagnation. The quantitative diagnosis was confirmed and CHD characteristics were explained. The laws of medicate administration for abovementioned 8 syndrome factors were summed up from prominent Chinese medical doctors for treating CHD [49].

5.3. Partially Observable Markov Decision Process (POMDP)

5.3.1. Principle. POMDP model is a dynamic decision model based on Markov process promoted by the Russian mathematician Markov after some improvements, which is the most common method in a dynamic programming strategy. Its purpose is to seek the best solution in many prescriptions applying optimization techniques [50].

5.3.2. Characteristics. Data acquisition is performed in parallel with the operation process, and the entire data acquisition does not require large-scale acquisition process as long as the regulation of data entry has been established, which greatly saves the cost of the study to facilitate the continuous optimization of the prescriptions. Furthermore, this dynamic process is a combination of man and machine model, and the strict mathematical operation is carried out at the same time while doing empirical evaluation.

5.3.3. Application Examples in CHD. Based on existing data, applying POMDP to compare the prescriptions of patients with same TCM syndrome element and no long-term endpoint event, it was found that optimizing prescription recommendation for "qi deficiency" patients is "Milkvetch root + Si junzi decoction without Radix Glycyrrhizae," prescription of "blood stasis" recommended: "Danshen root + Taohong Siwu decoction plus orange fruit without rehmanniae radix," and prescription of "turbid phlegm" recommended: "Gualou xiebai banxia decoction plus dried tangerine peel, Largehead Atractylodes rhizome, Platycodon root." The prescriptions derived from real clinical data are experiences and summaries of clinical practice with considerable clinical significance. The proposals using this rigorous mathematical comparison method are in conformity with the clinical normal circumstances and prove the reliability and operability of optimizing prescription method in efficacy evaluation on the other hand [51].

5.4. Random Walk Model

5.4.1. Principle. Random Walk Model is a commonly mathematical model simulating the statistical mathematics to provide the best possible state. Random walk model is a way of exploring the movement of things that set probability theory and dissipative structure theory in one. Its basic idea is, given a particle in space, and its moving vectors in space (including the direction and distance) are controlled by a random amount of transition probabilities, which can simulate complex process, such as the molecular Brownian motion of nature and electronic random motions in the metal.

5.4.2. Characteristics. It can compare and evaluate treatment options based on the dynamic changes after the intervention.

5.4.3. Application Examples in CHD. The study evaluates the clinical effects of Shengmai injection in treating CHD based on correct syndrome differentiation and incorrect syndrome differentiation and found that there were 273 patients in the correct syndrome group and 4 patients died (case-fatality rate was 1.47%). There were 297 patients in the incorrect syndrome group and 7 patients died (case-fatality rate was 2.36%). In the correct syndrome group, random fluctuation peak of comprehensive evaluation index, walk steps, positive growth rate of walk, ratio, random fluctuation power law, increase rate, and record times of comprehensive evaluation index were 1472, 13617, 0.1081, 9.25, 0.6742, 0.4706, and 3128, respectively, while, in the incorrect syndrome group, 1030, 14588, 0.0706, 14.16, 0.6606, 0.3128, and 3293, respectively. The random fluctuation power law in both groups exceeded 0.5. There is a long-range correlation between the comprehensive evaluation index and therapeutic method in the CHD patients were treated with Shengmai injection. The clinical therapeutic effects of Shengmai injection under

correct syndrome differentiation are better than the effects of Shengmai injection under incorrect syndrome differentiation [52].

6. Other Functions: Text Mining

It is a direction of data mining, in which you can find potential patterns and trends from millions of text data. In the field of IM, text mining can discover knowledge from amounts of literatures in order to promote development of clinical research and treatment programs in IM and provide new ideas and ways for IM studies with more objective and repeatable results [53].

6.1. Principle. text mining method is to find and mine inductive knowledge from the texts such as useful models, trends, and rules [54–57]. The text knowledge discovery technology, which is text mining technology, is the product of artificial intelligence, machine learning, natural language processing, data mining, and related automatic text processing such as information extraction, information retrieval, and text classification. Information extraction positions target data units from natural language texts and put the unstructured free texts into structured data that meet the application of requirements, which is extract free text data to fill predefined structured templates.

Traditional machine learning methods described above, such as neural networks, Bayes network, decision tree, k-nearest neighbor, and support vector machine, are all used for text classification and archiving [58]. The recent application of a relatively new model in diagnosis and treatment in the field of IM for CHD is the topic model technology.

6.2. Topic Model

6.2.1. Principle. As the product of text mining and natural language processing technology in recent years, *Topic Model* is a statistical model that can extract a class of topics implicit in the documentation set (not limited to text documents, it can be other discrete data sets), each of which is distribution of some words with related semantics. Topic models use the topic ideas of the text to make the text from the high dimension of the word to the low dimension of the topic, which reduces the high-dimensional sparse feature of the text and the effect of noise word processing in text information, as well as capturing the semantics of the text.

6.2.2. Characteristics. The characteristics of topic model are (1) easy for effective representation, organization, and storage of the text; (2) easy for semantic information retrieval, information extraction, automatic summarization extracts, and other operations according to the semantics; and (3) easy for effective text classification and clustering.

6.2.3. Application Examples in CHD. Based on topic model, the study analyzed the accompanied syndromes, comorbidities, and usage of Chinese herbal for preliminary optimization of the treatment regimen in different of accompanied syndromes and complication.

Seeing from the experiments of topic model, the obtained results are consistent with the actual situation in which the model can get the hierarchical relationship of the data. To set the accompanied syndromes and complications of a patient, the model can predict the corresponding use of Chinese herbs. Similarly, to give a combination of some Chinese herbs, the model can infer what accompanied syndromes and complication of the patient are. Topic model can extract the regularity of treatment regimens with clinical significance, provide a novel theoretical approach for the study of treatment regimen optimization, give objective evidence for the clinical syndrome and disease differentiation treatment, and set a new statistical analysis method for analysis of prescription with varied syndromes and diseases [59].

7. Conclusions

Real-world studies of modern medicine in CHD have got a rapid development in recent years with a wide range and large-scale covering and formed a multicountry research participation model such as GRACE. In contrast, real-world study of CHD in IM is just at the beginning, and the studies based on more than a million cases of CHD have not been reported, particularly lacking suitable real-world research methodology system of IM. It is undoubtedly a useful exploration to introduce cutting-edge technologies—data warehousing and data mining in the field of information to IM clinical studies, and establish massive data based, data-driven clinical research model to solve technical bottlenecks in IM clinical researches with individual clinics as a feature.

However, the application of data mining methods in clinical intervention to CHD for RWS problems of IM has established a relatively complete paradigm and technical support currently but is still in the developing stage, and its own methodology and practical application are continuously improved. How to get mining methods and clinical practice closely combined, how to better interpret and apply data mining results, and how to better improve the traditional mining methods also need further exploration to seek better solutions. Following the general rules of data mining methods [60], combining it with the characteristics of clinical practice in IM [61–63], continuing to explore suitable data mining methodology, and getting continuous optimization improvement in practice on basis of the existing database are the development direction of RWS of IM for CHD in the future.

Conflict of Interests

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

Acknowledgments

References

[1] Y.-T. Fang and S.-R. Wang, "An investigation on clinical studies of TCM in preventing and treating angina pectoris of coronary heart disease," *Chinese Journal of Integrated Traditional and Western Medicine* , vol. 23, no. 5, pp. 338–340, 2003.

[2] J. Tang and C. S. Tang, "Development trends of cardiovascular and cerebrovascular disease researches," *Journal of Peking University (Health Sciences)*, vol. 33, no. 4, pp. 289–291, 2001.

[3] Y. L. Liao, "Advance in external therapies of coronary heart disease," *Journal of Practical Traditional Chinese Medicine*, vol. 29, no. 2, pp. 152–153, 2005.

[4] K. J. Chen and Y. Lei, "Strengthen the prevention and treatment of unstable angina," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 16, no. 10, p. 579, 1996.

[5] J. M. Chen, *Data Warehousing and Data Mining Technology*, Electronic Industry Press, Beijing, China, 2002.

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

[7] A. Ceglar and J. F. Roddick, "Association mining," *ACM Computing Surveys*, vol. 38, no. 2, 2006.

[8] L. Cao, H. Zhang, Y. Zhao, D. Luo, and C. Zhang, "Combined mining: discovering informative knowledge in complex data," *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 41, no. 3, pp. 699–712, 2011.

[9] C. Ordonez and K. Zhao, "Evaluating association rules and decision trees to predict multiple target attributes," *Intelligent Data Analysis*, vol. 15, no. 2, pp. 173–192, 2011.

[10] C.-M. Wu and Y.-F. Huang, "Generalized association rule mining using an efficient data structure," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7277–7290, 2011.

[11] Y.-F. Huang and C.-M. Wu, "Preknowledge-based generalized association rules mining," *Journal of Intelligent and Fuzzy Systems*, vol. 22, no. 1, pp. 1–13, 2011.

[12] D. Taniar, W. Rahayu, V. Lee, and O. Daly, "Exception rules in association rule mining," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 735–750, 2008.

[13] Y.-L. Chen and C.-H. Weng, "Mining fuzzy association rules from questionnaire data," *Knowledge-Based Systems*, vol. 22, no. 1, pp. 46–56, 2009.

[14] J. C. Zhang and Y. H. Xie, "Statistical digging study on the academician Chen Keji's medical records on coronary heart disease," *World Journal of Integrated Traditional and Western Medicine*, vol. 3, no. 1, pp. 4–5, 2008.

[15] J. X. Chen, G. C. Xi, and W. Wang, "A comparison study of data mining algorithms in CHD clinical application," *Beijing Biomedical Engineering*, vol. 27, no. 3, pp. 249–252, 2008.

[16] K. Duan, J. H. Wu, and J. He, "The application of association rules in clinical data of small sample," *Shenzhen Journal of Integrated Traditional and Western Medicine*, vol. 17, no. 2, pp. 91–94, 2007.

[17] H. Wu, X. Liu, J. Wang, and X.-Z. Zhou, "Study on law using Chinese drug of famous old docter of traditional Chinese medicine to coronary heart disease based on association rules," *China Journal of Chinese Materia Medica*, vol. 32, no. 17, pp. 1786–1788, 2007.

[18] J. X. Yu, Z. H. Li, and G. M. Liu, "A data mining proxy approach for efficient frequent itemset mining," *VLDB Journal*, vol. 17, no. 4, pp. 947–970, 2008.

[19] M. Wojciechowski, K. Galecki, and K. Gawronek, "Three strategies for concurrent processing of frequent itemset queries using FP-gowth," in *Knowledge Discovery in Inductive Databases*, vol. 4747, pp. 240–258, 2007.

[20] X. Z. Zhou, B. Y. Liu, Y. H. Wang et al., "Study of method on Chinese herbal compound compatibility by complex networks," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 15, no. 11, pp. 98–100, 2008.

[21] Q. Ni, S. B. Chen, X. Z. Zhou et al., "Study of relationship between formula (herbs) and syndrome about type 2 diabetes mellitus affiliated metablic syndrome based on the free-scale network," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 13, no. 11, pp. 19–22, 2006.

[22] G. J. Ortega, R. G. Sola, and J. Pastor, "Complex network analysis of human ECoG data," *Neuroscience Letters*, vol. 447, no. 2-3, pp. 129–133, 2008.

[23] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

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

[25] Z.-Y. Gao, J.-C. Zhang, H. Xu et al., "Analysis of relationships among syndrome, therapeutic treatment, and Chinese herbal medicine in patients with coronary artery disease based on complex networks," *Journal of Chinese Integrative Medicine*, vol. 8, no. 3, pp. 238–243, 2010.

[26] L. Y. Luo and J. X. Chen, "Data mining and investigation of surgical operation information based on decision tree," *Medical Information*, vol. 21, no. 11, pp. 1936–1939, 2008.

[27] J. Q. Fang, J. Y. Luo, K. B. Yao et al., "Application of decision tree C5.0 in the pre-warning of birth," *Chinese Journal of Health Statistics*, vol. 26, no. 5, pp. 473–476, 2009.

[28] Y. Cheng and Y. T. Cui, "Application of decision tree algorithm based on PCA in the application of heart disease diagnosis," *Computer and Digital Engineering*, vol. 37, no. 10, pp. 171–174, 2009.

[29] H.-B. Qu, L.-F. Mao, and J. Wang, "Method for self-extracting diagnostic rules of blood stasis syndrome based on decision tree," *Chinese Journal of Biomedical Engineering*, vol. 24, no. 6, pp. 709–711, 2005.

[30] Y. Zhong, X. L. Hu, and J. F. Lu, "Diagnostic analysis of gastritis in traditional Chinese medicine based on association rules and decision tree," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 15, no. 8, pp. 97–99, 2008.

[31] Q.-L. Zha, Y.-T. He, and J.-P. Yu, "Correlations between diagnostic information and therapeutic efficacy in rheumatoid arthritis analyzed with decision tree model," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 26, no. 10, pp. 871–876, 2006.

[32] Q. Shi, J. X. Chen, H. H. Zhao et al., "Establishment of qi deficiency syndrome and physicochemical index association patterns in coronary heart disease with diabetes patients based on decision tree," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 27, no. 6, pp. 1538–1540, 2012.

[33] Q. Shi, J. X. Chen, H. H. Zhao et al., "Study on decision tree patterns of coronary heart disease patients with blood stasis syndrome," *Chinese Journal of Integrative Medicine on Cardio/Cerebrovascular Disease*, vol. 11, no. 8, pp. 897–900, 2013.

[34] J. X. Chen, C. Xi, W. Wang et al., "A comparison study of data mining algorithms in CHD clinical application," *Chinese*

Journal of Biomedical Engineering, vol. 27, no. 3, pp. 249–252, 2008.

[35] Q. Shi, J. X. Chen, H. H. Zhao et al., "Distribution mode of four-examination information based on complex network technique in CHD patients," *Journal of Beijing University of Traditional Chinese Medicine*, vol. 35, no. 3, pp. 183–188, 2012.

[36] Q. Shi, W. Wang, J. X. Chen et al., "Recognition patterns study of coronary heart disease patients with phlegm-blood stasis syndrome based on decision tree," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 28, no. 12, pp. 3523–3526, 2013.

[37] G. X. Sun, X. Y. Yao, Z. K. Yuan et al., "The realization of the BP neural network model based on the MATLAB coronary heart disease of TCM syndrome," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 29, no. 82, pp. 1774–1776, 2011.

[38] H.-L. Wu, X.-M. Ruan, and W.-J. Luo, "Cluster analysis on TCM syndromes in 319 coronary artery disease patients for establishment of syndrome diagnostic figure," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 27, no. 7, pp. 616–618, 2007.

[39] Y. P. Chang, *Factors of coronary heart disease syndromes, syndrome characteristics and evolution of the syndrome pathogenesis [Ph.D. thesis]*, Liaoning University of Traditional Chinese Medicine, 2011.

[40] X. X. Wu, B. J. Chen, R. F. Zeng et al., "Cluster analysis of clinical features and distribution of traditional Chinese medicine syndrome patterns in 154 cases of pre-thrombosis state," *Journal of New Chinese Medicine*, vol. 46, no. 4, pp. 77–79, 2014.

[41] C. Chen, Y. M. Meng, P. Zhang et al., "Diagnosis and treatment rule of traditional Chinese medicine for syndrome factors of chronic congestive heart failure: a study based on Shannon entropy method," *Journal of Chinese Integrative Medicine*, vol. 8, no. 11, pp. 1080–1083, 2010.

[42] Q. Y. He and J. Wang, "Research on TCM syndromes and diagnosis of patients after percutaneous coronary intervention based on cluster analysis," *Journal of Traditional Chinese Medicine*, vol. 49, no. 10, pp. 249–252, 2008.

[43] Y. N. Sun, S. Y. Ning, M. Y. Lu et al., "Chinese traditional medical clinical diagnosis for coronary heart disease based on Bayes classification," *Application Research of Computers*, vol. 23, no. 11, pp. 164–166, 2006.

[44] R. Wu, X. Y. Nie, J. Wang et al., "Coronary Heart Syndrome laws of old TCM doctors based on Bayesian network," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 17, no. 5, pp. 98–99, 2010.

[45] Z. Liu, G. M. Sang, M. Y. Lu et al., "Coronary heart disease diagnosis model based on weighted Bayesian categorization," *Journal of Guangxi Normal University: Natural Science*, vol. 26, no. 4, pp. 67–70, 2008.

[46] L. Z. Li, Z. Y. Gao, R. X. Xi et al., "Analysis of blood stasis of CHD based on Bayesian network," in *Proceedings of the 4th Academic Conference for World Federation of Chinese Medicine Societies Professional Committee of Cardiovascular Disease*, pp. 108–113, 2010.

[47] V. M. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2nd edition, 2000.

[48] Q. Yuan, C. Cai, H. Xiao, X. Liu, and Y. Wen, "SVM-aided cancer diagnosis based on the concentration of the macroelement and microelement in human blood," *Journal of Biomedical Engineering*, vol. 24, no. 3, pp. 513–518, 2007.

[49] J. Wang, R. Wu, and X. Z. Zhou, "Syndrome factors based on SVM from coronary heart disease treated by prominent TCM doctors," *Journal of Beijing University of Traditional Chinese Medicine*, vol. 31, no. 8, pp. 540–543, 2008.

[50] X. M. An and L. Lin, "Markov model research progress in the vital statistics," *Chinese Journal of Health Statistics*, vol. 24, no. 4, pp. 436–439, 2007.

[51] Y. Feng, H. Xu, K. Liu, X. Z. Zhou, and K. J. Chen, "Optimized treatment program for unstable angina by integrative medicine based on partially observable Markov decision process," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 33, no. 7, pp. 878–882, 2013.

[52] Z.-Y. Gao, H. Xu, K.-J. Chen, D.-Z. Shi, L.-Z. Li, and X.-Z. Zhou, "Efficacy evaluation of Shengmai Injection in treating coronary heart disease based on random walk model," *Journal of Chinese Integrative Medicine*, vol. 6, no. 9, pp. 902–906, 2008.

[53] T. Lv and Y. H. Jiang, "Application of text mining in biomedical field and its system tools," *Chinese Journal of Medical Library and Information Science*, vol. 19, no. 4, pp. 56–64, 2010.

[54] 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.

[55] H. Ahonen, O. Heinonen, M. Klemettinen et al., "Mining in the phrasal frontier," in *Principles of Data Mining and Knowledge Discovery*, vol. 1263, pp. 343–350, 1997.

[56] U. Y. Nahm and R. J. Mooney, "Text mining with information extraction to appear in the AAAI," in *Proceedings of the Spring Symposium on Mining Answers Grom Texts and Knowledge Bases*, 2002.

[57] U. Y. Nahm and J. Raymond, "Using information extraction to aid the discovery of prediction rules from text," in *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD '00) Workshop on Text Mining*, pp. 51–58, Boston, Mass, USA, 2000.

[58] Y. Kodratoff, "Knowledge discovery in texts: a definition, and applications," in *Foundations of Intelligent Systems: Proceedings of the 11th International Symposium, ISMIS'99 Warsaw, Poland, June 8–11, 1999*, vol. 1609 of *Lecture Notes in Computer Science*, pp. 16–29, Springer, Berlin, Germany, 1999.

[59] T. Joachims, "Text categorization with support vector machines: learning with many relevant features," LS VIII Technical Report no. 23, University of Dortmund, 1997.

[60] Y. Feng, *Integrative medicine program optimization based on Coronary syndrome factors [Ph.D. thesis]*, China Academy of Chinese Medical Sciences, 2012.

[61] S. C. Wang, X. Q. Wang, N. N. Xiong et al., "Countermeasure to ethical review development of TCM clinical research," *Journal of Beijing University of Traditional Chinese Medicine*, vol. 33, no. 3, pp. 156–158, 2010.

[62] Y. F. Zhao, L. Y. He, B. Y. Liu et al., "Syndrome classification based on manifold ranking for viral hepatitis," *Chinese Journal of Integrative Medicine*, vol. 20, no. 5, pp. 394–399, 2014.

[63] X. J. Fu, X. X. Song, L. B. Wei, and Z. G. Wang, "Study of the distribution patterns of the constituent herbs in classical Chinese medicine prescriptions treating respiratory disease by data mining methods," *Chinese Journal of Integrative Medicine*, vol. 19, no. 8, pp. 621–628, 2013.