Views & Comments

Construction of an Artificial Intelligence Traditional Chinese Medicine Diagnosis and Treatment Model Based on Syndrome Elements and Small-Sample Data

Jie Wang, Lian Duan, Hongzheng Li, Jinlei Liu, Hengwen Chen

Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China

1. Introduction

The combination of traditional Chinese medicine (TCM) and Western medicine has advantages in the treatment of chronic and complex diseases, for example, the introduction and application of TCM during the coronavirus disease 2019 (COVID-19) epidemic in Hubei Province, China, had a total effective rate of 90%, thus establishing the efficacy of integrating TCM and Western medicine. However, the integration of TCM diagnosis and treatment methods with modern medicine, as well as their further development, is limited due to several significant challenges. Since the 1950s, information technology has been used to explore the four core TCM diagnostics: inspection, listening and smelling, inquiring, and palpation. Since the 1970s, researchers worldwide have attempted to develop expert TCM systems in the context of clinical diagnosis and treatment, but have failed to adequately simulate the syndrome differentiation and treatment associated with TCM. Based on the continuous breakthroughs that have been achieved in artificial intelligence (AI), an expert TCM system has been integrated with technologies such as neural network fuzzy logic and relational databases as a means of advancing TCM research. Researchers have also attempted to use TCM big data—that is, data collected from TCM doctors across China—for performing objectification, standardization, and quantitation in research. However, it is difficult to integrate the knowledge obtained from the long-term clinical experience of TCM practitioners entirely using data-driven methods. Thus, imposing effective constraints on the learning process seems to be prohibitively difficult.

In recent years, artificial neural networks have emerged as an effective means of dealing with nonlinear problems. These neural networks, along with decision trees and random forest algorithms, which are highly effective for dealing with high-dimensional data, have been applied to TCM models [1,2]. These tools are particularly well suited for TCM because—unlike modern medicine, which has almost uniform and clear diagnosis and treatment rules—there is substantial overlap between diversified TCM theories, different schools of TCM, and various TCM practitioners, which makes it difficult to accumulate structurally similar and homogeneous

TCM data of high quality. The existing training samples for TCM are too small to properly train a model. To address this issue, standardized interpretable models can be used to integrate clinical knowledge with AI in order to achieve a functional combination of modern medicine and TCM.

2. Syndrome elements unify rules for diagnosis and treatment

Theoretical models and clinical applications of TCM treatment are primarily based on syndrome differentiation, which is a summary of the pathological and physiological differences in diseases that is determined based on etiology, disease site, and the nature of the disease [3]. In other words, syndrome differentiation summarizes the characteristics of various symptoms, and the treatment for any particular syndrome is based on specific principles and methods [4]. This process is characterized by multiple levels of complexity, abstraction, and high dimensionality [5]. However, there are problems in the application of the guiding principles of syndrome differentiation that preclude accurate diagnosis, such as fuzzy boundaries between syndromes, a high rate of overlap between application scenarios, and a superposition of syndrome differentiation parameters [6]. Thus, TCM methods that have been developed based on syndrome differentiation cannot establish effective connections between theories, methods, medicinal formulas, and materials, thereby reflecting the difficulties that are present in diagnosis standardization.

Any complex syndrome is determined by the intersection of specific symptom dimensions, such as the disease site, its specific nature, and the relationship between the pathogenic factors and body's natural resistances. The specific presentation of each dimension can be considered as an element of the syndrome differentiation process [7]. It has been reported that approximately 60 basic syndrome elements can be isolated by evaluating how the different characteristics of syndromes overlap. These basic syndrome differentiation elements can be arranged and combined to cover most of TCM syndrome types [8]. The smallest units of syndrome differentiation in TCM are called syndrome elements.

Syndrome elements should therefore be regarded as the cornerstone or fundamental building block when designing and applying AI models that use TCM to assess combinations of diseases and syndromes. This method of analysis enables connections to be made among symptoms, treatment methods, medicinal materials, and formulas during diagnosis and treatment. In contrast to the complexity and high dimensionality of syndromes, which are defined as a constellation of related symptoms, each syndrome element has a specific symptom group that specifies and differentiates it from other syndrome elements. By combining and superimposing these elements, clinically observed syndromes can be formulated as amalgamations of syndrome elements. Therefore, syndrome elements have low dimensionality, which allows for their relatively easy superimposition and combination. Each syndrome element can be treated using specific methods, medicinal materials, or drug combinations. In this way, broader syndrome treatments can be determined accordingly, once the syndrome elements and their combinations have been defined [9].

Syndrome elements should be considered as indispensable in the process of TCM diagnosis and treatment processes. We take coronary atherosclerotic disease (CAD) as an example. All clinical manifestations of patients with CAD can be classified into eight groups: blood stasis, qi deficiency, phlegm turbidity, yin deficiency, qi stagnation, yang deficiency, cold coagulation, and heat accumulation [10]. These elements can be combined to form syndrome types such as qi deficiency and blood stasis, qi and yin deficiency, intermingled phlegm and blood stasis, qi stagnation and blood stasis, phlegm obstruction and heat accumulation, and yang deficiency and cold coagulation [11]. According to the corresponding relationships between syndrome elements and their respective treatment methods, it can be inferred that TCM treatment methods for CAD may include promoting blood circulation, tonifying qi, eliminating phlegm, nourishing yin, regulating qi, warming yang, dispelling cold, and clearing heat. Subsequently, according to the treatment methods, the corresponding medicinal formulas, drug pairs, or medicinal materials can be obtained. This includes formulas such as Guanxin II formula; drug pairs such as *Codonopsis pilosula* and *Astragalus membranaceus*, or *Fructus piperis longi* and *Rhizoma alpiniae officinarum*; and single medicinal materials such as *Salvia miltiorrhiza* and *Panax notoginseng*. Then, according to the combinations of syndrome elements, the medicinal materials can be combined to form representative formulas for CAD treatment, such as Xuefu Zhuyu decoction and Gualou Xiebai Banxia decoction [12]. The superposition and combination of syndrome elements is the basis for changes, additions, and the removal of medicinal materials and formulas from the treatment plan. Syndrome elements are therefore helpful for integrating and unifying the diagnosis and treatment rules of TCM from different sources.

3. Integration of algorithms for a small-sample TCM data model

The use of efficient algorithms is crucial when applying AI systems to TCM. When devising AI algorithms for TCM, the applicability of the algorithm should be considered. Bayesian networks and support vector machines are ideal for considering the number of cases necessary for the identification of syndrome information using the four diagnostic methods [13,14], as they are accurate when the data used are small-scale and robust against noise. However, such algorithms are computationally expensive when applied to real-world and largescale datasets, which affects their practicability.

The complexity of syndrome differentiation in TCM requires the use of nonlinear models. Artificial neural networks, which have an excellent ability to analyze nonlinear problems, can simulate the structure and functioning of neural networks in the human brain

to effectively process data. This can help in the identification of complex, hidden patterns in the data, which is particularly advantageous when solving TCM syndrome problems [15]. However, neural network algorithms require a high level of standardization, and training such models requires a relatively large quantity of data. Thus, the lack of standardized TCM data limits the applicability of such models.

The complete simulation of the TCM diagnosis and treatment process and the processing of complex, multidimensional data are the biggest challenges in employing AI models for TCM. Decision trees and random forest algorithms are efficient at processing data with high dimensionality, exploring the interactions among data features, and meeting data requirements and contingencies when finding relationships among the data. However, decision trees can easily be over-fitted, which may reduce their accuracy. Random forest algorithms account for this problem to some extent by building multiple decision trees in a random manner; as a result, they are effective in dealing with unbalanced data. A study using random forest algorithms [16] to establish prediction models for TCM syndrome elements of chronic fatigue has reported high prediction accuracy. However, such models have limited adjustability and cannot accurately classify small-scale data. When applied to TCM, the limited availability of data results can be problematic.

Although AI models have led to advancements in the field of TCM, two major issues remain to be resolved. The first is the lack of standardized and objective data. AI models require large quantities of standardized and objective data for training. However, there are limited standardized datasets available for TCM, and the marked subjectivity and domain specificity in the data make the objectification process of TCM time-consuming. This limits the applicability of new AI algorithms such as artificial neural networks to TCM. The second issue is that existing studies have focused primarily on single application scenarios. The application of a single algorithm establishes prediction models based on only a specific type of data; as a result, the universality and portability of the established models are generally insufficient. This leads to an inability to formulate universal and systematic results for multiple diseases, as the focus of the model is narrow. Thus, formulating AI methods for diagnosis and disease characterization using TCM by integrating multiple algorithms and data standardization is an important research problem in the field.

4. Strategies and directions for AI in TCM

Based on the most recent advancements in AI algorithms, we propose an AI model for TCM to explore syndrome elements and their related rules. Using syndrome elements as the starting point, rules integration as the path variable, and symptom scores provided by established diagnosis and treatment guidelines as the weights, recommended prescriptions can be formed through symptom combination, disease identification, syndrome derivation, treatment methods and rules, formation of medicinal formulas, addition and removal of medicinal materials, and assignment of dosage values (Fig. 1). In addition, to improve the accuracy of the syndrome calculations, a small sample of high-quality TCM diagnosis and treatment data should be used, and a feedback mechanism should be established. TCM practitioners can adjust these parameters based on the output results, with the adjustment being added as a new rule to the rule dataset.

By leveraging the learnability, expandability, practicality, and iterability of AI TCM models combined with a thorough understanding of syndrome elements, syndromes, medicinal materials, medicinal formulas, clinical cases, classical texts, and medical records, TCM diagnosis and treatment knowledge systems
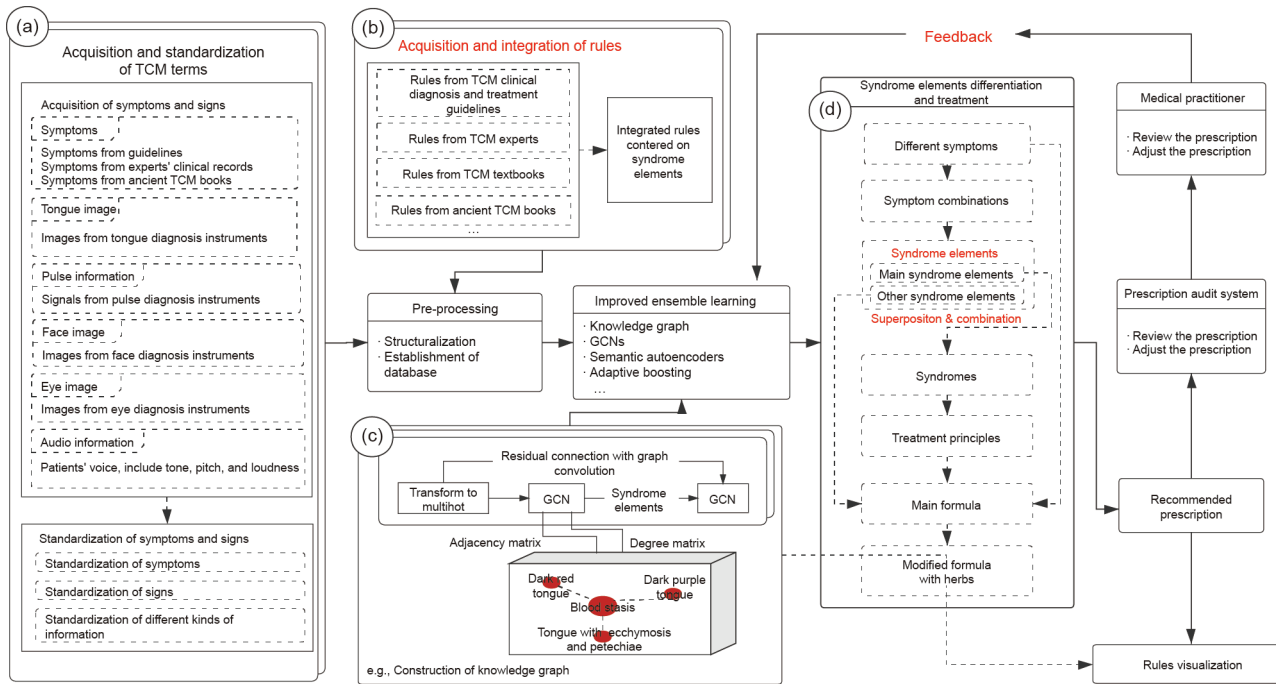
Fig. 1. Path of an AI diagnosis and treatment model for an exemplar disease for TCM. (a) Acquisition and standardization of TCM terms. We acquire the term information for the symptoms, including from the tongue, pulse, face, visual image, voice, and other demographic patient information. This data is then labeled to complete the data standardization. (b) Acquisition of TCM diagnosis and treatment rules. TCM diagnosis and treatment rules can be derived from TCM guidelines, expert experiences, teaching materials, ancient classics, and so forth. These diagnosis and treatment rules are associated with and integrated from different sources through syndrome elements. Integrated diagnosis and treatment rules with syndrome elements as their core can be formed from this basis. (c) Constructing a knowledge graph. Taking the construction of a graph convolutional network (GCN) and a knowledge map as an example, this figure illustrates the schema for integrating algorithm rules to form a visual model. A knowledge map is constructed with syndrome elements and symptoms as the nodes and with the correlations between symptoms and syndrome elements as the boundaries. The corresponding weights are derived from the symptom scores obtained from standard diagnosis and treatment guidelines for particular diseases. Then, a convolution operation is used for weighted summation, and the results are obtained in the form of output to establish a visual model. (d) Integrating the diagnosis and treatment rules of various sources to identify syndrome elements and output the appropriate prescription. Taking standardized symptoms as the input layer, we first judge the syndrome elements corresponding to standardized symptoms, and then form a hypothesized syndrome through the superposition and combination of the various syndrome elements. Through the corresponding relationship between syndrome and treatment, we output the appropriate prescriptions and TCM practices relevant for that combination of main syndrome elements, and then use the formula and TCM to identify the remaining syndrome elements. Other syndrome elements and symptoms result in their own corresponding TCM recommendations and contribute to the drug prescriptions for the main set of prescriptions as the output layer.

originating from different sources and medical fields can help to integrate the disparate bodies of knowledge in the medical field. Corresponding relationships among the symptoms, syndrome elements, syndromes, treatment methods, medicinal formulas, and medicinal materials can be established to improve the accuracy of syndrome differentiation. Graph convolutional networks and knowledge graphs can help to integrate rules. Based on this method, a knowledge map is constructed with the syndrome elements and symptoms as the nodes and the correlations between symptoms and syndrome elements as the boundaries. The corresponding weights of these correlations are derived from the symptom scores obtained from the standard diagnosis and treatment guidelines for any particular disease. By constructing an adjacency matrix and degree matrix, the Laplacian matrix is then calculated to represent the weight of different symptoms under the different syndrome elements. Finally, a convolution operation is used to perform a weighted summation, and the results are used to establish a visual model. The visual display of these rules can convey the rule data more intuitively than purely numerical or text-based descriptions, and can therefore help TCM doctors learn from the experience of well-known doctors.

Based on the abovementioned design, an AI-based TCM system that conforms to industry standards, classical texts, and expert opinions was formulated (Fig. 1). This research strategy can serve as a benchmark for the clinical application of guidelines and can provide a new direction for research on standardized and intelligent AI-based TCM diagnosis and treatment systems.

5. Outlook

In our research, CAD was taken as a breakthrough example of the integration of TCM and AI for the diagnosis and treatment of a single disease. We analyzed the possible syndrome elements through clinical symptoms, as well as through tongue and pulse information, and then obtained reliable prescriptions using syndrome element as the core variable. Thus, based on syndrome elements and artificial prior knowledge, we designed a TCM diagnosis and treatment model based on AI. This model is capable of training even when the quantity of sample data is important. The proposed model has higher accuracy in AI prescription generation than that of the existing big data TCM models. We expect that an AI diagnosis and treatment model similar to the one we developed in our study can be implemented in most TCM contexts based on the characteristics of TCM and the experience of TCM experts. We predict that such an AI diagnosis and treatment model that uses prior knowledge can continue to improve the rationality and stability of syndrome diagnosis and further increase the accuracy of medical prescriptions.

Regarding potential future directions, we anticipate the conversion of the TCM medical system from a single-disease model to a multi-disease model, as well as from a pure TCM system to one that integrates TCM and Western medicine into a single diagnosis and treatment system. We hope that, in the near future, this approach can also simplify the medical service process, improve the efficiency of diagnosis and treatment, and allow patients to

receive AI-based diagnoses and treatments in order to help offset the increasing scarcity of highly experienced TCM experts. Under such a system, even patients who are far from a hospital will still be able to receive the best medical services in a more accessible manner.

Acknowledgments

References

[1] Zhang F, Wu C, Jia C, Gao K, Wang J, Zhao H, et al. Artificial intelligence based discovery of the association between depression and chronic fatigue syndrome. J Affect Disorders 2019;250:380–90.

[2] Wang Z, Li L, Song M, Yan J, Shi J, Yao Y. Evaluating the traditional Chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning. J Ethnopharmacol 2021;272:113957.

[3] Sun XL, Jiang WW, Yu DL, Lin X, Ding BG, Liu MA. Research on basic path and inherent law in the process of Chinese medicine treatment based on syndrome differentiation. J Tradit Chin Med 2016;57(4):289–94. Chinese.

[4] Wang YY. Suggestions on improving the system of traditional Chinese medicine syndrome differentiation. J Tradit Chin Med 2004;45(10):729–31. Chinese.

[5] Li S. Discussion on the characteristics of TCM syndromes from the concepts of dimension and ranks: approaches to the standardization of TCM syndromes. J Beijing University Tradit Chin Med 2003;26(3):1–4. Chinese.

[6] Zhang ZB, Wang YY. Establishment of a new TCM syndrome differentiation system. J Beijing University Tradit Chin Med 2005;28(1):1–3. Chinese.

[7] Zhang ZB, Wang YY. Research on TCM syndrome nomenclature and classification: review and hypothesis. J Beijing University Tradit Chin Med 2003;26(2):1–5. Chinese.

[8] Zhu WF. Standardization research of differentiation system of symptoms and signs and "syndrome" in TCM. Tianjin J Tradit Chin Med 2002;19(5):1–4. Chinese.

[9] Wang J, Tang YL, He QY, Xiong XJ. Thinking of prescriptions corresponding to syndromes in construction of syndrome differentiation system. J Tradit Chin Med 2009;24(7):837–9. Chinese.

[10] Wang J, Li J, Yao KW, Zhong JB. Study on syndrome elements and their combination laws in patients with angina pectoris. J Tradit Chin Med 2007;48 (10):920–2. Chinese.

[11] Li J, Wang J. Literature case analysis of 5099 cases of syndrome elements and syndrome combinations of angina pectoris of coronary heart disease. Chin J Basic Med Tradit Chin Med 2007;13(12):926–7. Chinese.

[12] China Association of Chinese Medicine. Guidelines for diagnosis and treatment of stable angina pectoris of coronary heart disease (T/CACM1325–2019). J Tradit Chin Med 2019;60(21):1880–90. Chinese.

[13] Zhang H, Ni W, Li J, Zhang J. Artificial intelligence-based traditional Chinese medicine assistive diagnostic system: validation study. JMIR Med Inform 2020;8(6):e17608.

[14] Xu F, Xu ZX, Xu WJ, Wang YW, Liu T, Xia CM, et al. Classification of TCM syndromes in 835 cases of coronary heart disease: on the basis of Bayesian networks principle. Shanghai J Tradit Chin Med 2014;48(1):10–3. Chinese.

[15] Zhang NL, Yuan S, Chen T, Wang Y. Statistical validation of traditional Chinese medicine theories. J Altern Complem Med 2008;14(5):583–7.

[16] Hong YZ, Zhou CL, Zhang ZF, Xu JT. Selection of characteristic symptoms of chronic fatigue syndrome elements based on random forest method. J Tradit Chin Med 2010;7:634–8. Chinese.