Exploratory Research and Hypothesis in Medicine **2023** vol. 000(000) | 000–000 Epub
DOI: 10.14218/ERHM.2023.00009

Original Article

Construction of Chinese Medicine Prescription Data Mining Based on Association Rules of the System

Jian-She Yang[1,2,3*], Qiang Wang[2], Si-Yang Zhou[2], Fen-Fen Zhang[2] and Zhong-Wei Lv[1]

[1]Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China; [2]Gansu Medical College, Pingliang, China; [3]Lanzhou University Second Hospital, Lanzhou, China

Received: February 21, 2023 | **Revised:** May 04, 2023 | **Accepted:** May 30, 2023 | **Published online:** July 7, 2023

Abstract

Background and objectives: Traditional Chinese medicine (TCM) prescriptions are complex compositions according to certain compatibility rules to treat different diseases. Therefore, understanding the superposition, inhibition, and interaction between the effective components is essential. This study aimed to use modern information technology and the established TCM databases to analyze and explore the effectiveness of TCM compositions in terms of prescription and regulation.

Methods: We adopted data mining technology to analyze the composition rules and active components among the spleen and stomach prescriptions that mainly includes the core algorithms of the apriori method and the frequent pattern (FP)-growth method.

Results: In the present work, a new algorithm was developed that uses the modified FP-tree and the head table structure, producing only the FP-tree once and the head table structure at each recursion.

Conclusions: The new algorithm can be used to obtain the same results as the original algorithm of frequent item set mining, but it is at least two times faster than the FP-growth method. Thereafter, it can be used to screen and preprocess other agent data as well as to create a database by using the improved algorithm, thus establishing a simulation system for effective component and computer-composing principles of TCM that is characterized by frequent item sets, association rules, and the clustering analysis algorithm and consequently undergoes analysis for the frequency of medicinal herbal ingredients, frequency of symptoms with certain compounds, medicine-sickness connection, and pharmaceutical composition clustering to determine the related effective ingredients to treat disorders of the spleen and stomach.

Introduction

A medicinal herbal prescription is an important method to treat conditions by traditional Chinese medicine (TCM). Exploring the effective ingredients of a medicinal herbal prescription will help to clarify the scientific nature of Chinese medicine prescriptions. A TCM prescription consists of a combination of various medicinal herbs following the principle of compatibility, that is, according to the disease needs and medicinal herbal characteristics, two or more combinations of medicinal herbs are selected. Based on the rich clinical experience in the past dynasties, the principles of prescription compatibility such as "king and minister" and "seven emotions and harmony" were obtained in order to reduce the toxicity of medicinal herbs and to enhance their therapeutic effect. Numerous clinical medication experiences and facts have proven the compatibility relationship between TCM prescriptions and the effective ingredients.

In a TCM prescription, there are as few as two or as many as dozens of medicinal herbs, including hundreds of ingredients. Due to the compatibility principle, there are very complex interactions between the ingredients. From the chemical point of view, mutual reactions widely occur between the ingredients in the prescription, which may inhibit or catalyze each other. It is difficult to deal with such a complex reaction system by conventional analysis means; therefore, based on the previous research data of TCM prescriptions, we introduce a method of computer-aided data analysis to explore correlations between prescription ingredients and disease symptoms.

Keywords: Traditional Chinese medicine; Compound; Data mining; Frequent; Item set; Association rules; Compatibility relation.

Abbreviations: FP, frequent pattern; TCM, traditional Chinese medicine.

[*]**Correspondence to:** Jian-She Yang, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China. ORCID: https://orcid.org/0000-0001-7069-6072. Tel: +86-21-66302721, E-mail: yangjs@impcas.ac.cn

How to cite this article: Yang J-S, Wang Q, Zhou SY, Zhang FF, Lv ZW. Construction of Chinese Medicine Prescription Data Mining Based on Association Rules of the System. *Explor Res Hypothesis Med* 2023;000(000):000–000. doi: 10.14218/ERHM.2023.00009.

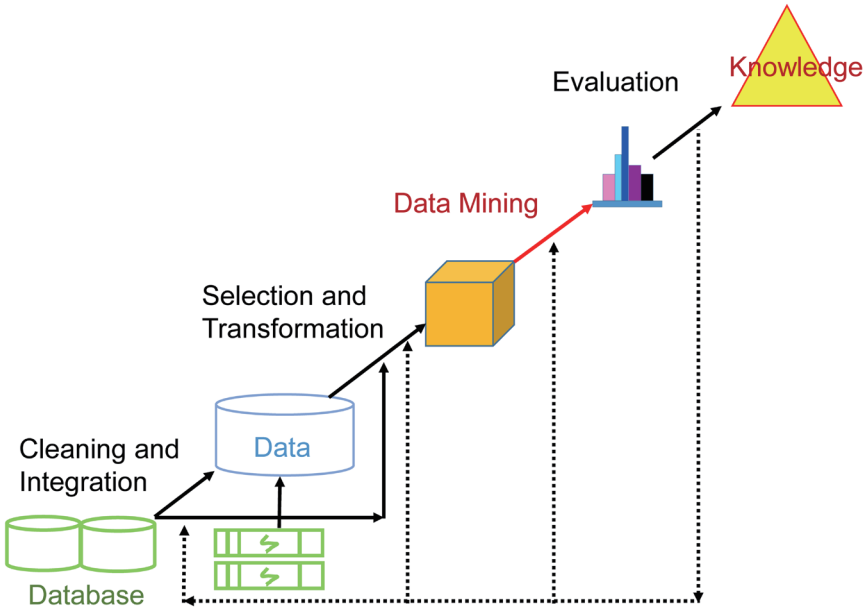

Fig. 1. Process of data mining.

Data mining technology

The advantage of computer technology lies in data processing. With the advancement of mathematics, statistics, and other methods in the field of computer science, big data processing and database technology have developed rapidly. In addition to the traditional database query function, database mining technology has arisen in order to summarize the implicit information and knowledge behind the big data.

Concept of data mining

Data mining, also called "database knowledge discovery," is a complex process of extracting and mining unknown and valuable patterns or regular knowledge from a large amount of data. Simply put, it is extracting or "mining" knowledge from a large amount of data. The cross-combination of database technology, statistical principles, pattern recognition, artificial neural networks, visualization technology, and other multidisciplinary technologies extracts useful knowledge, rules, or high-level information from massive data to provide support services for various work. The data mining process generally consists of three main stages: data preparation, data mining, and result expression and interpretation. The process of knowledge discovery consists of the following seven steps, which are described in Figure 1:

1. Data cleaning: Eliminating data interference noise or wrong data in the data source;
2. Data integration: Combining a variety of different types of data;
3. Data selection: Selecting the data related to the analytical purpose in the source database;
4. Data transformation: The data are transformed into a form suitable for mining and analytical algorithms;
5. Data mining: Using statistical algorithms and other artificial intelligence technologies to determine the hidden law patterns in the data;
6. Pattern assessment: Identifying truly interesting patterns that provide knowledge based on a certain measure of interest;
7. Knowledge representation: A visual knowledge representation interface that provides users with mining knowledge.

Technology of data mining

There are many kinds of data mining methods, such as classification, estimation, prediction, association rules, clustering, coarse set, complex data type mining, etc. The association rule is one of the most commonly used and mature methods in the field of data mining. The core algorithm adopted in this study was the association rule. The application of the association rule is mainly to mine the association between medicinal herbs and symptoms, medicinal herbs with each other, and medicinal herbal ingredients and symptoms. Association refers to the interdependence between the data. The specific definition is as follows: Suppose I = {I1, I2, …, IM} is a set of items. Given a database D, where each transaction is a nonempty set of I, i.e., each item corresponds to a unique identifier TID. The support of the association rule in D is the percentage of items in D containing both IX and IY, namely the probability; the confidence is the percentage of items in D already containing X and Y, i.e., the conditional probability. At this point, the association rule is considered strong if a minimum support and confidence value is predetermined and when the minimum support threshold and minimum confidence threshold are met.

Application of computer technology and data mining in the pharmaceutical field

Identification of Chinese medicinal materials

China is rich in TCM resources, but the evaluation method for quality standards is an urgent problem to be solved in the TCM industry. Image quantitative analysis technology has been used to establish a microscopic identification pattern identification system for *Coptis chinensis* tissue cells, which provide a new three-dimensional quantitative research technology and visual data for the identification of authentic medicinal herbs. In addition, the triangular surface element and nonuniform rational basis spline surface reconstruction algorithm as well as the software-generated interrupt three-dimensional real-time graphics card realize the three-dimensional reconstruction and dynamic display of Chinese

medicinal herbs.[1,2] *Asarum heterotropoides Fr. Schmidt var. mandshuricum (Maxim.) Kitag*, *Asarum sieboldi Miq. var. seoulense Nakai*, and *Asarum sieboldii Miq.*, three *Asarum* genera, were identified. A total of 26 samples were selected as the training set, 19 *Asarum* samples were used as the test set for data mining, and quantitative classification features were obtained from *Asarum* essential oil using gas chromatography-mass spectrometry analysis. The results are consistent with those reported previously.[3]

However, the quality of TCM is determined by the type and content of its chemical components. The qualitative and quantitative analyses of only a few of the active ingredients cannot fully reflect the quality difference of TCM due to the overall effect of the TCM theoretical system, the synergistic effect of the TCM, and the compatibility relationship of the medicinal flavor. Regarding the quantitative study of the prescription compatibility principle, fuzzy mathematics quantitative tools are useful, and the integrated use of clustering analysis, pattern recognition technology, and statistical analytical methods with TCM prescriptions split and analyze the interactions between medicinal herbs in order to find the best compatibility relationship and dose, and then clarify the prescription.

Pattern recognition technology[4] is becoming one of the most scientific and effective methods for TCM quality evaluation and TCM variety classification. Su *et al*.[5] took the content of 0 macro and trace elements of 20 species as the classification characteristics and used the nonlinear reflection method in pattern identification to identify and classify 78 samples of broadleaf holly leaf, and the results were consistent with the actual situation. Moreover, Guo *et al*.[6] used 3H-geniside as the tracer to observe the quantitative distribution of geniposide in mice, and they discussed the relationship between this change and the theory of *gardenia*. The results show that the distribution characteristics of the same organ are basically consistent with the relationship between *gardenia* and the viscera, thus providing a morphological basis for the traditional theory of *gardenia*. Furthermore, a quantitative method of finding alternatives for medicinal materials has been established: quantify the nature, taste, and meridian of TCM, and investigate the similarity between medicinal materials in order to find alternative medicinal materials.[7] Additionally, through atomic absorption spectroscopy of 10 kinds of Xinwen Jiebiao medicinal herbs and 7 kinds of Wenli medicinal herbs, the content of 15 trace elements was determined, and the relationship between their efficacy and content was analyzed. The results showed that the efficacy of these two kinds of medicinal herbs was related to the contents of manganese, barium, and other elements, and the discrimination model of these two kinds of medicinal herbs was successfully established.[8]

Application of computer technology and data mining in TCM and a prescription database

Through the computer retrieval system of Atlas Classics and Materia Medica developed by Nanjing University of Chinese Medicine, readers can browse all of the original text of Materia Medica or the original text of a certain Chinese medicine. Based on National Chinese Herbal Medicine research, a large database was established that comprises 13,268 Chinese herbal medicine records (772 families), including 11,471 plant medicines, 1,634 animal medicines, and 163 mineral medicines. Each record contains information about the class, Latin scientific name, plant-animal-mineral, medicine, literature, location, efficacy, and other basic information.[9] The "Chinese Herbal Medicine Database Retrieval System," established by the Fujian College of TCM, adopts the mode of block design and uses the dictionary library structure to realize the intensive management of TCM special terms; in addition, it provides modern retrieval tools for scientific research, clinical practice, and teaching of Chinese medicinal herbs.[10] The "TCM Prescription Coding and Literature Database System," developed by the Nanjing University of Chinese Medicine, includes 101,903 ancient and modern prescriptions; this database can be used to search the name of the prescription, title, prescription medicinal herbs, functions, indications, etc. There are more than 40 large-scale TCM databases, including about 1.1 million pieces of information, such as TCM journal literature databases, disease diagnosis and treatment databases, various kinds of TCM prescription databases, ethnic medicine databases, all kinds of pharmaceutical enterprises, and a national standard database. This system can realize the single-library and multi-library selection query. Moreover, the "Chinese Medicine Commonly Used Prescription Database Retrieval System" includes Chinese medicines as well as the retrieval prescription, author, medicine, function, indications, pharmacological effect, and usage. Furthermore, it can be used to investigate the evolution of the prescription and compare the compatibility of the prescription, according to the function and pharmacological query of the corresponding prescription, etc. Previous study[11] used the FoxBase + database system and the Universal Chinese Disk Operation System as the Chinese character support environment, and they developed an ancient Chinese medicine management and analysis system that can provide information regarding the medicinal herbal ingredients and other data according to various situations such as dynasty, disease name, and disease certificate. Through the analysis of the medical case prescriptions,[12] including 416 prescriptions and 465 types of medicinal material, 23 kinds of "core prescriptions" were found, among which Siwu soup, Liujunzi soup, and Buzhongyiqi soup were given priority, in addition to "core medicinal herbs" such as *licorice*, *ginseng*, *bighead atractylodes rhizome*, *angelica*, and *poria cocos* as well as 13 kinds of medicine. These results are consistent with clinical medications and experience. Computer technology also plays a role in the expiration date of medicinal herbs. Chen *et al*.[13] used ultraviolet spectrophotometry and HPLC-DAD-MS/MS to determine the absorption of silver yellow injection, predict the stability of chlorogenic acid and baicalin, and then imported the experimental data to the computer for calculation to predict the expiration date of silver yellow injection. Another research team applied TCM medical record management platform and SAS statistical software to analyze the cases of a professor in the treatment of type 2 diabetes, in order to explore the medication rules for the treatment of type 2 diabetes, so as to enrich and optimize the diagnosis and treatment plan for type 2 diabetes based on the experience of famous doctors.[14] Therefore, it can be seen that computer technology is playing an increasingly important role in the study of TCM, and data mining technology provides a new and effective method for the objectification and standardized research of TCM. Combined with the TCM database, through the statistical analysis of the medicinal herbs in the top 20 dynasties, the changed rules of ancient and modern medicinal herbs were analyzed by data mining.[15] Li *et al*. used the cluster analysis method to statistically analyze the medication rules of Banxiaxiexin soup, discussed the distribution and characteristics of clinical cases, and concluded that there are four main combinations of medicinal herbs for the treatment of digestive diseases and that their medication characteristics can reflect the treatment strategy of TCM syndrome differentiation.[16] A total of 43 prescriptions were collected, and the data mining and analysis of the correlation rules and frequency analysis were adopted. It was found that 24 commonly used medicinal herbs mainly cure heat and dampness, guide Qi stagnation, and reconcile Qi and blood; these findings were in line with the common clinical rules for the treatment of

dysentery in ancient and modern times.[17] The prescriptions used for the treatment of thirst elimination were collected, and certain scientific compatibilities between single medicines and prescriptions were found through data mining technology; the conclusion was basically consistent with the principle of syndrome differentiation and treatment.[18] Similarly, Dai *et al*. referred to the Dictionary of Traditional Chinese Medicine, selected 1,355 prescriptions of the spleen and stomach, and used data mining methods such as cluster analysis, corresponding analysis, and frequent collection methods to statistically analyze commonly used medicinal herbs and closely related cluster parties. They concluded that the basic prescription for replenishing the spleen and Qi is represented by the soup. In recent years, with the development of computer technology, molecular biology knowledge, and combined with data analysis, the diagnosis of spleen deficiency in TCM can be made more reasonable and systematic. For example, the assumption-based truth maintenance system (ATMS) artificial intelligence method has been applied to the mining research of clinical data of spleen deficiency in TCM.[19] In addition, Cao *et al*.[20] used data mining technology to calculate the contribution rate of the syndrome and syndrome group to diagnose spleen deficiency syndrome, and they recorded 1,564 cases to establish a mathematical model. Moreover, another study[21] analyzed the gene expression profiles of the samples from patients with spleen deficiency, and they found type 2 diabetes mellitus-spleen deficiency pattern patients experienced significant hypoimmunity and/or immune dysfunctions, and possessed a specific gene expression profile, therefore they concluded that the gene expression profile was helpful for the differentiation and diagnosis of spleen deficiency syndrome. In the feature extraction stage, the Wilcoxon signed-rank test as well as the between-group and within-group sum of square ratio, respectively, were used. Furthermore, Liu *et al*.[22] used data from 324 samples as the industry assessment samples and data from 99 new samples collected as the external assessment samples, and the degree of agreement between the analysis experts and computer simulation expert diagnostic procedures was assessed to evaluate the practical application effect of the measurement and diagnostic methods of the dialectical scale of spleen and stomach diseases. The positive predictive value was 77.2%, 93%, the total compliance rate was 88%, the compliance rate of the main certificate was 93.8%, and the compliance rate of the concurrent certificate was 79.7%. The positive predictive value of the external assessment was 100%, the real predicted value was 85.5%, the total coincidence rate of the false certificate was 87.9%, the coincidence rate of the main certificate was 90.9%, and the coincidence rate of the concurrent certificate was 73.8%. These results show that the computer measurement method of spleen and stomach diseases has a good diagnostic effect. Tang *et al*.[23] used Bayesian discriminant analysis and a simplified scoring method to establish a differential diagnostic system of spleen and stomach Yin deficiency. According to the 13 main symptoms, the frequency of symptoms, and the contribution rate to the diagnosis, clinical research was conducted. In the systematic identification results, the contribution rate of abdominal distension, unregulated stool, epigastric discomfort, and hunger was >10%, which was considered to have great differential diagnostic significance; while the contribution rate of troublesome fever, fatigue, and thirst was <10%, which had a relatively little significance on the differential diagnosis. The total contribution rate of the differential diagnosis of all 13 symptoms was 92%, and the sensitivity and specificity of 25 cases of Yin deficiency of the spleen and stomach exceeded 90%, thus demonstrating the differential diagnosis by TCM. Jiang *et al*. first preprocessed the original data (1,355 spleen and stomach prescriptions) to standardize, structure, and digitize the prescription data. Then, according to the data characteristics of the prescription, cluster analysis, correspondence analysis, and frequent set methods were selected for quantitative analysis. The results show that through the data mining of the core medicinal herbs, medicinal herbs with each other, and the "prescription syndrome" and the formula structure, the correlation results are basically consistent with the general rules and characteristics of TCM spleen and stomach prescriptions.[24] Zha *et al*. analyzed 292 clinical studies on TCM treatment of chronic gastritis. After information digitization processing, a total of 28 symptoms and a lingual pulse with a frequency of more than nine times were counted. These symptoms and lingual pulse were clustered, and 28 symptoms and lingual pulse were clustered into three categories. The first category included hiccups, fatigue, loss of appetite, and other symptoms, according to the theory of TCM, which can be judged as spleen and stomach Qi deficiency; the second category included dry mouth, epigastric pain, constipation, and other symptoms, which can differentiate between the liver and the stomach; and the third category included abdominal pain, abdominal pain, acid swallowing, abdominal distension, vomiting, etc.; such symptoms can be liver depression and fire, spleen, and stomach. These statistical results are basically consistent with the actual clinical symptoms.[25] More than 600 clinical medical cases collected by Li *et al*. were standardized and processed, and the data used association analysis and the frequent pattern (FP)-tree algorithm to mine the association between symptoms and prescriptions, symptoms and syndrome, and medicinal herbs and syndrome. The results showed 151 association rules between symptoms and prescriptions, 116 association rules between symptoms and syndrome types, and 144 association rules between medicinal herbs and syndrome types.[26] A new method for inheriting the academic thoughts and clinical experience of famous veteran doctors with the help of artificial intelligence technology was explored and a prototype system framework was proposed based on rules and deep learning models, which was demonstrated as a feasible way.[27] A decision tree method based on information entropy was applied to explore TCM syndrome differentiation of chronic gastritis. Using the bootstrap method to amplify 406 cases, the decision tree algorithm C4.5 was used to determine the coincidence rate of model classification, and the results were as follows: 83.60% for the training set, 80.67%, and 81.25% for the test set. The sensitivity and specificity of the model were also high, which can be applied to the differential diagnosis of TCM syndrome forms of chronic gastritis.[28] The symptoms of patients were grouped according to the stomach status, abdominal status, diet, excretion, tongue diagnosis, etc. First, the hierarchical clustering method was used to group symptoms, and the principal components of the symptoms in each group were used as the input to learn the Bayes network in order to analyze the symptoms after grouping. With 2,021 medical cases, the identification of chronic gastritis also achieved a high accuracy. In addition, according to the learning method of the Bayes network under incomplete data, an improved structural equation modeling algorithm combining simulated annealing and the Bayes classifier (BC) algorithm was proposed,where simulated annealing was used for structure selection and the BC algorithm for initial parameter estimation was able to improve the learning ability of Bayes networks with such data.[29]

At present, the data mining techniques used for chronic gastritis mainly include association rule mining, cluster mining, the decision tree algorithm, factor analysis, etc. These methods have expanded the methodology to TCM clinical diagnosis and treat-

ment information mining as well as provide great technical support for the inheritance and mining of TCM diagnosis and treatment experience.[30]

Currently, domestic data mining technology analysis research is still in its infancy. For researchers in the field of medicine, many data processing software programs (such as Weka, B Miner, SPSS Clementine, SAS Enterprise Miner, etc.) contain the function of commonly used data mining methods. As the researchers' understanding of "data mining" and its applications increases, these novel data analysis tools will have a positive role in promoting medical research.

Data mining system of TCM prescriptions

The Chinese medicine prescription data mining system database is called the Chinese medicine prescription database, and the construction of the system is composed of two parts: First, it establishes the Chinese medicine prescription database; and then it builds the Chinese medicine prescription system analysis module. Mining was done with the mining analysis module of the Chinese medicine prescription database. With the help of the data mining technology of the related analytical algorithm, computer automatic data resources, analysis, and mining hidden-related rules, the prescription analytical results were obtained.

Establishment of the TCM prescription database

Data mining is built on the basis of a large number of high-quality data information, the amount of data, and the integrity and authenticity of the quality standards, which are directly related to the knowledge as well as the accuracy and effectiveness of the conclusion. Therefore, the establishment of the Chinese medicine prescription database is based on the whole Chinese medicine prescription analytical system, and the data in the database must be complete, true, and effective.

Data preparation

In order to ensure the scientific accuracy of the prescription data, we selected the data sources of the prescription database, including the 2010 Chinese Pharmacopoeia (seventh edition) and the Summary of Chinese Chamber, as the main sources of the prescriptions. Although there is a lot of relevant scientific research literature, due to the fact that the study of Chinese medicine is still in the exploratory stage, most of the research results of Chinese medicine composition have not yet formed a unified standard. Therefore, we first followed the 2010 edition of the Pharmacopoeia of the People's Republic of China for the data selection of Chinese medicine ingredients, then chose the Pharmacopoeia of Chinese medicine extract, and screened some of the relevant Chinese medicine composition research studies of major ingredients by searching the literature through the China National Knowledge Infrastructure database.

Data cleaning

Due to the development and inheritance of thousands of years of TCM, the wide application area, and the fact that the literature records are mostly influenced by regional dialects and the language expression mode of the past dynasties, ancient Chinese medicine literature often has many meanings and synonymous words. These phenomena have led to a large amount of information with ambiguous word meanings, inconsistent records, missing content, data noise, and redundancy during data screening. Therefore, in order to ensure that the database records are consistent, clear, complete, and effective, the collected TCM data information needs to be pretreated before establishing the database, and the key data of the Chinese medicine prescription need to be cleaned and filtered so that the information is suitable for database construction and mining in a standardized form in order to ensure the effectiveness of the latest database mining results. Data cleaning first standardizes the data source data as well as standardizes and unifies the prescription names, symptom terms, efficacy records, and medicinal herbal names, mainly according to the research needs. The process is described in following sections.

Standardization of prescription names

Since TCM prescriptions are the accumulation of clinical practice experience of TCM workers in the past dynasties, the inherited ones are basically recorded in ancient documents, and the data types belong to the form of text, coupled with the different habits of human description. As a result, there may be different names for the same prescription. Therefore, we uniformly selected the names recorded in the Pharmacopoeia of the People's Republic of China as the standard name. If it is not included in the Pharmacopoeia, the one with complete significance and the most concise text is used as the standard record.

Standardization of medicinal herbal names

Chinese herbal medicine is influenced by different varieties, origins, and climate factors and there are different qualities, so the theory of TCM must pay attention to authentic medicines. For example, many medicines have a similar "Sichuan," "cloud," "wide," or "dian" origin abbreviation, plus the origin, dialect text description, and Chinese herbal medicine phenomenon.

Standardization of symptoms and efficacy

In terms of symptom description, there are also a large number of semantic fuzzy repeated language descriptions, so they are unified using specific subject words to define them.

Data conversion

Data conversion is the TCM data format operation, TCM data are provided as text type of data, and the computer needs to identify and transform text data processing, without wasting operation resources, in order to facilitate the computer in the mining operation identification and statistical data. In addition, the computer needs to convert text type data into numerical type data, digitize the data to the text data, and make various types of data conversions in the mathematical analysis of data types. Considering the different degree gradients, we took the normal body temperature "flat" as the origin, divided the four characteristics into eight gradient indicators, and digitized them, respectively. Moreover, we divided the five flavors into seven gradients.

Furthermore, in order to facilitate the program identification and operation, the toxicity, meridian, efficacy, and symptoms of TCM were digitized accordingly, and a unified digital coding and identification format was set.

Database establishment

According to Microsoft Office Access database management software, the whole database consists of the following parts: Chinese medicine prescription, essential medicine, medicinal herbal composition, and basic symptoms. At the same time, the data table is used to identify the data in the table. The coding adopts a five-digit code, which is divided into two parts: prefix and suffix. The prefix consists of two digits indicating the data category, and the suffix

represents the data serial number, which includes the following important parts.

The TCM prescription table is the core of the whole database, which inputs the basic information of the TCM prescription, such as the prescription name, medicinal herbs, prescription category, dosage form, symptoms, efficacy, etc.

The essential medicine table is used to retrieve medicinal herbal information, which includes the names of the medicinal herbs, sex, five tastes, meridian, efficacy, medicinal herbs type, etc.

The basic symptoms table contains standard symptoms as well as similar symptoms.

The medicinal herbal composition table mainly records the clear ingredients of the TCM extracts, and this part of the data is included in the pharmacopoeia and reported in the relevant research literature. It is mainly composed of compound components and extracted medicinal herbs.

Construction of the TCM prescription mining system

Introduction to the system

The Chinese medicine prescription mining system is based on the analysis of the Chinese medicine prescription database system, through the previous data preparation. Computer data mining technology was used to build the mining of Chinese medicine prescription module, which was used to analyze the data of the Chinese medicine prescription database, in order to find useful knowledge and information. The system was developed using Visual C++ 6.0.

The system interface mainly consists of three modules: medicinal herbal analysis module, symptom analysis module, and square-like analysis module. The medicinal herbal analysis module functions according to the input medicinal herbal name, retrieved from the Chinese medicine prescription database, then the prescription is found, using the data mining algorithm to analyze the prescription data, and the input medicinal herb correlation and high confidence are determined. Additionally, the prescription of the medicinal herbs, medicinal herb correlation analysis, component correlation analysis, etc. are performed. The symptom analysis module uses the input symptoms to screen the symptomatic prescriptions, based on the selected prescription group, using the mining algorithm to conduct medicinal herb correlation analysis, medicinal herb to medicinal herb correlation analysis, medicinal herb group correlation analysis, component correlation analysis, etc. A group of prescriptions with similar effects was selected to investigate the effective ingredients of the corresponding prescriptions by screening analysis of the spleen and stomach prescriptions.

Selection of mining rules

The data mining rules used in this system include the high-frequency item set method and the association rule method.

High-frequency term set

The high-frequency item set is a statistical method based on the frequency, specifically referring to the sample, and the high frequency of the project set, assuming it defines a sample project set (both in all items, including the sample project set percentage). When the percentage is greater than the minimum support threshold, we think the sample project set frequency is higher, and the sample project set is named as the "high-frequency item set." The statistical method of the high-frequency term set can be used for "medicinal herbal frequency analysis," "component frequency analysis," "symptom frequency analysis," and some data analysis with the same term.

Association rules

Association rules are mainly used to discover the strong associations between project sets in a large amount of data. Association rules are very mature calculation methods in data mining technology, which are specifically defined as follows:

1. Let I = {i1, i2,…, im} be the term set, where the term ik (k = 1,2,…, m) can be a medicinal herb in the prescription or an ingredient in the medicinal herbs. Let the task-related data D be a set of transactions, where each transaction T is a set of items, making $T \subseteq I$. Let A be a set of items, and $A \subseteq T$.
2. Association rules are logical implications: $A \Rightarrow B$, $A \subset I$, and $A \cap B = \Phi$. The association rule has two important attributes:
   - Support: P (A ∪ B), the probability that the set of A and B occurs in the transaction set D.
   - Confidence: P (B | A), the probability that item set B also appears simultaneously in the transaction set D of the item set A.
3. Rules artificially given the minimum support threshold and the minimum confidence threshold, while satisfying the threshold, are called strong rules. Given a transaction set D, the problem of mining association rules is to produce association rules that support and credibility are greater than the minimum support and minimum credibility given by the user, that is, the problem of generating strong rules.

The apriori algorithm and the FP-tree-frequency algorithm

The apriori algorithm uses classic mining Boolean association rules and a frequency set algorithm. Its biggest advantage is that the algorithm process is simple and intuitive, and the principle is easy to understand. The apriori algorithm mining process mainly contains two stages: the first stage involves iterative calculation, finding all high-frequency terms from the data set, and the support is not lower than the user threshold set. The second stage involves the high-frequency term in the associated rules. Among them, the mining set of all frequent items is the core of the algorithm, accounting for the majority of the total computation. The specific algorithm is as follows:

1. L1 = {large 1-itemsets};
2. for ($k = 2$; $L_{k-1} \neq \Phi$; k++), do begin
3. Ck = apriori-gen (Lk−1); / / New candidate set
4. for all transactions t ∈ D, do begin
5. C_t = subset(C_k, t); / / Candidate set included in transaction t
6. for all candidates c ∈ Ct, do
7. c.count++;
8. end
9. Lk={c ∈ Ck |c.count ≥ minsup}
10. end
11. Answer = $\cup_k L_k$.

First produce frequent 1 item set L1, then frequent 2 item set L2, until there is some r value making the calculations terminate when Lr is empty. The term set in Ck is the candidate set used to generate the frequency set, and the final frequency set Lk must be a subset of Ck. Each element in Ck is verified in the transaction database to determine whether to join Lk, and the validation process here requires multiple scans of the available database. The program code of the apriori algorithm is accessable by contacting the corresponding author. .

Due to Boolean correlation rules of the apriori algorithm, in order to generate all the frequency set, the method of recursive operation, which requires the entire database, is used until the K frequency set is an empty set, and the algorithm is stopped. A large candidate set is produced, in theory, and the candidate set scale can reach the total number of geometric multiples. These will require

Table 1. Comparison of two algorithms

Item	Resource need	Processing time	Accuracy
FP-tree-frequency algorithm	not large	short	high
Apriori algorithm	very large	long	high

a large number of computer system resources, leading to a time-consuming operation process.

Therefore, in order to overcome the defects of the apriori algorithm, the system chooses the FP-tree frequency set algorithm, and the FP-tree frequency algorithm does not produce the candidate set algorithm, according to the strategy of division and cure, using a tight data structure to store all of the information needed to find the frequent item set. The algorithm only needs two database scans; the first scan gets a one-dimensional frequent item set. The database is scanned a second time with the one-dimensional frequent item set to filter nonfrequent items in the database and to generate the FP tree.

The FP-tree-frequency algorithm is described as follows:

- Input: Things database D, minimum support minsupport.
- Output: The complete set of frequent modes.

Build the FP tree

To build the FP tree, the database was scanned, the mobile phone frequent item set F was scanned, the support degree was counted, and the support degree for F was arranged in descending order to obtain the arrangement item table L.

Next, the root node of the FP tree was created and marked as "null." For each item T in D, the following steps were executed: Select the frequent items in T and rank them in order in L. Let the ranked frequency item table be [p / P], where p is the first element and P is the table of the remaining elements. Call insert_tree ([p / P], T). The procedure was performed as follows: (1) if T has a child N makes N.itemName = p.itemName, the count of N adds 1; (2) otherwise create a new node N, set its count to 1, link to its parent node T, and link it to the node with the same itemName through the node chain structure. (3) If P is not empty, the recursive call is insert_tree ([p / P], T).

Rule mining of the FP tree

With the FP-growth (Tree, α) function, the initial call is the FP-growth (Tree, null):

If Tree contains a single path, P

Then

{Each combination of nodes in the path P (noted as β)

Production mode β ∪ α, whose support = minimum support of the node in β}

else for each αi is on Tree's head

do{Generation mode β = αi ∪ α with support support= αi.support;

Conform the conditional pattern base of β, and then construct the conditional FP tree Tree β;

if Treeβ = an empty set, then call the FP-growth (Tree β, β)}

End

FP-Tree algorithm flow:

- Step 1: Scan the transaction database, sort each item by frequency, and delete items with a frequency less than the minimum support MinSup. The frequent 1-item set was generated simultaneously.
- Step 2: For each item record, reorder the order of one frequent item, with the second and last scan;
- Step 3: Insert each record from step 2 into the FP-Tree;
- Step 4: Find the frequent items from the FP-Tree.

To sum up, by selecting the FP tree frequency set algorithm, it overcomes the shortage of large resources and the long processing time of the apriori algorithm system, which finally constitutes the core of the TCM prescription data mining system, and completes the construction of the whole data mining system (Table 1).

Research on the mining of spleen and stomach prescriptions and heat-clearing prescriptions

Using the constructed TCM prescription database and the TCM prescription data mining system, the data mining research was conducted on the spleen and stomach prescriptions and heat-clearing prescriptions. In addition, the core medicinal herbs, compatibility rules, and main medicinal substances of these two types of medicinal herbs were analyzed as follows:

- Step 1: Use the statistical method to calculate the medicinal herb frequency of each medicinal herb in the formula, determine the medicinal herbs with a high medicinal herb frequency for analysis, and determine the core medicinal herbs of the prescription;
- Step 2: Using the frequency rule of the item set, determine the frequency set of medicinal herb pairs composed of two medicinal herbs as well as the frequency set of medicinal herb groups composed of three medicinal herbs and four medicinal herbs, and calculate the set frequency of the TCM pairs;
- Step 3: Use the association rules to calculate the medicinal herb-medicinal herb correlation confidence, medicinal herb-symptom-related confidence, medicinal herb-medicinal herb group correlation confidence, and medicinal herb group-symptom correlation confidence as well as to analyze the medication rules of the formula;
- Step 4: Statistically analyze the occurrence frequency of medicinal herbal ingredients in the formula, determine the frequent item set of the ingredient group, and analyze the core component group of the formula;
- Step 5: Calculate the component-symptom correlation confidence through the association rules and analyze the symptomatic prescription-like medicinal substances.

Mining and pharmaceutical analysis of the spleen and stomach prescriptions

Using the database and data mining system constructed in this study, the spleen and stomach medicinal herbs were mined and analyzed, and the core medicinal herbs of the spleen and stomach prescriptions were determined. Through the medicinal herb analysis, the rules of the spleen and stomach prescriptions were studied, and the medicinal substances of the spleen and stomach prescriptions were studied through the symptomatic analysis of component groups. Through the screening database, 581 records of spleen and stomach prescriptions and 355 samples of medicinal herbs were obtained.

Analysis of medicinal herb frequency

Through the mining system, 355 spleen and stomach prescriptions were counted, the medicinal herbs with high medication frequency

Table 2. Frequency analysis of medicinal herbs

Number	Name	Frequency	Support degree	Category
1	*Ginseng*	192	33.08%	Buqi
2	*Bighead atractylodes rhizome*	182	31.36%	Buqi
3	*Orange peel*	178	30.69%	Liqi
4	*Prepared Radix Glycyrrhizae*	174	29.93%	Buqi
5	*Tuckahoe*	162	27.83%	Lishui
6	*Pinellia ternata*	110	18.93%	Huatan
7	*Chinese herbaceous peony*	106	18.26%	Buxue
8	*Ginger*	97	16.73%	Fasan
9	*Radices saussureae*	91	15.59%	Liqi
10	*Mangnolia officinalis*	83	14.34%	Huashi
11	*Rhizoma coptidis*	71	12.14%	Qingre
12	*Cinnamomum cassia*	70	12.05%	Wenli
13	*Astragalus mongholicus*	64	11.00%	Buqi
14	*Fructus amomi*	61	10.42%	Huashi
15	*Rhizoma zingiberis*	59	10.23%	Wenli
16	*Ligusticum wallichii*	58	9.94%	Huoxue
17	*Scutellaria baicalensis*	55	9.47%	Qingre
18	*Radix Rehmanniae Praeparata*	55	9.47%	Buxue
19	*Rhizoma Dioscoreae*	52	8.99%	Buqi
20	*Radix sileris*	51	8.80%	Fasan
21	*Rhizoma atractylodis*	51	8.70%	Huashi
22	*Baked ginger*	50	8.61%	Wenli
23	*Radix aconiti carmichaeli*	47	8.03%	Wenli
24	*Radix Rehmanniae Recens*	46	7.94%	Qingre
25	*Radix bupleuri*	45	7.75%	Fasan
26	*Fructus Ziziphi Jujubae*	43	7.46%	Buqi
27	*Fructus Gardeniae*	42	7.17%	Qingre
28	*Rhizoma cimicifugae*	42	7.27%	Fasan
29	*Medicated leaven*	40	6.88%	Xiaoshi

were identified, the medication trend of the spleen and stomach prescriptions was analyzed, the core medicinal herbs of the spleen and stomach prescriptions were determined, the medicinal herbs with an occurrence frequency of 40 were taken as the threshold, and the research support degree was more than 5%. The core medicinal herbs for the use of spleen and stomach prescriptions were mainly focused on Qi tonic, Qi management, blood tonic, and warm medicinal herbs, supplemented by heat-clearing and blood circulation medicinal herbs (Table 2).

Analysis of medicinal herb pairs in medicinal herb groups

The medicinal herb pair is mainly a commonly used medicinal herb combination composed of medicinal herbs. Some known and unknown medicinal herb pairs can be found by statistical analysis of frequent item sets, which is of positive significance for the study of the prescription rules of spleen and stomach prescriptions. The minimum threshold value of the medicinal herb pair use frequency was set to 70 times, and the minimum threshold value of the support degree was 10%. The mining system mined 10 medicinal herb pairs that met the support degree threshold (Table 3).

Through the analysis of spleen and stomach prescriptions, it was determined that the most frequent medicinal herbs are ginseng-atractylodes, atractylodes-Poria, and ginseng-zhi licorice, which are typical medicine pairs of TCM, such as Atractylodes-Poria prescription of TCM, which treat the symptoms of spleen deficiency, dampness-heat, diarrhea, and so on.

The support threshold of the medicinal herb group composed of the three medicinal herbs was set to be 8%, and the mining system mined eight medicinal herb groups that met the support threshold. The support threshold of the medicinal herb group composed of four medicinal herbs was set to be 8%, and the mining system found eight medicinal herb groups that met the support threshold.

Table 3. Analysis of medicinal herb pairs

Number	Medicinal herb pairs	Frequency	Support degree
1	*Ginseng*:*Bighead atractylodes rhizome*	129	22.20%
2	*Bighead atractylodes rhizome*: *Tuckahoe*	118	20.31%
3	*Ginseng*:*Prepared Radix Glycyrrhizae*	106	18.24%
4	*Bighead atractylodes rhizome*:*Prepared Radix Glycyrrhizae*	99	17.04%
5	*Ginseng*:*Tuckahoe*	95	16.35%
6	*Bighead atractylodes rhizom*:*Orange peel*	89	15.39%
7	*Orange peel*:*Prepared Radix Glycyrrhizae*	85	14.63%
8	*Orange peel*:*Tuckahoe*	80	13.77%
9	*Ginseng*:*Orange peel*	77	13.25%
10	*Prepared Radix Glycyrrhizae*:*Tuckahoe*	75	12.91%

According to the mining system operation, the development of the prescription rules of spleen and stomach medicinal herbs is based on the treatment of the main diseases of the spleen and stomach. Around these four medicinal herbs, the treatment of other complications is added to form a series of spleen and stomach prescriptions.

Component group analysis

This system has a database of medicinal herbal ingredients. By screening the chemical components of the corresponding medicinal materials in the database and mining the chemical components of the prescription medicinal herbs in the spleen and stomach prescriptions, the frequency and support of these related components were obtained.

Through the spleen and stomach prescriptions of TCM chemical data mining, it can be seen that the spleen and stomach prescriptions contain the highest frequency of the following chemicals: ginsenosides, pale alcohol, licorice sugar, glycyrrhizic acid, poria cocos glycan, lemon, etc., which are commonly used in the spleen and stomach prescriptions, and the main therapeutic effect should be reflected by the pharmacological effect of these medicinal herbal ingredients. Since the pharmaceutical chemical composition depends on the medicinal herb composition of the prescription compatibility, there is a strong consistent relationship between the composition group and the association between the prescription medicinal herbs (Table 4).

Analysis of the association between spleen and stomach prescriptions and symptoms

The spleen and stomach prescriptions were mainly used to treat vomiting, anorexia, abdominal pain, chest tightness, diarrhea, abdominal distension, cough, upset, poor, thin, thirst, fatigue, etc. Through the analysis of the corresponding symptoms with medicinal herbs and the medicinal herb group by correlation analysis, one can determine the strongest association with symptoms. The correlation analysis of symptoms and medicinal herbal components allows the treatment of the corresponding symptoms.

Symptom-medicinal herb association analysis

Using the correlation between a single symptom and the association rule whose support is greater than 8%, it can be seen that vomiting, diarrhea, and anorexia reached 15.32%, 14.63%, and 13.39%, respectively. TCM theory holds that ginseng has the effect of treating all deficiency diseases, nausea, vomiting, and diarrhea. Atractylodes is used for spleen deficiency, abdominal distension, diarrhea, phlegm, edema, sweating, fetal movement, and other symptoms. The results of this system on the correlation analysis between medicinal herbs and symptoms are consistent with the TCM theory (Table 5).

Analysis of symptomatic medicinal herbal groups

The association rule was used for the association between a single symptom, and the association rule was more than 10%. Using this mining system, with vomiting and diarrhea symptoms as the goal, the spleen and stomach prescription symptom-medicinal herbal group association analysis showed the association strength to be more than 10%, which can be seen from the analytical results. Vomiting symptoms had the highest correlation strength with ginseng, white, and licorice group (14.21%), followed by orange

Table 4. Chemical component analysis of three medicinal herb groups

Number	Components of medicinal herb groups	Frequency	Support degree
1	*Ginsenoside:Atractylol:Cajuputene*	72	12.39%
2	*Ginsenoside:Atractylol:Glycyrrhizin*	70	12.05%
3	*Ginsenoside:Atractylol:Pachyman*	69	11.88%
4	*Ginsenoside:Enoxolone:Pachyman*	67	11.53%
5	*Atractylol:Cajuputene:Pachyman*	61	10.50%
6	*Atractylol:Glycyrrhizin:Pachyman*	59	10.15%
7	*Atractylol:Cajuputene:Glycyrrhizin*	55	9.74%

Table 5. Association analysis of symptoms and medicinal herb frequency

Number	Symptoms and medicinal herb association	Frequency	Support degree
1	*Ginsen*⇔vomit	65	11.19%
2	*Ginsen*⇔diarrhea	62	10.67%
3	*Orange peel*⇔vomit	60	10.33%
4	*Tuckahoe*⇔vomit	59	10.15%
5	*Bighead atractylodes rhizome*⇔vomit	57	9.81%
6	*Bighead atractylodes rhizome*⇔diarrhea	55	9.47%
7	*Ginsen*⇔apocleisis	52	8.95%
8	*Orange peel*⇔apocleisis	49	8.43%

peel, hot licorice, and ginger medicine group (13.16%). From the diarrhea symptoms associated with the strongest medicinal herb group analysis, diarrhea can be treated by ginseng, atractylodes, red licorice, angelica, tuckahoe, and their combination. For the treatment of diarrhea and vomiting symptoms, which have mutual synergy and a synergistic reduction-related effect, these herbs are consistent with the commonly used spleen and stomach prescriptions. Therefore, using the symptoms and medicinal herbal group correlation analysis can provide a reference basis for the compatibility of clinical medicinal herbs (Table 6).

Discussion

Through the mining system constructed in this paper, the data mining research of spleen and stomach prescriptions and heat-clearing prescriptions was conducted. First, the frequent set principal analysis was performed, and the spleen and stomach prescription medicinal herbal type was determined through medicinal herbal group association rule analysis. Next, the core of the spleen and stomach prescription as well as the common medicinal herbs were determined and analyzed by medicinal herb association studies with symptoms. The corresponding vomiting and diarrhea symptoms of the spleen and stomach medicinal herbal group compatibility were then summarized. Using the medicinal herbal composition database corresponding to vomiting symptoms and diarrhea symptoms, the medicinal herbal component group with strong association rules with symptoms was determined. Finally, it was found that there may be effective substances to treat the symptoms of vomiting and diarrhea.

Future directions

There are still some deficiencies in this study that must be mentioned. For example, the system did not involve the mining of the dose-effect relationship, there was a lack of statistical data of the medicinal herb dosage, and it was difficult to analyze the symptomatic principle of the prescription at different medicinal herb doses. In addition, the dependence of data mining on the database determined that the results of mining cannot go beyond the limitations of the data sources. Due to the fact that the TCM data are mostly recorded in ancient writings, despite the data noise cleaning and standardization sorting, some noise interference still remained in the data, such as foreign bodies of the same name of ancient and modern herbs. Moreover, the complexities of pharmaceutical composition restrict the deeper data mining analysis. In future work, the data mining of prescriptions should be more in-depth by increasing the analysis and research of the correlation law of TCM. The study of negative association rules can also be carried out to solve the taboo problem of TCM compatibility. At the same time, data mining technology is constantly developing and progressing, and many new data mining algorithms, such as neural network technology and decision tree technology, can be applied to lead to new discoveries.

Table 6. Association analysis of symptoms and medicinal herb frequency

Number		Symptoms and medicinal herb group association	Frequency	Support degree
1	*Diarrhea*	*Ginseng:Bighead atractylodes rhizome:Prepared Radix Glycyrrhizae*	29	16.20%
2		*Ginseng:Bighead atractylodes rhizome:Tuckahoe*	26	14.53%
3		*Ginseng:Bighead atractylodes rhizome:Angelica sinensis*	25	13.97%
4		*Bighead atractylodes rhizome:Tuckahoe: Radices saussureae*	23	12.85%
5		*Bighead atractylodes rhizome:Orange peel: Radices saussureae*	21	11.73%
6	vomit	*Ginseng:Bighead atractylodes rhizome:Prepared Radix Glycyrrhizae*	27	14.21%
7		*Orange peel:Prepared Radix Glycyrrhizae:Ginger*	25	13.16%
8		*Ginseng:Prepared Radix Glycyrrhizae: Tuckahoe*	24	12.63%
9		*Orange peel: Mangnolia officinalis:Rhizome atractylodis*	23	12.11%
10		*Prepared Radix Glycyrrhizae: Tuckahoe:Pinellia ternata:Ginger*	22	11.58%
11		*Ginseng:Tuckahoe: Pinellia ternata: Orange peel:Tuckahoe:Ginger*	21	11.05%

Conclusions

In this study, by using Microsoft Office Access database management software, the TCM prescription database was established. Based on the data mining technology, the TCM prescription mining system was developed. For the first time, the medicinal herbal composition data were entered into the prescription database system, and the correlation between the symptoms and the effective component group of the prescription for treating the corresponding symptoms was analyzed.

Through the mining experiments on the spleen and stomach prescriptions, we determined that the core and common medicinal herb group and pairs of spleen and stomach prescriptions mainly include Yiqi and Liqi medicinal herbs, Fasan medicinal herbs, Huashi medicinal herbs, and Wenli Huoxue medicinal herbs, such as ginseng, atractylodes, poria cocos, angelica, red licorice, etc. We also analyzed the prescription compatibility rules of the spleen and stomach prescriptions. At the same time, through disease-medicinal herb association analysis, disease-medicinal herb group association analysis, and disease-medicinal herb composition group association analysis of vomiting and diarrhea, the prescription compatibility rule of the corresponding symptom was explored, and the effective component group of the prescription for the treatment of each symptom was speculated. However, due to the very complexity of the TCM system, the dataset construction and mining could never cover all of the information at present; hence, more precise algorithms need to be developed.

Acknowledgments

None.

Funding

This research was funded by the National Natural Science Foundation of China (72171170), Shanghai Municipal Health Commission (GWV-10.1-XK09), Shanghai Shenkang Center (SHDC-2020CR2054B), University Teachers Innovation Fund Project of Gansu Province (2023A-182), and the Key Research Project of Pingliang Science and Technology (PL-STK-2021A-004).

Conflict of interest

Dr. Jian-She Yang has been an editorial board member of *Exploratory Research and Hypothesis in Medicine* since March 2020. The other authors have no other conflicts of interest.

Author contributions

Conceptualization (YJS, LZW); methodology (YJS); software (WQ, ZSY, ZFF); validation (YJS, LZW); formal analysis (WQ, ZFF); investigation (WQ, ZSY, ZFF); resources (WQ, ZSY); data curation (YJS); original draft preparation (WQ, ZSY, ZFF, YIS); review and editing (YJS, LZW); visualization (YIS); supervision (YJS, LZW); project administration (YJS, LZW); funding acquisition (WQ, LZW, YJS). All authors have read and agreed to the published version of this manuscript.

Data sharing statement

The technical appendix and dataset are available from the corresponding author at yangjs@impcas.ac.cn.

References

[1] Zhao C, Cai Z. Three-dimensional quantitative mass spectrometry imaging in complex system: From subcellular to whole organism. Mass Spectrom Rev 2022;41(3):469–487. doi:10.1002/mas.21674, PMID: 33300181.
[2] Weiss TL, Zieselman A, Hill DP, Diamond SG, Shen L, Saykin AJ, *et al*. Alzheimer's Disease Neuroimaging Initiative. The role of visualization and 3-D printing in biological data mining. BioData Min 2015;8:22. doi:10.1186/s13040-015-0056-2, PMID:26246856.
[3] Su W, Wu Z, Chen J, He X, Li J. Chemical pattern recognition of traditional Chinese medicine kudingcha (I) (in Chinese). Zhong Yao Cai 1998;21(3):115–119. PMID:12567936.
[4] Ma D, Wang L, Jin Y, Gu L, Yu X, Xie X, *et al*. Application of UHPLC Fingerprints Combined with Chemical Pattern Recognition Analysis in the Differentiation of Six Rhodiola Species. Molecules 2021;26(22):6855. doi:10.3390/molecules26226855, PMID:34833946.
[5] Su W, Wu Z, He X, Chen J. Chemical pattern recognition of traditional Chinese medicine kudingcha (II) (in Chinese). Zhong Yao Cai 1998;21(4):170–173. PMID:12567945.
[6] Guo SG, Ben CN, Zhao LY, Yang MJ, Yu SY, Chen S. Quantitative Dynamic Study on the Whole-Body Autoradiography and Image Analysis of 3H-Gardenoside in Rats and the Relationship of the Channel Tropism (in Chinese). Journal of Beijing University of Traditional Chinese Medicine 1996;19(4):28–31. doi:10.3321/j.issn:1006-2157.1996.04.011.
[7] Li M, Bai L, Peng S, Sun F, Wang L, Liu H, *et al*. Simple quantitative analytical methods for the determination of alkaloids from medicinal and edible plant foods using a homemade chromatographic monolithic column. J Chromatogr B Analyt Technol Biomed Life Sci 2019;1128:121784. doi:10.1016/j.jchromb.2019.121784, PMID:31518898.
[8] Chen HL, Sun LC. Research on the efficacy of the drug and the content of 15 life elements (in Chinese). Jiangxi Journal of Traditional Chinese Medicine 1996;8(4):26–27.
[9] Yu YQ, Xie ZW, Shi YH, Gao SY, Ma FY, Wang J. Development and application of the national Chinese herbal medicine database (in Chinese). China Journal of Chinese Materia Medica 1993;18(9):106–108.
[10] Cai GD, Chen XL, Lin BH, Gui XM, Chen L. Development and application of retrieval systems of Chinese herbs database (in Chinese). Journal of Fujian College of Traditional Chinese Medicine 1994;4(3):23–25.
[11] Zhao W, Lu W, Li Z, Zhou C, Fan H, Yang Z, *et al*. TCM herbal prescription recommendation model based on multi-graph convolutional network. J Ethnopharmacol 2022;297:115109. doi:10.1016/j.jep.2022.115109, PMID:35227780.
[12] Zhang Y, Hou J, Zeng Z. Analysis of Prescription Medication Rules of Traditional Chinese Medicine for Diabetes Treatment Based on Data Mining. J Healthc Eng 2022;2022:7653765. doi:10.1155/2022/7653765, PMID:35140904.
[13] Chen H, Chen X, Han Q, Wu J, Tang DQ, Du Q, *et al*. A new strategy for quality control and qualitative analysis of Yinhuang preparations by HPLC-DAD-MS/MS. Anal Bioanal Chem 2012;404(6-7):1851–1865. doi:10.1007/s00216-012-6281-3, PMID:22885972.
[14] Zhang Y, Hou J, Zeng Z. Analysis of prescription medication rules of traditional Chinese medicine for diabetes treatment based on data mining. J Healthc Eng 2022;2022:7653765. doi:10.1155/2022/7653765, PMID:35140904.
[15] Xu HY, Liu ZM, Fu Y, Zhang YQ, Yu JJ, Guo FF, *et al*. Exploiture and application of an internet-based computation platform for integrative pharmacology of traditional Chinese medicine. Zhongguo Zhong Yao Za Zhi 2017;42(18):3633–3638. doi:10.19540/j.cnki.cjcmm.2017.0141, PMID:29218953.
[16] Li GC, Shi XD. Cluster analysis of clinical cases of pinellia Pinellia (in Chinese). Chinese Archives of Traditional Chinese Medicine 2005;23(5):836–838. doi:10.13193/j.archtcm.2005.05.68.ligch.028.
[17] Zhou SC. Data mining and analysis of the compatibility law of traditional Chinese medicines based on FP-growth algorithm. J Mathematics 2021;2021:1–10. doi:10.1155/2021/1045152.
[18] Sun X, Zhang B, Wang S, Liu S, Zhou Q. Analysis of the rule of TCM compatibility in TCM prescriptions containing *Ginseng Radix ET Rhizoma* in ancient books for Xiaoke Bing. Evid Based Complement Alternat Med 2020;2020:9472304. doi:10.1155/2020/9472304, PMID:32308724.

[19] Dai H, Fang SX. Application of ATMS in computer-assisted TCM diagnosis in Traditional Chinese Medicine (in Chinese). Application Research of Computers 2007;24(1):269–270.
[20] Cao LH, Liang JX. The Pilot Study of Diagnosis of Spleen Deficiency Syndrome by Means of Data Mining (in Chinese). Journal of Mathematical Medicine 2010;13(2):234–236. doi:10.3969/j.issn.1004-4337.2010.02.045.
[21] Chen L, Yang Z, Chen W, Li R, Lin C, Guan L, *et al*. Differential expression of immune-related genes between healthy volunteers and type 2 diabetic patients with spleen-deficiency pattern. J Tradit Chin Med 2015;35(6):646–52. doi:10.1016/s0254-6272(15)30154-0, PMID:26742309.
[22] Liu FB, Hao YT, Liu XL, Li Q, Pan ZH, *et al*. Assessment on Syndrome Differential Scale of Spleen-stomach Diseases Used for Computer Aided Expert Diagnosis System (in Chinese). Acad J SU MS 2002;23(5):401–404. doi:10.13471/j.cnki.j.sun.yat-sen.univ(med.sci).2002.0137.
[23] Wei MX, Tang YX, Chen DZ, Bei SY, Guan XZ. A new approach to establish measurable differential diagnosis criterion of Yin deficiency in stomach or spleen with information science and computer technology (in Chinese). Liaoning Journal of Traditional Chinese Medicine 2001;28(7):388–390. doi:10.13192/j.ljtcm.2001.07.5.weimx.002.
[24] Jiang YG, Li L, Li RS, Li HQ, Chen B. Experiment on data mining in compatibility law of Spleen-stomach prescription in Traditional Chinese medicine (in Chinese). World Science and Technology 2003;5(3):33–37.
[25] Zha QL, Yu JP, Wang J, Ye Y, Liu XW, Lv AP. The regularity of TCM information of gastritis from the literature of TCM treatment of chronic gastritis (in Chinese). Chinese Journal of Information on TCM 2007;14(5):102–103. doi:10.3969/j.issn.1005-5304.2007.05.058.
[26] Li WL, Zhao GP, Lu JF, Li X. Application of Association Rule in Analysis and Exploration of Famous Doctors' Experience (in Chinese). Journal of Nanjing TCM University 2008;24(1):21–24.
[27] Ren X, Guo Y, Wang H, Gao X, Chen W, Wang T. The intelligent experience inheritance system for traditional Chinese medicine. J Evid Based Med 2023;16(1):91–100. doi:10.1111/jebm.12517, PMID:36938964.
[28] Xu L, He J, Meng H. Application of decision tree based on information entropy in TCM syndrome differentiation of chronic gastritis (in Chinese). Acad J Sec Mil Med Univ 2004;25(9):1009–1013. doi:10.16781/j.0258-879x.2004.09.02.
[29] Pei C, Ruan C, Zhang Y, Yang Y. Bayes classifier chain based on SVM for traditional Chinese medical prescription generation. In: Wang X (ed). Web and Big Data. Cham: Springer; 2020:748–763. doi:10.1007/978-3-030-60259-8_55.
[30] Zhuang Y, Yan J, Chen Y, Chen J, Johnson N, Bo R, *et al*. Research on the experience of National Master of Traditional Chinese Medicine Professor Lu ZhiZheng in treating chronic atrophic gastritis based on data mining technology. Minerva Gastroenterol (Torino) 2023. doi:10.23736/S2724-5985.23.03308-9, PMID:36943205.