Application of data mining technology for inheriting medical experience of famous traditional Chinese medicine veterans in the diagnosis and treatment of diabetes mellitus

Yi Li1, Gui-Jie Wang2, Bo Pang2*

1Guang’anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China. 2Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing 100091, China.

Abstract
The inheritance of academic experience of famous traditional Chinese medicine (TCM) practitioners is the living soul of TCM. Applying systematic and scientific inheritance methods is an important means to continue the characteristic theory and unique experience of TCM. Data mining technology is a scientific research method based on the interdisciplinary content of statistics, informatics and other disciplines, and has played an important role in the inheritance of academic experience of famous TCM veterans in recent years. With the challenges of the information revolution and artificial intelligence era, TCM needs to use information technology’s unique advantages in inheriting essence and maintaining integrity and innovation to realize a high degree of integration of knowledge discovery and learning, experience, and inheritance. Taking diabetes mellitus, with high incidence in our country, as an example. This study commented on the research of data mining technology in the diagnosis and treatment inheritance of diabetes among famous traditional Chinese medicine veterans, systematically discussed the current situation, dilemma, prospect and outlook to provide new ideas for the inheritance of the diagnosis and treatment experience of famous TCM veterans with diabetes.

Keywords: data mining; famous traditional Chinese medicine veterans; diabetes mellitus; learning and experience inheritance; traditional Chinese medicine
Background

Diabetes mellitus is a disorder disease of glucose metabolism characterized by chronic hyperglycemia caused by multiple factors such as genetic factors, environmental effects, dietary preferences, and drug effects. There are about 463 million diabetics worldwide, and the global prevalence of diabetes is increasing yearly [1]. In China, the international authoritative medical journal Journal of the American Medical Association (JAMA) reported that the prevalence of diabetes and prediabetes among Chinese adults rose by nearly 4% from 2013 to 2018. The prevalence of prediabetes has increased significantly, which has become an important factor threatening the public health and economic development of China [2]. Diabetes is a worldwide problem, especially because diabetes-related risk factors have not been effectively controlled [3]. Chinese medicine, as an important means of diabetes treatment, has significant advantages in lowering blood glucose and improving patients’ symptoms and has become a superior strategy for clinical intervention in diabetes.

The rich academic experience of famous Chinese medicine practitioners provides valuable medical data for the clinical treatment of diabetes, which is an important part of the innovative development of diabetes diagnosis and treatment, as well as an important foundation for the scientific interpretation of Chinese medicine principles and the continuous innovation of Chinese medicine superior technology. Based on these backgrounds, data mining technology has played a key role in organizing and mining medical cases of diabetes “prescription-medication-evidence-symptom-sickness-person” since its introduction and more and more clinical workers have started to use data mining methods to carry out the famous old Chinese medicine experimental inheritance project objectively and efficiently. Our team has been devoted to researching traditional Chinese medicine (TCM) data mining technology for a long time, based on human experience and complex data, and has researched expanding the type of algorithm and digital intelligence inheritance [4-6]. In this paper, we review the current application of data mining technology in the field of diabetes mellitus to provide guidance and reference for the transmission of the diagnosing and treating experience of famous old Chinese medicine practitioners in diabetes mellitus. This article makes up for the shortcomings of previous research, and the framework structure of this paper is shown in Figure 1.

Application Status of data mining technology

Data mining software commonly used in TCM diabetes research

Data mining technology, also known as data mining, explores the data value laws implied in random and fuzzy data by relying on machine learning, statistical algorithms, data warehouse management, pattern recognition systems, etc. It is a multidisciplinary interdisciplinary scientific research tool for discovering unknown new knowledge based on existing data [7]. At present, the data mining software applied to diabetes research in TCM mainly includes (1) Traditional Chinese Medicine Inheritance Support System (TCMISS), which is a famous old Chinese medicine experience inheritance system jointly developed by the Institute of Automation of the Chinese Academy of Sciences and the Institute of TCM of the China Academy of Chinese Medical Sciences and can analyze and manage existing information by collecting clinical case data [8]. Since its establishment, more than 700 papers have used the system for TCM data mining. By using the TCMISS system, Professor Shi Yan’s team used the system’s built-in mutual information method, association rules, cluster analysis, complex system entropy clustering and other data processing methods to study the rules of medicine used by contemporary famous TCM doctors in treating diabetes mellitus with deficiency of both Qi and Yin (pathological changes in which Qi and Yin’s deficiencies coexist), phlegm (mucous material accumulated at the site of the lesion) and dampness, and heat that injures fluid (injury of fluid due to exuberant heat), etc. [9]. (2) TCM clinical research information sharing system. This system, led by the Chinese Academy of TCM, can realize the integration with the medical record system of partner hospitals and then realize the automatic, standardized, and visualized processing of data through structured data collection to form a multifaceted analysis network with the academic experience of famous veteran TCM practitioners as the basis. Fang, Zhouhui et al. completed the analysis of complications, combined diseases and TCM interventions in 1337 diabetic patients with the help of this system, and the results showed that benefiting Qi, nourishing Yin and activating blood is the common treatment method of famous veteran TCM doctors in treating diabetes [10]. (3) Cloud platform of ancient and modern medical cases. This system was developed by the Institute of TCM Information, China Academy of Chinese Medical Sciences, and it contains an automatic analysis module of the prescription Chinese herbal medicine’s attribution, four Qi, five tastes and efficacy, which is a technical tool that integrates big data and cloud computing model. Wang Xuanquan used this system to analyze 141 cases of prescribed drugs for type 2 sugar diabetes patients clinically treated by Professor Zhao Jinxi and found that blood stasis (pathological changes due to stagnation of Qi, Qi deficiency, blood deficiency, trauma, internalization of Yin and cold, and various other reasons, resulting in stagnation of blood in a certain part of the pathology) evidence accounted for the largest proportion of the included cases (81.56%). In contrast, the core drugs used by Professor Zhao for the treatment of type 2 diabetes, such as Gegen (Pueraria Lobatae Radix), Danshen (Salvia miltiorrhiza Radix et Rhizoma), Huangqin (Scutellariae Radix), Sanqi (Notoginseng Radix et Rhizoma), Xuanshen (Scrophulariae Radix), Chenpi (Citri Reticulatae Pericarpium), Shanyao (Dioscoreae Rhizoma), and Bei/bashen (Glehniae Radix), were excavated to provide reference for passing on academic experience of diagnosing and treating diabetes by famous veteran TCM doctors [11]. (4) Other platforms. Some inheritance platforms developed by scholar teams have been applied to diabetes treatment inheritance by famous old Chinese doctors, such as the database of Chinese medical cases and the famous old Chinese doctors’ website [12,13].

Common data mining methods for TCM diabetes research

The team systematically combed through the literature related to the research on the diabetic experience of famous veteran Chinese medicine practitioners based on data mining technology at this stage and found that the common mining methods mainly focus on frequency statistics, cluster analysis, association rules, complex network analysis, and other levels. Frequency statistics is often used to deal with the number of occurrences of target data or directional data, which is the most common and intuitive reflection of the experience of famous veteran Chinese medicine practitioners. It is often used as the initial processing step of data mining to pave the way for subsequent research based on cluster analysis and association rules [14]. Cluster analysis is an unsupervised learning model that divides data into several different categories (subsets) according to the data attributes and analyzes the categories of famous veteran Chinese medicine practitioners’ medication patterns for diabetes treatment by combining the gaps between the divided groups and also classifies the patient symptoms, which is an important analytical method for carrying out new prescription mining and implicit knowledge processing [15]. Association rules are algorithms that deal with dependencies between multiple variables, often expressed as drug-drug associations in diabetes research. Confidence and support are the prerequisites for applying association rules; confidence reflects the dependence between things, which is an expression of strength characteristics; support reflects the frequency of association rules, which is a manifestation of data prevalence characteristics. In addition, association rules can be used for the intrinsic connection between drugs and efficacy [16]. Complex network is a special network structure that represents the

Submit a manuscript: https://www.tmrjournals.com/th
drug law matrix, and in the prescription drug relationship graph, complex network technology often takes a drug element as the initial node and connects diabetes prescription drugs through nodes and edges by the data relationship between nodes and nodes [19,20]. In the context of diabetes diagnosis and treatment by famous Chinese medicine practitioners, this method can be used to further explore the intrinsic connection between “evidence and medicine”, “evidence and evidence”, and “evidence and prescription”. With the help of visualization software such as Gephi, the treatment ideas of famous veteran TCM doctors can be concretized, thus realizing the re-mining and re-organization of tacit knowledge, and providing ideas for the dispensing of medicine, combination of disease and evidence, and compound prescriptions in diabetes clinical research [21,22].

The current situation of the famous old Chinese medicine doctors’ diagnosis and treatment of diabetes mellitus

Clinical symptoms research
By analyzing the clinical symptoms, signs, pulse, and tongue of diabetic patients, the four diagnostic information and evidence elements of patients can be better summarized to tap the implicit knowledge from the medical cases of famous veteran Chinese medicine practitioners and expand their treatment experience. Xing Ying et al. used data mining methods such as Bayesian networks, cluster analysis, association rules, hidden structure models, and artificial neural networks to study the distribution characteristics of TCM evidence in 2848 patients with type 2 diabetes [23]. In the symptom distribution study, it was found that the disease features of pulse and tongue in type 2 diabetic patients were the most frequent among the symptoms recorded, with red tongue, yellow greasy coating, and stagnant veins under the tongue being the most common tongue features and string pulse being the most common pulse feature. In terms of specific symptoms, dry stools, difficulty falling asleep, frequent urination, and numbness in the limbs were the core of the complex network presented by data mining. Liu Houli et al. compiled 309 diabetes treatment cases of Director Xu Jianqin. They screened 178 cases that were clinically recognized as being consistent with damp-heat internalized diabetes mellitus, and based on the TCM inheritance assistance system, 142 symptoms were compiled, of which 20 were frequent $\geq 30$ times. The top 5 were abnormal sweating, dizziness, body fatigue, and dark red tongue, indicating that patients with damp-heat internalized diabetes mellitus were in the clinical evidence type [24]. At the same time, researchers pointed out that the clinical treatment of this type of patient should be based on the characteristics of their symptoms. The legislation should be based on the identification of the Middle jiao (below the diaphragm and above the umbilicus), and it is appropriate to use both clearing heat and strengthening the spleen. Hu Yuanzhan et al. extracted symptom and medication information from 719 real clinical cases of diabetes mellitus treated by famous TCM practitioners with the help of the TCM big data platform [25]. The clinical symptoms with a high frequency of occurrence were collated by relying on the unary and binary frequent item sets. The top five symptoms in the one-dimensional frequent item set are weakness, thirst, physical wasting, excessive urination and yellow moss; the top five symptoms in the two-dimensional frequent item set are physical wasting and yellow moss, dry stool and physical wasting, yellow moss and easy to eat, dry stool and thirst, soreness of the waist and knees and tinnitus, which objectively, comprehensively and accurately reflect the characteristics of clinical symptoms of diabetes mellitus.

Study of medication pattern
Regarding the research on the drug use pattern of famous and veteran Chinese medicine practitioners, most scholars have used association rules, cluster analysis, factor analysis and other research methods to mine the characteristic drug pair and core prescriptions of famous and veteran Chinese medicine practitioners. Hong Junjie et al. collected 63 Chinese medicine prescriptions from 18 masters of Chinese medicine to treat diabetes. They conducted association analysis on drug frequency and core drug groups to conclude that masters of Chinese medicine mostly treated diabetes by clearing heat and generating fluid, benefiting Qi and nourishing Yin, and activating blood circulation and removing blood stasis, and excavated new prescriptions for diabetes treatment, including Salvia miltiorrhiza Radix et Rhizoma, Scrophulariae Radix, Shudihuang (Rehmanniae Radix Prasiparae), Fuling (Porzia), Taoren (Persicis Semen), and Honghua (Carthami Flos) [26]. Wang Gaolei et al. collected clinical data of 300 type 2 diabetic patients treated by Director Lu Bo, analyzed the frequency of drugs, their attribution and prescription pattern, and obtained that the top five Chinese herbs used in frequency were Scutellariae Radix, Huanglian (Coptidis Rhizoma), Danggui (Angelicae Sinensis Radix), Wumei (Mume Fructus), and Ganjiang (Zingiberis Rhizoma), reflecting that Director Lu inherited the essence of Chang'an Mi's school of internal medicine in treating diabetes, starting from “deficiency (ZhengQi) insufficiency, with ZhengQi deficiency as the main aspect of the contradiction of pathological reactions, manifested as the body's essence, qi, blood, fluid deficiency and functional weakness, the organs and meridians of the function is low, the ability to resist disease is reduced, can be seen in a variety of weakness and insufficiency of the evidence), stasis (blood stagnation or coagulation in the body, including blood overflowing outside the meridians and stagnation, as well as stagnation in the lumen of the meridians due to obstruction of blood flow, is both a pathological product and can be a secondary pathogenic factor), and toxicity” and making good use of the self-prepared reduced-flavored Wumei pill to treat diabetes mellitus with deficiency of both Qi and Yin, Stasis and toxicity [27,28]. Ni Linlin et al. collected 120 prescriptions of Professor Feng Jianhua for the treatment of diabetes mellitus and obtained the core drug relationship network by correlation analysis, and concluded that the core prescriptions of Professor Feng were Shengshigao (Gypsym Fibrosam), Salvia miltiorrhiza Radix et Rhizoma, Shendihuang (Rehmanniae Radix), Coptidis Rhizoma, Zhiwu (Anemarrhenae Rhizoma), Wuweizi (Schisandrae Chinensis Fructus), Puerariae Lobate Radix, and Zou mays, and concluded that Professor Feng's diabetes diagnosis and treatment habitually used sweet and bitter drugs. Treatment was often based on the legislation of nourishing Yin and
clearing heat [29]. Xiong Guohui et al. collected 75 core prescriptions from Professor Fang Chaohui for the treatment of type 2 diabetes mellitus with Qi and Yin deficiency, applied association rules (second-order and third-order) and cluster analysis to obtain common drug pairs and new prescription combinations, and concluded that Professor Fang's diagnosis and treatment of diabetes mellitus often took the core treatment of benefit Qi, nourishing Yin and generating fluid, and the treatment was characterized by emphasizing the functions of the three organs of the lungs, spleen and kidneys, and treating the upper, middle and lower extremities together [30]. Wan Anqing et al. used an improved Apriori algorithm to analyze the prescriptions in the National Famous Old Chinese Medicine Practitioners' Good Diabetes Cases, and among the 402 included prescriptions, the total number of high-frequency drugs was 24 [31]. At the same time, based on C + + language combined with R data mining software using the ADPM algorithm, 15 high-frequency core medicine groups were obtained, and Danpi (Moutan Cortex), Luron (Cerci Cornu Pantotrichum), Ezhui (Curcuma Rhizoma), and Portia were the top-ranked drugs in terms of confidence. Most of the medicines used by famous old Chinese doctors for diabetes were used to clear heat, tonic deficiency, diuretic and dampness, and blood circulation and blood stasis. The improved algorithm speeds up the data mining iteration process and avoids the shortage of frequent pattern mining efficiency.

**Study of drug-syndrome/evidence patterns**

Exploring the distribution pattern of drugs and symptoms and evidence patterns in the diagnosis and treatment of diabetes by famous veteran TCM practitioners is a vital manifestation to show the unique advantages of TCM in the treatment of diabetes. It is also an essential way to refine the academic ideas of famous veteran TCM practitioners, to inherit the experience of combining disease and evidence, corresponding prescriptions and evidence and drugs, to improve the clinical efficacy of TCM in the diagnosis and treatment of diabetes, and to build a scientific and guiding clinical decision support system [32]. With the help of the PageRank algorithm, Ni Boran et al. extracted the prescription pattern of Professor Lu Renhe, the master of Chinese medicine, in the treatment of type 2 diabetes mellitus from four Qi and five tastes, medicine pair groups, core prescriptions and medicine dose in multiple dimensions and obtained the 'drug-symptom' linkage network [33]. We summarized the characteristics of Professor Lu's prescriptions, such as Ganzhuzi (Atractylodis Rhizoma) and Baizhuzi (Atractylodis Macrocephalae Rhizoma) for diahrrea and bloating, and Capitidis Rhizoma and Muxiang (Aucklandiae Radix) for red tongue with yellow coating and bitter mouth. In addition, combined with drug weighting levels, Professor Lu's medicine is systematically interpreted with the combination of disease, evidence, and symptoms and the rules of staged diagnosis and use of medicine.

**Prospects for the application of data mining technology in diabetes etiology test heritage**

Data mining technology is widely used in Chinese medicine diagnosis and treatment of diabetes and the treatment of famous and old Chinese medicine. With diversified mining methods, systematic research can be conducted on prediabetes, diabetes and diabetes complications. Most of the studies are focused on patients' basic information, symptoms and signs, evidence, prescription drugs (four Qi, five tastes, generalization and efficacy), core prescriptions and test indicators. In recent years, most of the research has been focused on medicine and evidence, symptoms, prescriptions and treatments, and the mining methods have been extended to more complex and refined multi-natural language processing methods, reflecting the rapid development of data mining technology. The development of data mining technology is an important force for TCM inheritance, and it provides technical support for optimizing TCM research methods and better inheriting the experience of famous masters. Currently, data mining technology still has certain shortcomings in transmitting diabetesetology tests. Firstl, Chinese medicine is a practical medicine, and its thinking method comes from inductive deduction, logical reasoning, and enlightened judgment of clinicians on the treatment process. The transmission of the experience of famous veteran TCM doctors in treating diabetes requires the guidance of the TCM thinking system, and the thinking system of diabetes diagnosis and treatment is complex. Famous veteran TCM doctors often develop unique patterns of thinking in treating diabetes in terms of identification and use of medicine. The establishment of such thinking is based on the clinical cases and experience in heritage and the process of famous veteran TCM doctors' collecting information, analyzing information, summarizing information, and writing information, whether the research method relying on computer mining can avoid the drawbacks of the western thinking research model is still a key element that needs to be solved. Secondly, the results of some diabetes data mining studies are still defective, either because the mining process is not standardized or because fewer case studies are included. This results in certain discrepancies between the core prescriptions or drug-evidence relationships mined and the academic ideas of the famous Chinese medicine practitioners analyzed. In addition, diabetes data mining research methods are currently limited. The algorithm used is still slightly single, but the contemporary development of computer information technology is an important opportunity for TCM Shouzheng innovation research, relying on the progress of modern information technology field, the application of data mining technology in diabetes test heritage will be pre-broad and promising.

**References**

5. Liu FD, Jiang XC, Wang GR, Pang B, Hua BJ. Application and reflection of machine learning methods in the study of Chinese medicine inheritance. J Tradit Chin Med 2022;63(8):720–724 + 738. (Chinese) Available at: http://kns.cnki.net/kcms2/article/abstract?v=WgOwWnP4PgkPF5r6T1-r6HaHt5N9kS0tUq-RADB01sRWNWlwtkGKjMyEooZJqGRAF3rNWRJqap-dlreYgKgFYSH1-KplvYRYb1y8YM4hhATBvYmgnk4lmlG9Q9NqUkJU&uniplatform=NZKPT6 lanugage=CHS


18. Chen KL, Fan YP. Classification algorithms in data mining and their applications in Chinese medicine evidence-based medicine. Chin J Tradit Chin Med 2011;26(3):469–473. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMapW6l8S5bp1a1gVhFo-xlC5ptkvTWWbp-PI-NZ-rbB54vzu2ZNDyC rYlxymwumZ2r6xu5V0OaooQy2J0zStqVYXc5bSVg0YJGFHGe -KJCh3r10cok4qYr&uniplatform=NZKPT&language=CHS


20. Yang YN, Li YJ, Chen YP, Ni Q. Dosing rules and core prescriptions of marketed proprietary Chinese medicines for the treatment of diabetes mellitus. World Tradit Chin Med 2022;17(3):418–423. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMae ATHHOD0FkmHSMMCuLVBJRBXcAn4HH2ZJ5cx6s14TrkV pC683FntvzFBI2I8JAJA3lElk77C7Cqnxawd6M0n3yanP9lnhV4l KA40mmPD7MYHPolgLOuRQaAsi &uniplatform=NZKPT&lan guage=CHS


26. Hong JJ, Zheng SQ. Research on the data mining based on the medication pattern of national medical masters in treating diabetes mellitus. Shandong J Tradit Chin Med 2019;38(5):444–449. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMaqZB8Rs-DuNQW_0xvXNispOCQ1vHmLqUkQ0UuMBySPoCRe nNSQleadpHT79yBEnOQA5QIurOFwGW41rveGED3Sm1x0sXnYrD6aAnWi12bSh4R2a0WXEk-GKus90GGe &uniplatform=NZKPT&Tlanguage=CHS

28. Zhang GJ, Lu B. Lu Bo’s experience in the treatment of achlorhydria from syncope. Shanxi TCM 2013;29(11):8-9. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMaazo4Vhn92kI6844i41cmHv9IQCl0FzGxxqKt_jsvQSUKtDq19XdPrXBNwwZwA3HfVUGF8jXTSFR1zvH9Q8WwHKz7rlytOyeioOvZTW1a9iPF0&uniplatform=NZKPT&language=CHS

29. Ni LL, Xu YS, Zhou HL. Data mining-based analysis of the medication pattern of Feng Jiahua famous veteran Chinese medicine practitioners for the treatment of type 2 diabetes mellitus. Chin Med Herald 2022;19(18):102–105+118. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMarq30uK0WesNC-Ir6VrLoIj19muI5Vni8BOW9gB8f88ROG3mMkJn3TunxTL72XwupUIXOZKjS5mZcuAxCnC9zmkI5H9It8JZriri953U19vdOrHkkQcBlVUKC4=&uniplatform=NZKPT&language=CHS


31. Wan AQ, Bao JY, Hu KF. Apriori improvement algorithm-based study on the medication pattern of famous veteran Chinese medicine practitioners for the treatment of diabetes mellitus. Chin J Tradit Chin Med Inf 2017;24(12):97–101. (Chinese) Available at: https://kns.cnki.net/kcms2/article/abstract?v=rHLknEdMarq31u3BHs0w5h5eGhboxxGErCY11G1c_C0tS57v7kkaL5gSNMFJXW16v5SvQvUn8UXGt5y0HQyQ3xrekjW0HsLiWxXfctBll0hbtUoiChHd02jHvRRe2R118qkpry069_10zRAeCAV=&uniplatform=NZKPT&language=CHS
