





## Research Article

# Knowledge Graph Construction of Chronic Kidney Disease Diagnosis and Treatment with Traditional Chinese Medicine

Yiadong Xie <sup>1,2</sup>, Jiayi He <sup>1</sup>, Ping Xia,<sup>3</sup> Weiming He <sup>4</sup>, Min Huang,<sup>1</sup> Peipei Fang,<sup>1</sup> Chenjun Hu,<sup>1</sup> and Kongfa Hu <sup>1,5</sup>

<sup>1</sup>School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China

<sup>2</sup>Institute of Literature in Chinese Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China

<sup>3</sup>Department of Traditional Chinese Medicine, First Affiliated Hospital of Soochow University, Suzhou 215006, China

<sup>4</sup>Division of Nephrology, Affiliated Hospital of Nanjing University of Chinese Medicine, Jiangsu Province Hospital of Chinese Medicine, Nanjing 210029, China

<sup>5</sup>Jiangsu Collaborative Innovation Center of Traditional Chinese Medicine in Prevention and Treatment of Tumor, Nanjing 210023, China

Correspondence should be addressed to Weiming He; yfy0019@njucm.edu.cn and Kongfa Hu; kfhu@njucm.edu.cn

Received 22 October 2022; Revised 5 December 2022; Accepted 18 April 2023; Published 12 May 2023

Academic Editor: Xuezhong Zhou

Copyright © 2023 Yiadong Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** The purpose of this study is to construct a knowledge graph of chronic kidney disease (CKD) diagnosis and treatment with traditional Chinese medicine (TCM), reorganize its knowledge, and display it. It allows the inheritance, development, and utilization of CKD diagnosis and treatment experiences with TCM in a standard and scientific manner. **Methods.** First, we constructed a knowledge framework for TCM diagnosis and treatment on the basis of the Chinese Pharmacopoeia, government projected textbook, and the current TCM diagnosis and treatment standards. Then, we collected and sorted the electronic medical records of TCM inpatients, extracting and normalizing the diagnoses, symptoms, syndromes, prescriptions, and other diagnosis and treatment information, creating the knowledge base of TCM diagnosis and treatment for CKD. Finally, we stored TCM diagnosis and treatment CKD knowledge in Neo4j graph database, which refers to the knowledge framework and knowledge base. The frequent patterns and complex network knowledge mining methods are integrated to construct the TCM diagnosis and treatment CKD knowledge graph. **Results.** The knowledge graph of CKD diagnosis and treatment with TCM was constructed, including 807 nodes and 10476 relationships, which are 273 diagnoses, 130 symptoms, 34 syndromes, 370 Chinese herbal medicine (CHM) nodes, and 5483 diagnosis-symptom, 1349 diagnosis-syndrome, 3644 syndrome-CHM relationships. **Conclusion.** The knowledge graph provides rich knowledge of TCM diagnosis and treatment of CKD, which is helpful to inherit the clinical experience of TCM diagnosis and treatment of CKD and assist clinical diagnosis and treatment of CKD.

## 1. Introduction

Around the world, CKD has become a major global public health challenge that poses a serious threat to human health. According to survey data, the prevalence of CKD in adults is 10.2%–13% in the United States, Norway, and other Western developed countries [1, 2]. The situation in China is similar. According to the results of the “Epidemiological Survey of Chronic Kidney Disease in China” led by Professor Haiyan Wang of Peking University First Hospital, the prevalence of

CKD in Chinese adults over 18 years old is 10.8%, which estimated that there are approximately 120 million CKD patients in China [3]. Although there is no record of CKD in ancient Chinese medicine books, based on its clinical manifestations, this disease belongs to the category of “oedema, hemuresis, lumbago, deficiency, uroschesis, drowning poison, obstruction, and rejection” of TCM [4]. In clinical practice, TCM has beneficial therapeutic effects for CKD, which include delaying disease progression and dialysis initiation [5, 6]. The Nephrology Department of

Jiangsu Provincial Hospital of Chinese Medicine was founded in 1954. It is the earliest medical group of nephropathy with TCM in China. It has unique advantages and characteristics at the academic level and academic innovation, including innovative theories, technologies, and new drugs. They formulated clinical diagnosis and treatment plans for three dominant diseases: renal fatigue, chronic renal wind, and gonorrhoea. Professor Yunxiang Zou treated nephrotic oedema by activating blood and invigorating water from the lung, spleen, and kidney [7]. In the case of nephrotic asthenia, the nephron should be toned, turbidity cleared, and collaterals cleared. Professor Yanqin Zou, following his father Yunxiang Zou, applied the He-Luo method based on syndrome differentiation and treatment to smooth the veins of patients with CKD and improve the stagnation of qi and blood circulation [8]. Professor Wei Sun [9] believes that the basic pathogenesis of CKD is kidney deficiency, dampness (heat), and blood stasis. Kidney deficiency is the basis for the occurrence and development of the disease, and dampness and blood stasis are important factors in disease progression. He proposed “benefiting kidney, clearing away the damp heat, and promoting blood circulation,” which is the basic treatment for CKD.

In 2012, Google introduced the knowledge graph to raise the search engine efficiency. Subsequently, knowledge graphs show rich application value in assisting intelligent question answering, natural language understanding, big data analysis, recommendation calculation, and other aspects, and provide technical support for knowledge graph research in e-commerce, finance, medical, and other specific fields. In the field of TCM, the application and construction of knowledge graph face the difficulties of diverse knowledge sources, complex knowledge structures, and high requirements for knowledge quality. Some researchers have constructed relevant knowledge graphs through electronic medical records, expert interviews, and the literature. Xuezhong Zhou et al. work developed a unified traditional Chinese medical language system (UTCMLS) through an ontology approach that sorted out and improved the concepts and terms of traditional Chinese medicine [10]. Combining UMLS with the characteristics of Chinese medicine language, they constructed a large corpus database and semantic network, which integrate linguistics with the knowledge system of traditional Chinese medicine. Tong Yu et al. work constructed a large-scale knowledge graph, which integrates terms, documents, databases, and other knowledge resources to facilitate various knowledge services such as knowledge visualization, knowledge retrieval, and knowledge recommendation, and help the sharing, interpretation, and utilization of TCM health care knowledge [11]. Xinhong Jia et al. used BILSTM-CRF to extract knowledge from the electronic medical records of patients with CKD and used Neo4j for knowledge representation and storage to construct a slow obstructive pulmonary knowledge graph [12]. Xinlong Li examined the case of syndrome differentiation and treatment of insomnia by three TCM physicians, used Gephi to build a knowledge graph of clinical personalized syndrome differentiation and treatment [13].

He further revised and optimized the knowledge graph through comprehensively applied expert interviews, literature research, and complex network.

Based on the diagnosis and treatment norms, Chinese Pharmacopoeia, planning textbooks, and other literature, taking the clinical diagnosis and treatment of CKD as the core, and assisted by data mining analysis methods, this study further summarizes the knowledge of TCM diagnosis and treatment of CKD, and shows the relationship between syndrome, TCM prescription, and CHM through data visualization technology so as to provide researchers, medical workers, and students with rich knowledge of TCM diagnosis and treatment of CKD. It is helpful to inherit and develop the clinical experience of TCM and assistance in the diagnosis and treatment of CKD.

## 2. Materials and Methods

*2.1. Data Resource.* The data are gathered from the inpatient electronic medical record of the Nephrology Department and chronic disease management system of Jiangsu Provincial Hospital of Chinese Medicine, TCM diagnosis and treatment norms, Chinese Pharmacopoeia, and government projected textbooks.

*2.1.1. Electronic Medical Records of TCM Inpatients.* 8,017 cases of chronic kidney disease were collected in Jiangsu Hospital of Traditional Chinese Medicine from January 1999 to October 2020. The data are stored in semistructured and text forms, including personal information, diagnosis and treatment information, first course record, ward round record, and discharge record. Among them, the four diagnosis summaries contained in diagnosis and treatment information and the TCM prescription contained in ward round records are the most critical.

Inclusion criteria are as follows: (1) CKD is diagnosed clinically, and the stage is 1~5; (2) the case information is complete, including at least basic personal information, diagnoses, symptoms, syndromes, and prescriptions; (3) having more than three prescriptions of TCM.

Exclusion criteria are as follows: (1) poor compliance; (2) incomplete case data.

*2.1.2. TCM Diagnosis and Treatment Standard.* TCM disease syndrome and clinical diagnosis and treatment terms refer to classification and codes of diseases and patterns of traditional Chinese medicine [14], Clinic terminology of traditional Chinese medical diagnosis and treatment diseases, Clinic terminology of traditional Chinese medical diagnosis and treatment Part 1: Diseases and Clinic terminology of traditional Chinese medical diagnosis and treatment; Part 2: Syndromes and Clinic terminology of traditional Chinese medical diagnosis and treatment; Part 3: Therapeutic methods. CKD diagnosis and clinical staging refer to the KDIGO Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease (2012 Edition) [15].

**2.1.3. CHM Knowledge Base.** Based on the medicines included in “Chinese Pharmacopoeia (2020 Edition),” for those not included in actual studies, we supplemented them according to Chinese Materia Medica. The CHM knowledge base is established, including Chinese name, pinyin, English name, description, character, identification, inspection, extract, content determination, processing, nature, taste and meridian, function, usage and dosage, attention, storage, use, preparation, specification, taboo, preparation, indication, characteristic map, and fingerprint map.

**2.2. Methods.** Figure 1 shows the technology roadmap of constructing a knowledge graph for CKD diagnosis and treatment in TCM. First, we established the framework of TCM diagnosis and treatment knowledge according to TCM diagnosis and treatment norms, CHM knowledge base, and TCM syndrome differentiation and treatment principles. The framework was validated and approved by TCM experts. Then, we used natural language processing methods KS-CCD [16], BERT-BILSTM-CRF, and CMF-SEC [17] to extract the diagnosis and treatment knowledge related to CKD, such as diagnoses, symptoms, syndromes, and prescriptions, and standardized the diagnosis and treatment data. Combined with manual audit, we stored the standardized data in the Neo4j database to build the knowledge base of TCM diagnosis and treatment of CKD. Frequent pattern and complex network were used to mine core CHM, core syndromes, and the relationship among syndromes, prescription, and CHM. Thus, a knowledge graph of CKD diagnosis and treatment with TCM was constructed.

**2.2.1. Construction of the TCM Diagnosis and Treatment Knowledge Framework.** This study constructs a knowledge framework for TCM diagnosis and treatment based on ontology. The concept of ontology originates from the field of philosophy and is used to represent the essence of the world. In 1993, Gruber of Stanford University defined “ontology is a clear specification of conceptualization.” Based on the ontology construction method and the diagnosis and treatment approach of TCM syndrome differentiation, combined with diagnosis and treatment norms and CHM knowledge base, we comprehensively sorted the knowledge ontology of diseases, CHM, treatment principles, and methods to construct the TCM diagnosis and treatment knowledge framework by combining top-down and bottom-up methods. Figure 2 shows the knowledge framework. The top-level ontology includes diagnosis, symptom, syndrome, treatment method, prescription, and CHM. A semantic relationship is a relationship between ontology concepts. Based on the semantic relationship proposed in UMLS and the principle of TCM syndrome differentiation and treatment, the semantic relationship between concepts is determined.

In the selection of research data, only the earliest case records with identical diagnosis, syndrome, and prescription are kept. Store entities and relationships are extracted from diagnoses, symptoms, syndromes, and CHM. For one syndrome, the treatment and prescription can be the same or

different. The patient data show the characteristics of different syndrome with same diagnosis, different diagnosis with same syndrome, diagnosis and syndrome are same, but the prescription is different. Here, it is necessary to mention the important theories of Syndrome Differentiation of Traditional Chinese Medicine: “Treat different diseases with the same method, and treat same disease with different methods” [18]. “Treat different diseases with the same method” refers to different diseases are treated with same treatment due to the existence of same syndrome in the disease development process. “Treat same disease with different methods” refers to the treatment of the same disease using different methods due to the presence of different syndromes in the disease development process.

**2.2.2. CKD Knowledge Extraction.** TCM medical records are one of the main carriers of TCM inheritance and the innovation and contain rich medical knowledge. The purpose of CKD knowledge extraction is to extract diagnoses, symptoms, syndromes, prescriptions, and methods information related to CKD diagnosis and treatment from the electronic medical records of TCM inpatients, in order to realize the structuring and standardization of CKD diagnosis and treatment information and build a CKD knowledge base. Figure 3 shows the CKD knowledge extraction process.

(1) *KS-CCD.* First, based on the “WS 445-2014 Basic Dataset of Electronic Medical Records” and the user-defined keyword set, the keyword set is matched. Levenshtein string similarity algorithm and entity matching technology are the methods used in this process. After the keyword set is matched, the index of each keyword in the electronic medical record of TCM hospitalization is determined. Next, sort the keyword set. Similarity calculation is performed for keywords that do not appear. Select the keyword with the highest similarity to replace and extract the corresponding key information. If the similarity is too low and no valid key information can be extracted, the original keyword is retained and the key information corresponding to the keyword is displayed as empty. The key-value pair <key, value> is formed by the keyword set and key information to complete the structured extraction of diagnoses, symptoms, syndromes, prescriptions, and treatment methods.

(2) *BERT-BILSTM-CRF.* The BERT-BILSTM-CRF model identifies CHM and syndrome entity and extracts them. The model consists of an input layer, a BERT layer [19], a BILSTM layer (forward LSTM and backward LSTM) [20], and a CRF layer [21]. The input layer is the diagnosis and treatment information and ward round records in medical records. The BERT layer is mainly responsible for converting the input words into low dimensional vectors. The BILSTM layer is connected to the BERT layer, processing the results of the previous layer, and performing layer normalization. Finally, the CRF layer obtains the dependency relationship between labels according to the order relationship between labels, and then processes the results of the two layers to obtain the most reasonable output.

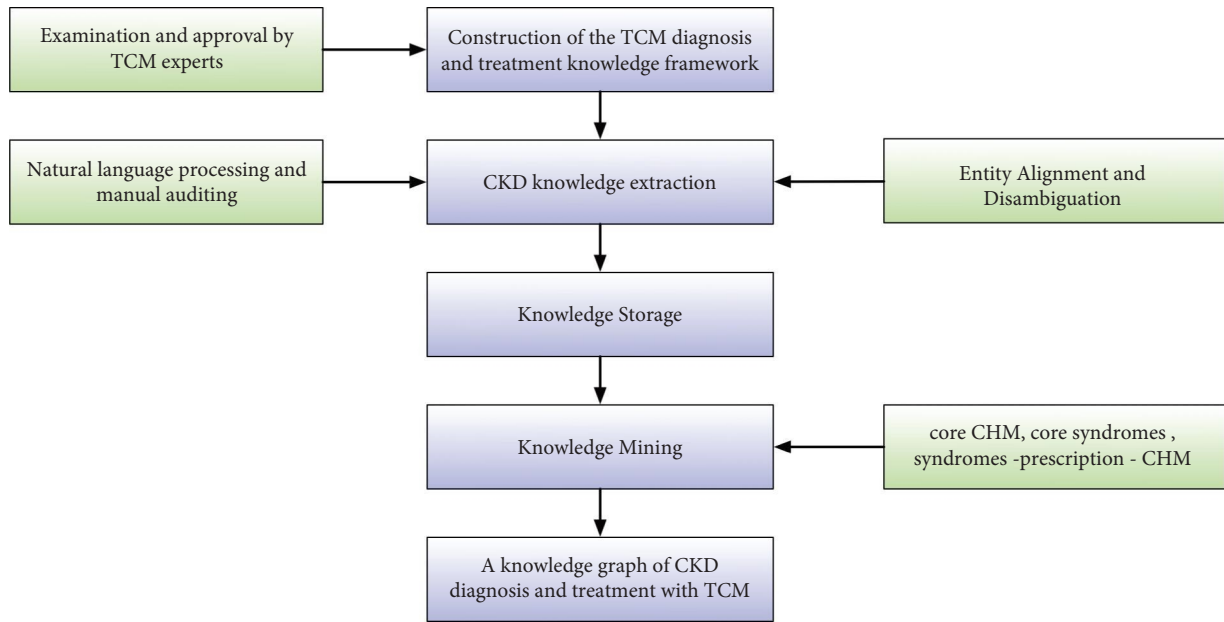


FIGURE 1: Technology roadmap for constructing the knowledge graph. It includes the construction of the TCM diagnosis and treatment knowledge framework, CKD knowledge extraction, knowledge storage, and knowledge mining. Finally, a knowledge graph of CKD diagnosis and treatment with TCM was constructed.

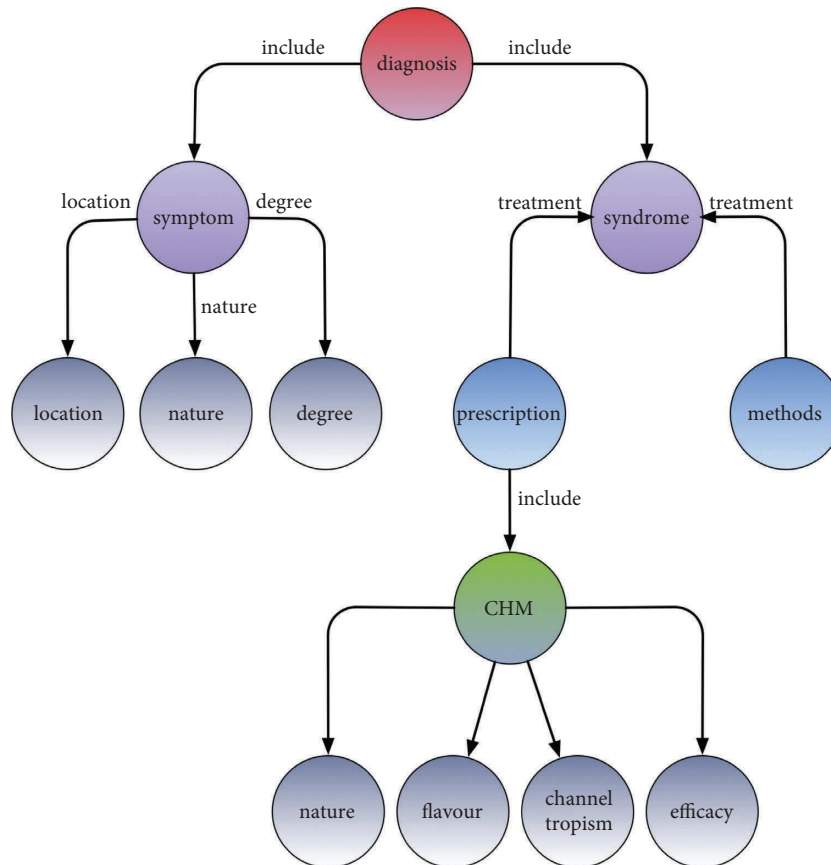


FIGURE 2: TCM diagnosis and treatment knowledge framework. The top-level ontology includes diagnosis, symptom, syndrome, treatment method, prescription, and CHM. Symptom has location, nature, and degree attributes. CHM has nature, flavour, channel tropism, and efficacy attributes.

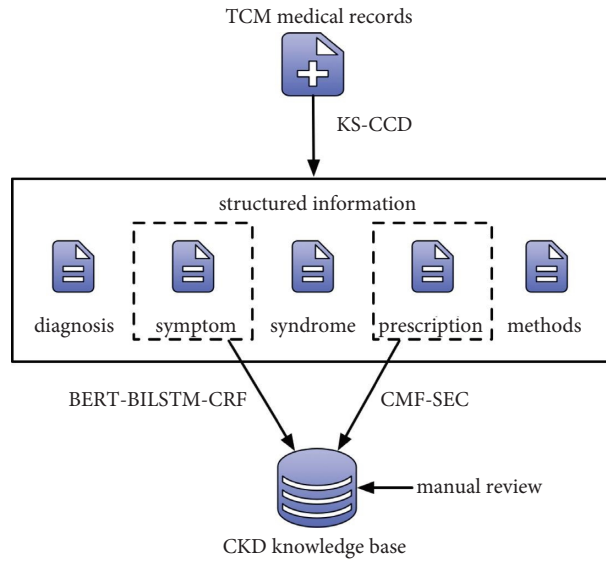


FIGURE 3: CKD knowledge extraction flowchart. We use KS-CCD to extract structured information from TCM inpatient electronic medical records, including diagnoses, symptoms, syndromes, prescriptions, and treatment methods. Then, BERT-BILSTM-CRF and CMF-SEC are used to clean the symptom and prescription, respectively. Finally, combined with manual audit, the CKD knowledge base is constructed.

Take the sentence “the patient’s tongue is pale, and there are cracks in the thin yellow coating, the syndrome still belongs to Qi and Yin deficiency, wet stasis resistance network. The treatment is to tonify Qi and nourishing Yin, clear and harmonize collaterals, and add 30g of hedyotis diffusa (Baihuasheshicao) on the top” as an example. As shown in Figure 4, first, input this sentence into BERT to generate a dynamic feature representation. BERT is composed of multiple Transformer Encoder terminals, and Transformer was first used to solve sequential tasks (such as machine translation) [22]. The encoder can obtain the semantic features of the text, and the decoder generates the corresponding text based on the feature vector for output. BERT adopts two tasks in pretraining: Masked Language Model and Next Sentence Prediction.

Masked Language Model is to mask 15% of words randomly during training and use context to predict them. Therefore, BERT can make full use of context information to generate dynamic word vectors, which solves the problem of polysemy in traditional word vector models. Next Sentence Prediction is a dichotomous task, that is, to determine whether sentence  $B$  is the following sentence of sentence  $A$ , which helps the model improve the feature representation at the sentence level. Therefore, BERT can generate dynamic representation of compound sentence features.

On this basis, in order to understand the text context semantic information, the feature vector generated by BERT is used to obtain the relative position information of the text through BILSTM so as to further improve the representation effect of the feature vector on the text. Finally, considering the global dependency between tags, in order to avoid the confusion of  $B$  and  $I$ , and different words of the same entity correspond to different types of tags, conditional random fields (CRF) are used to constrain the final output structure, and then the most reasonable sequence tags are decoded.

(3) *CMF-SEC*. The same CHM often has an official name, alias, different name, interpretation, etc. The same CHM may have different names in different regions, and different CHM may also have the same name in different regions. To solve the problem of name variety and confusion, we use CMF-SEC to clean up and standardize the prescription data, including prescription text structure standardization, prescription information completion, duplicate removal, CHM combination splitting, TCM prescription structure and synonymy standardization, and error correction.

The above results are examined manually to ensure the accuracy of the knowledge extraction results of TCM diagnosis and treatment of CKD to build the knowledge base of TCM diagnosis and treatment of CKD.

**2.2.3. Entity Alignment and Disambiguation.** We mainly aligned and disambiguated entity of CHM and syndrome. The clinical diagnosis is standardized according to the ICD-10. TCM syndrome requires entity alignment and disambiguation, because it is not uniform in specification. CHM often involves many synonyms in its actual use. To solve the above problems, we built a named entity knowledge base of CHM based on Chinese Pharmacopoeia and Chinese Materia Medica (see Section 2.1.3); based on TCM diagnosis and treatment specifications and customized extended syndrome of departments, we built a named entity knowledge base of syndrome (see Section 2.1.2).

**2.2.4. Knowledge Storage.** Integrating the TCM diagnosis and treatment knowledge framework and CKD knowledge base, the TCM diagnosis and treatment CKD knowledge is stored in the Neo4j graph database. Neo4j is a popular graph database management system implemented in Java. Compared with other graph databases, Neo4j is highlighted by its

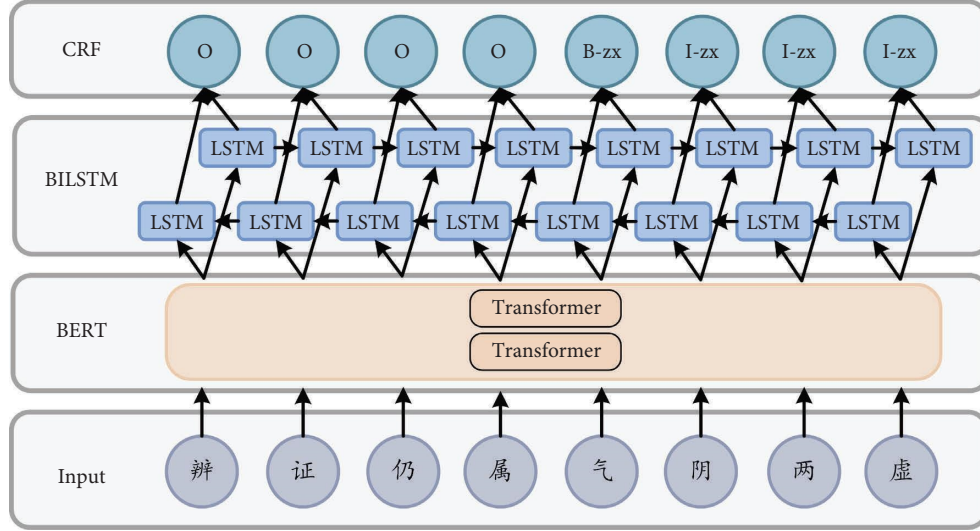


FIGURE 4: BERT-BiLSTM-CRF. Use BERT-BiLSTM-CRF for further CHM and syndrome extraction, including input layer, BERT, BiLSTM, and CRF.

active and vibrant developer communities and the number of use cases. Unusually, Neo4j uses its own query syntax, the Cypher query language [23].

**2.2.5. Knowledge Mining.** This study applied frequent pattern mining and complex network analysis in data mining to conclude the past experience and discover new knowledge of CKD diagnosis and treatment with TCM. It allows the diagnosis and treatment knowledge and deep rules to be expressed in an objective form, mines the relationship among syndrome, prescription, and medicine, and uncovers the CKD diagnosis and treatment knowledge of TCM. Frequent pattern mining is helpful for discovering the regular CHM combinations and the combinations under the certain syndrome, obtaining association rules by adjusting the support and confidence and further exploring the compatibility knowledge with high correlation; build a complex network for TCM diagnosis and treatment of CKD, extract the diagnosis, syndrome, and prescription of TCM into nodes; draw the relationship between entities into connections, and show the potential relationship among diagnosis, syndrome, prescription, syndrome, and prescription with the help of data visualization by adjusting node degree, clustering coefficient, and other statistical indicators [24].

(1) *Frequent Pattern.* Frequent pattern mining can effectively mine the implied correlation features in data and reveal the internal relationship. Association Rule  $X \rightarrow Y$  represents a frequent pattern. The evaluation indices are equations (1) and (2). The apriori algorithm is used to explore the core syndromes and core CHM in the diagnosis and treatment.

$$\text{Support}(X \rightarrow Y) = P(X, Y), \quad (1)$$

$$\text{Confidence}(X \rightarrow Y) = \frac{P(X, Y)}{P(X)}. \quad (2)$$

(2) *Complex Network.* According to the number of node types and edge types in the network, complex networks can be divided into homogeneous networks and heterogeneous networks [25, 26]. A homogeneous network refers to a network with the number of node types  $|a|=1$  and the number of edge types  $|R|=1$ . A heterogeneous network refers to a network with the number of node types  $|a|>1$  or the number of edge types  $|R|>1$ .

Let the network have  $N$  nodes, and  $k_i$  be the number of nodes directly connected to node  $v_i$ ; then, the node degrees of  $v_i$  are defined as equation (3). In an undirected network with  $N$  nodes, the degree of one node will not exceed  $N-1$ , so the normalized degree is defined as equation (4). Frequency refers to the number of joint occurrences of node  $v_i$  and node  $v_j$ . We used the Python Networkx2.6.3 to build a homogeneous complex network and heterogeneous complex network, mining the relationship among syndrome, prescription, and medicine, and assigning normalized node degree and frequency as the weights of nodes and edges.

$$C_d(v_i) = k_i, \quad (3)$$

$$C_D(v_i) = \frac{k_i}{(N-1)}. \quad (4)$$



### 3. Results

In the process of constructing the graph, some data with incomplete information are filtered, only keeping 8855 records in 1383 patients. The knowledge graph of CKD diagnosis and treatment with TCM includes 807 nodes and 10476 relationships, which are 273 diagnoses, 130 symptoms, 34 syndromes, 370 CHM nodes, 5483 diagnosis-symptom, 1349 diagnosis-syndrome, and 3644 syndrome-CHM relationships. Medical record information is stored for each patient in the knowledge graph.

**3.1. Knowledge Graph of the Diagnosis and Treatment of CKD with TCM.** Constructing the knowledge graph of TCM diagnosis and treatment of CKD based on Neo4j. Figures 5 and 6 reflect the relationship among “diagnosis, syndrome, prescription, and CHM attribute” in the knowledge graph. The red node represents diagnosis, the purple node represents TCM syndrome, the green node represents CHM, and the grey node represents the nature, taste, and meridian attributes of CHM. It contains rich semantics while maintaining a decently legible structure, which provides high retrieval efficiency.

**3.2. Application of Knowledge Graph of TCM Diagnosis and Treatment of CKD.** Based on the knowledge graph of the diagnosis and treatment of CKD with TCM, a TCM diagnosis and treatment of CKD knowledge service platform including knowledge query, knowledge mining, and other functions is designed and developed. The platform UI adopts the Bootstrap4 streaming grid system, which adopts different layouts according to the screen width to fit different screen sizes. Following the design principle of “easy to see, easy to learn, and easy to use,” the interface is simple, friendly, and convenient for users.

**3.2.1. Knowledge Query.** The “knowledge query” module of the CKD knowledge service platform for TCM diagnosis and treatment includes diagnoses, syndromes, symptoms, prescriptions knowledge query, knowledge relation visualization display, and knowledge card functions. Users can set the frequency parameter to control the result range by selecting the knowledge query of different categories of diagnoses, syndromes, symptoms, and prescriptions. Figures 7 and 8 show the query of “chronic nephritis” knowledge and set the frequency parameters of diagnosis, syndrome, symptom, and prescription as 1, 0.85, 0.95, and 0.96, respectively. The results show that the TCM syndromes related to “chronic nephritis” are “spleen-kidney qi deficiency, dampness-turbidity syndrome, dampness-heat syndrome, and blood stasis syndrome.” By clicking on “spleen-kidney qi deficiency,” the platform displays CHM related to “spleen-kidney qi deficiency,” and the knowledge card of “spleen-kidney qi deficiency” is given on the right side of the interface. The CHMs related to “spleen-kidney qi deficiency” are Astragali Radix (Huangqi), Atractylodis Macrocephalae Rhizoma (Baizhu), Pyrrosiae Folium (Shiwei), Polygoni

Cuspidati Rhizoma Et Radix (Huzhang), Eucommiae Cortex (Duzhong), Perillae Caulis (Zisugeng), Codonopsis Radix (Dangshen), Curcumae Radix (Yujin), Angelicae Sinensis Radix (Danggui), Smilacis Glabrae Rhizoma (Tufuling), Dioscoreae Nipponicae Rhizoma (Chuanshanlong), and Dipsaci Radix (Xuduan). Click “Astragali Radix (Huangqi),” which shows that the attributes are “warm in nature, sweet in taste, belonging to the kidney, spleen, lung, and heart meridian.” Additionally, the knowledge card “Astragali Radix (Huangqi)” is displayed on the right side, and “details” can be clicked for further information.

**3.2.2. Knowledge Mining.** For example, the relationship is mined among syndromes, prescriptions, and CHM of chronic nephritis. The setting support was 0.65, the confidence was 0.8, the support was 0.1, and the confidence was 0.8; the association rule method was used to mine core CHM and core syndromes, as shown in Figures 9 and 10. The core CHMs are “Astragali Radix (Huangqi), Atractylodis Macrocephalae Rhizoma (Baizhu), Folium (Shiwei), Eucommiae Cortex (Duzhong), Polygoni Cuspidati Rhizoma Et Radix (Huzhang), Perillae Caulis (Zisugeng), Curcumae Radix (Yujin), Codonopsis Radix (Dangshen), Angelicae Sinensis Radix (Danggui), Smilacis Glabrae Rhizoma (Tufuling), Dioscoreae Nipponicae Rhizoma (Chuanshanlong), Dipsaci Radix (Xuduan), Chuanxiong Rhizoma (Chuanxiong), Atractylodis Rhizoma (Cangzhu), and Hedyotis diffusa (Baihuasheshecao).” The core syndrome is “spleen and kidney qi deficiency, blood stasis, damp turbidity, and damp heat syndrome.”

We set the node degree to 10 and frequency to 500 to build a syndrome complex network and set the node degree to 32 and frequency to 6,250 to build a syndrome CHM complex network. Figures 11–13 show the complex network of syndrome and the complex network of syndrome and CHM. “Spleen and kidney qi deficiency, damp turbidity syndrome, and blood stasis syndrome” are the most frequent and closely related. The CHM corresponding to “spleen and kidney qi deficiency” is Angelicae Sinensis Radix (Danggui), Polygoni Cuspidati Rhizoma Et Radix (Huzhang), Codonopsis Radix (Dangshen), Atractylodis Macrocephalae Rhizoma (Baizhu), Astragali Radix (Huangqi), Perillae Caulis (Zisugeng), Pyrrosiae Folium (Shiwei), Curcumae Radix (Yujin), Eucommiae Cortex (Duzhong), Smilacis Glabrae Rhizoma (Tufuling), Dioscoreae Nipponicae Rhizoma (Chuanshanlong), Chuanxiong Rhizoma (Chuanxiong), and Dipsaci Radix (Xuduan). “Damp turbidity syndrome” corresponds to the CHM “Polygoni Cuspidati Rhizoma Et Radix (Huzhang), Codonopsis Radix (Dangshen), Atractylodis Macrocephalae Rhizoma (Baizhu), Astragali Radix (Huangqi), Perillae Caulis (Zisugeng), Pyrrosiae Folium (Shiwei), and Eucommiae Cortex (Duzhong).” “Blood stasis syndrome” corresponds to the CHM “Angelicae Sinensis Radix (Danggui), Polygoni Cuspidati Rhizoma Et Radix (Huzhang), Codonopsis Radix (Dangshen), Atractylodis Macrocephalae Rhizoma (Baizhu), Astragali Radix (Huangqi), Perillae Caulis (Zisugeng), Pyrrosiae Folium (Shiwei), Curcumae Radix (Yujin),

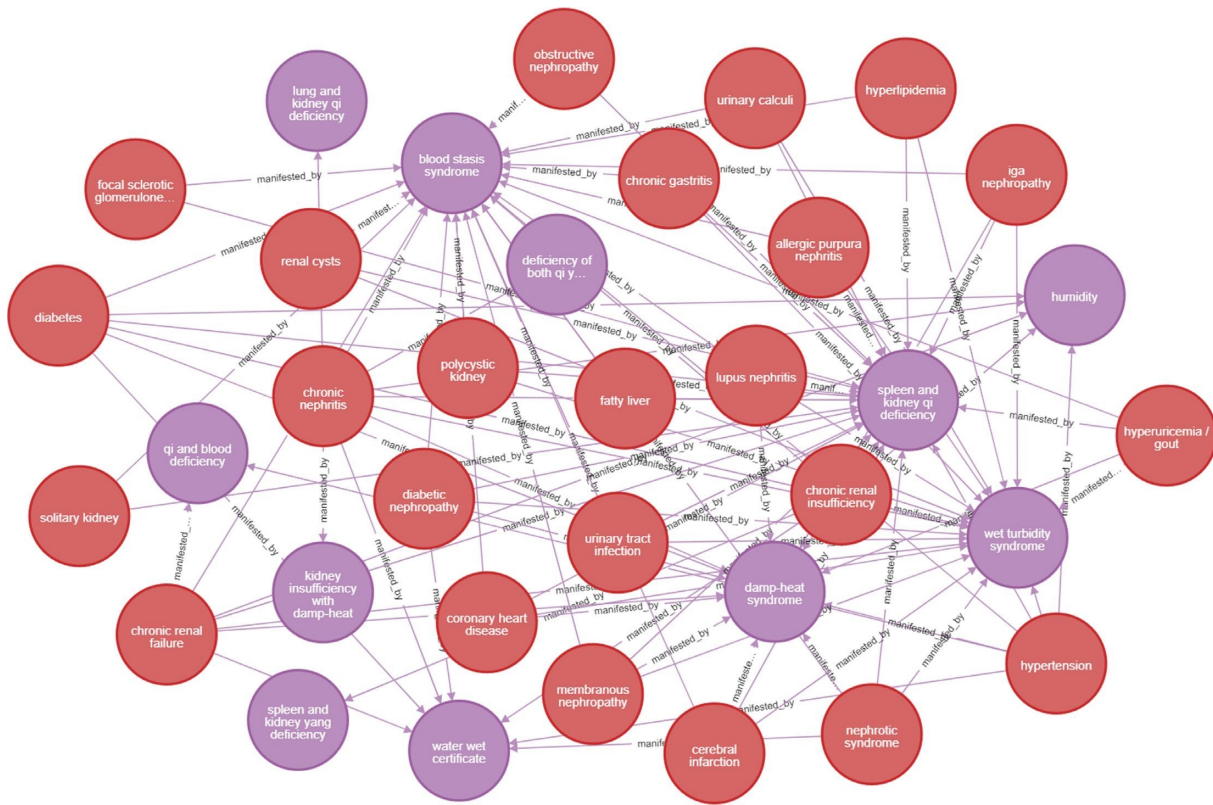


FIGURE 5: Knowledge graph (diagnosis-syndrome). The figure shows the disease diagnosis and corresponding syndrome of CKD. For example, chronic nephritis corresponds to the spleen and kidney qi deficiency, damp-heat syndrome, and other syndromes, vividly showing the relationship between diagnosis and syndrome.

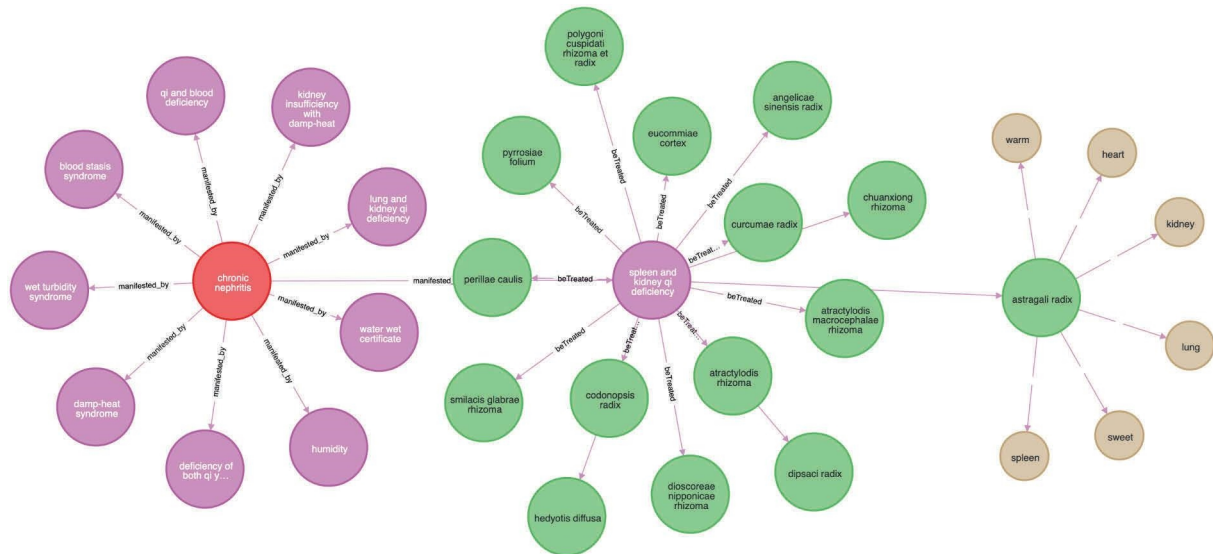


FIGURE 6: Knowledge graph (diagnosis-syndrome-CHM-CHM attributes). On the basis of Figure 5, the CHM and its attributes are further shown that are related to syndromes. Taking the spleen and kidney qi deficiency corresponding to chronic nephritis as an example, the associated CHM includes Astragali Radix (Huangqi) and Codonopsis Radix (Dangshen). Astragali Radix (Huangqi) is warm in nature and sweet in taste, belonging to the spleen, lung, kidney, and heart meridian.



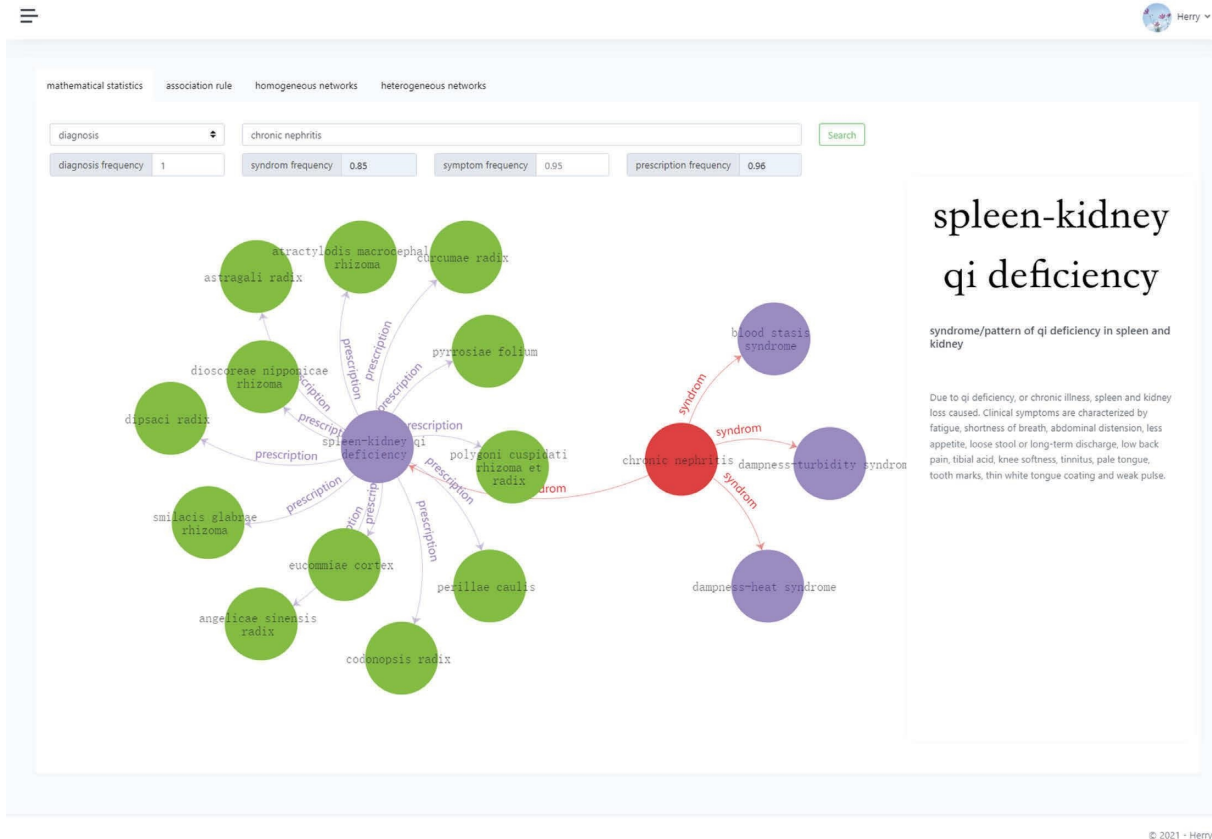


FIGURE 7: Knowledge query (spleen-kidney qi deficiency). Search the knowledge of syndrome and prescriptions related to chronic nephritis on the platform.

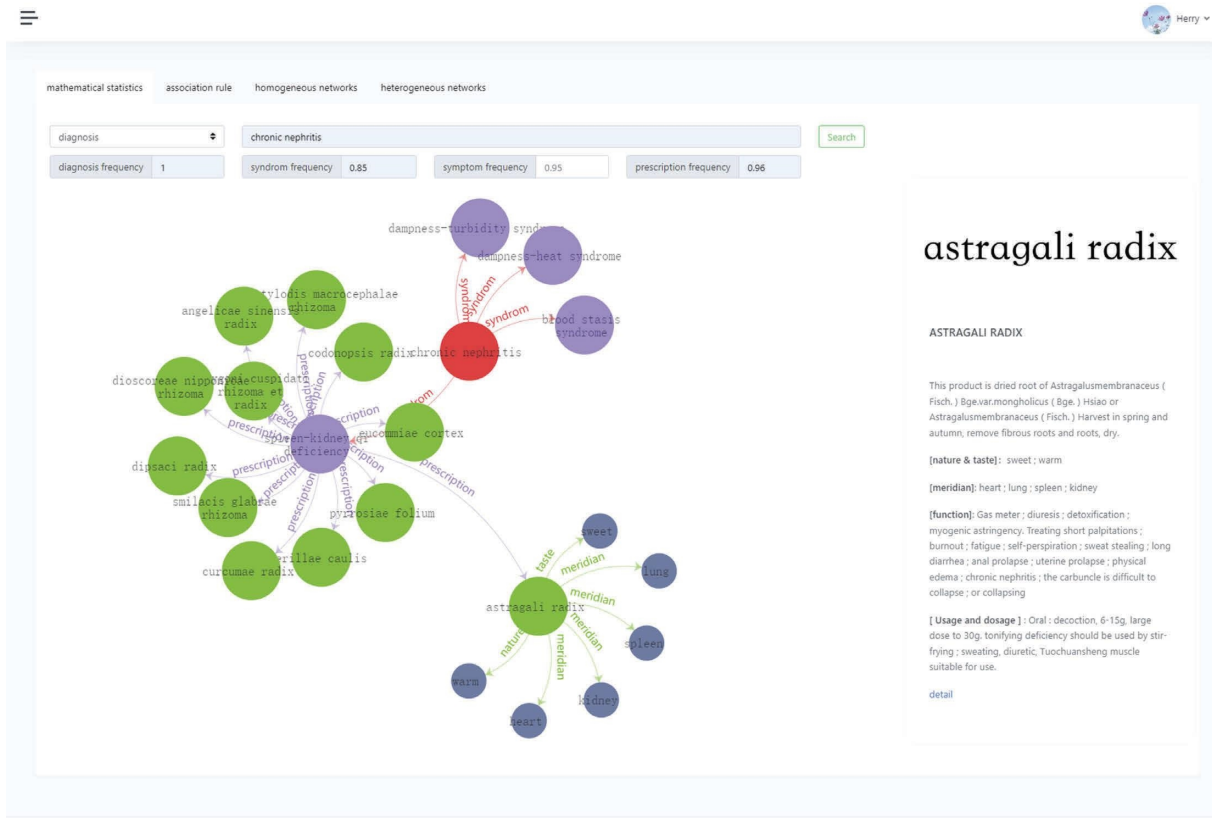
Eucommiae Cortex (Duzhong), Smilacis Glabrae Rhizoma (Tufuling), Dioscoreae Nipponicae Rhizoma (Chuanshanlong), and Dipsaci Radix (Xuduan).”

#### 4. Discussion

After years of development and update in the field of nephropathy with TCM, a variety of characteristic therapies have been formed. Years of clinical practice have proven that TCM has certain advantages in CKD prevention and treatment. As the national TCM nephropathy medical centre and regional TCM (specialist) diagnosis and treatment centre, the Nephrology Department of Jiangsu Provincial Hospital of Chinese Medicine has accumulated rich experience in the diagnosis and treatment of renal diseases. There are a group of famous doctors. The academic leader is Professor Yanqin Zou, a master of national medicine, and the discipline leader is Professor Wei Sun, a famous TCM doctor in Jiangsu Province. In this study, we selected the real clinical diagnosis and treatment data of the Nephrology Department of Jiangsu Province Hospital of Chinese Medicine and constructed a knowledge graph of TCM diagnosis and treatment of CKD with the help of knowledge graph, data mining technology, and expert consultation.

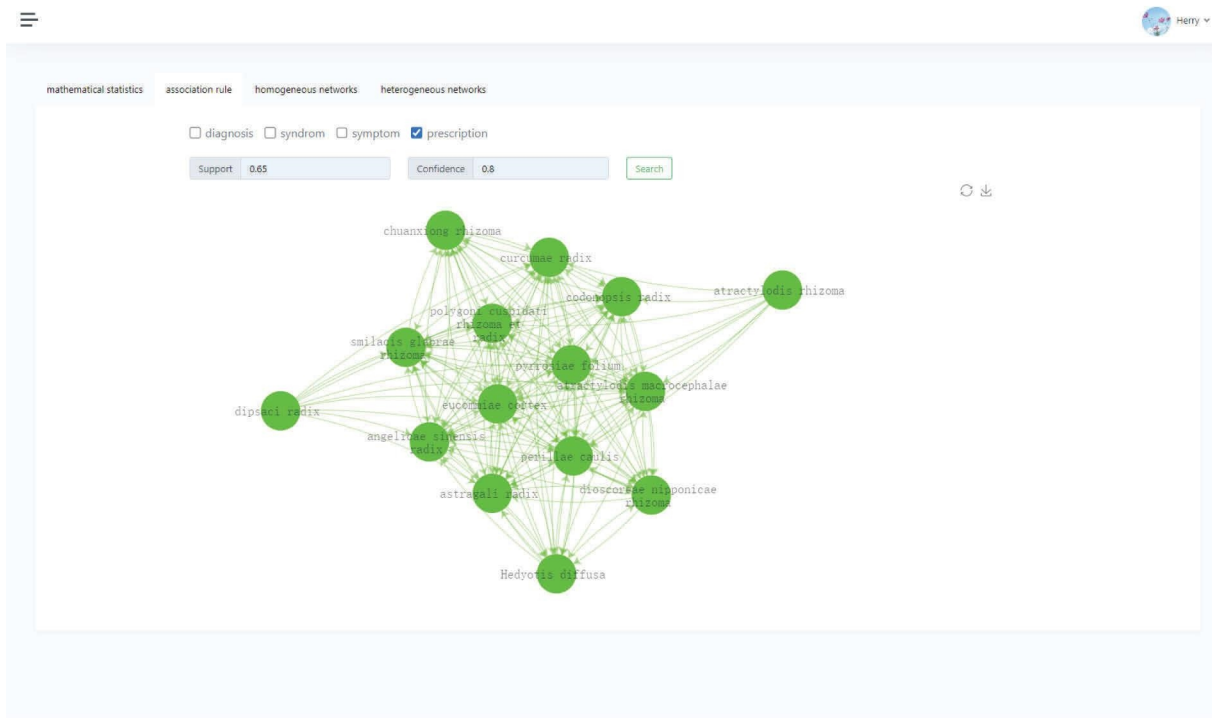
We use several knowledge extraction methods, such as KS-CCD, BERT-BILSTM-CRF, and CMF-SEC, to complete

the structural processing and cleaning of CKD clinical medical records. The medical records mainly include text information on diagnoses, symptoms, syndromes, prescriptions, and treatment methods. The entities are formed by combining manual annotation, including characters (patients and doctors), disease, symptom, diagnosis, CKD staging, pathogenesis, syndrome, treatment, prescription, CHM, CHM nature, CHM taste, CHM channel tropism, prescription, and physical and chemical examination. The semantic relationship includes “have,” “include,” “syndrome,” “pathogenesis,” “symptom,” “treatment,” “CHM nature,” “CHM taste,” and “CHM channel tropism.”. Store the extracted entities and semantic relationships, based on Neo4j database, to build a knowledge graph of TCM diagnosis and treatment CKD. Knowledge mining methods include frequent pattern and complex network analysis, and the frequent pattern is based on statistics. We used frequent pattern to find that the core CHM for CKD is “Astragali Radix (Huangqi), Atractylodes Macrocephala Rhizoma (Baizhu), Folium (Shiwei), Eucommiae Cortex (Duzhong), Polygoni Cuspidati Rhizoma Et Radix (Huzhang), Perillae Caulis (Zisugeng), Curcuma Radix (Yujin), Codonopsis Radix (Dangshen), Angelica Sinensis Radix (Danggui), Smilacis Glabrae Rhizoma (Tufuling), Dioscoreae Nipponicae Rhizoma (Chuanshanlong), Dipsaci Radix (Xuduan), Chuanxiong Rhizoma (Chuanxiong), Atractylodes Rhizoma



© 2021 - Herry

FIGURE 8: Knowledge query (Astragali Radix (Huangqi)). Show the attributes and knowledge card of Astragali Radix (Huangqi).



© 2021 - Herry

FIGURE 9: Knowledge mining-core CHM. Use association rules to mine core CHM, set the support of 0.65, and the confidence of 0.8.

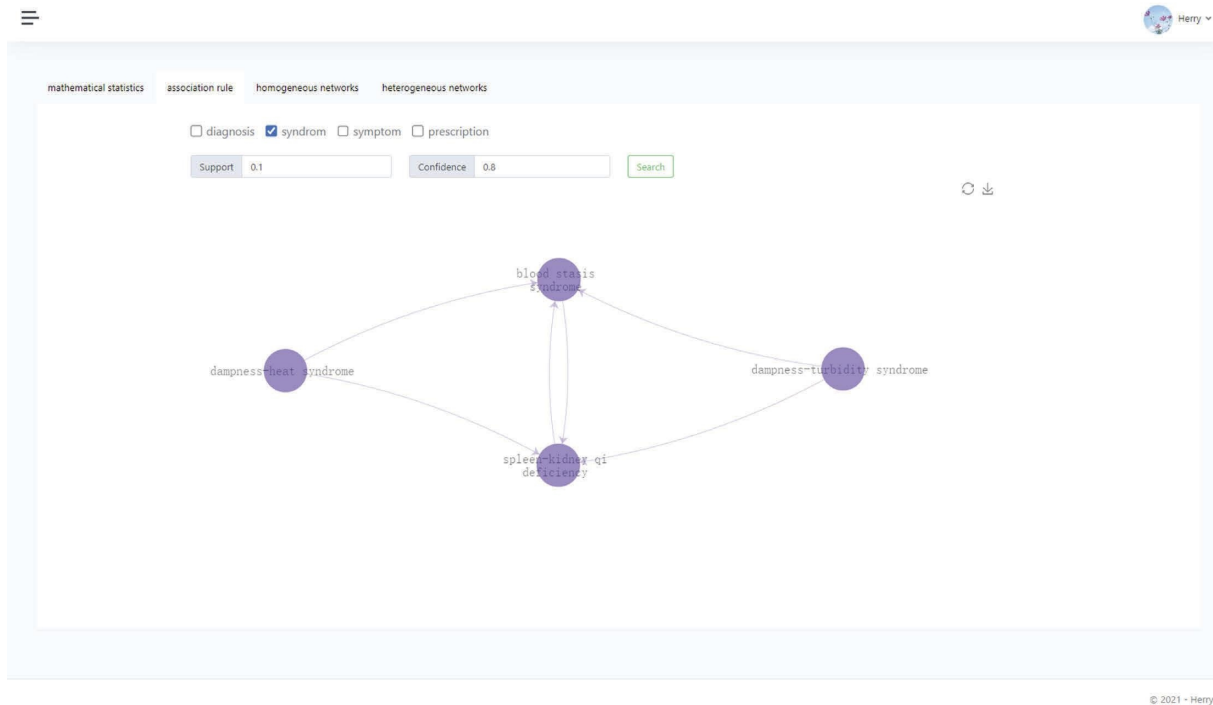


FIGURE 10: Knowledge mining-core syndromes. Use association rules to mine core certificate types, and set support of 0.1 and confidence of 0.8.

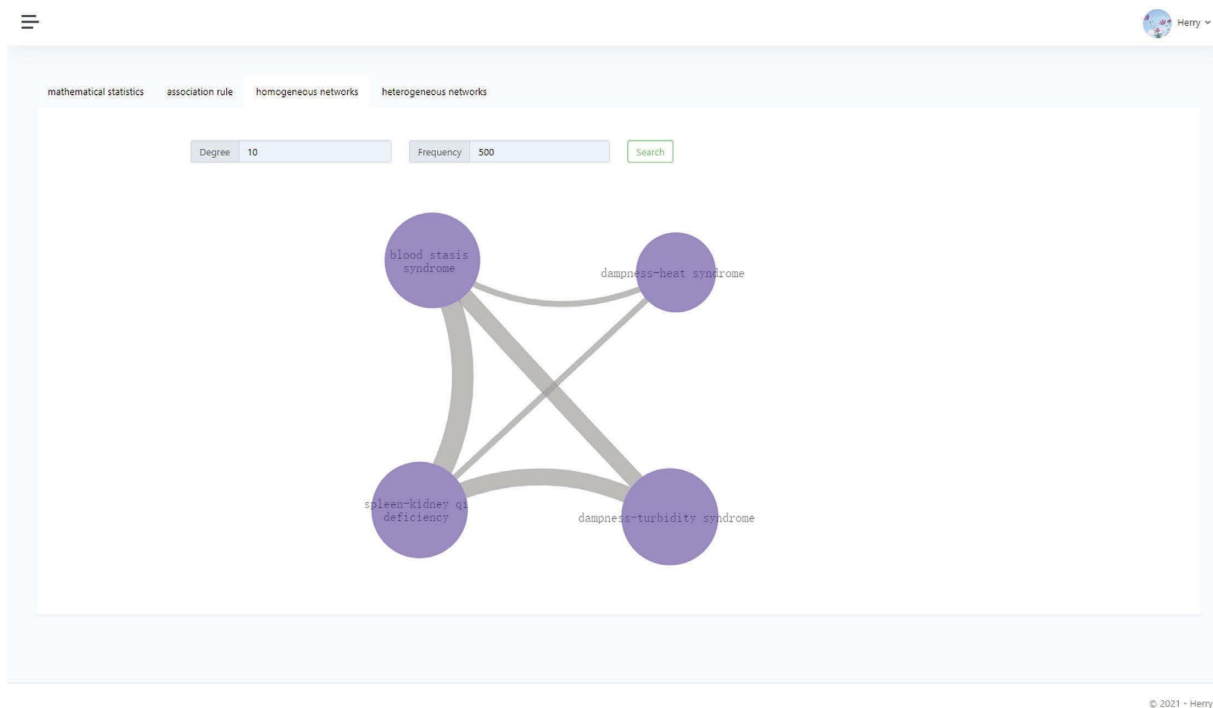
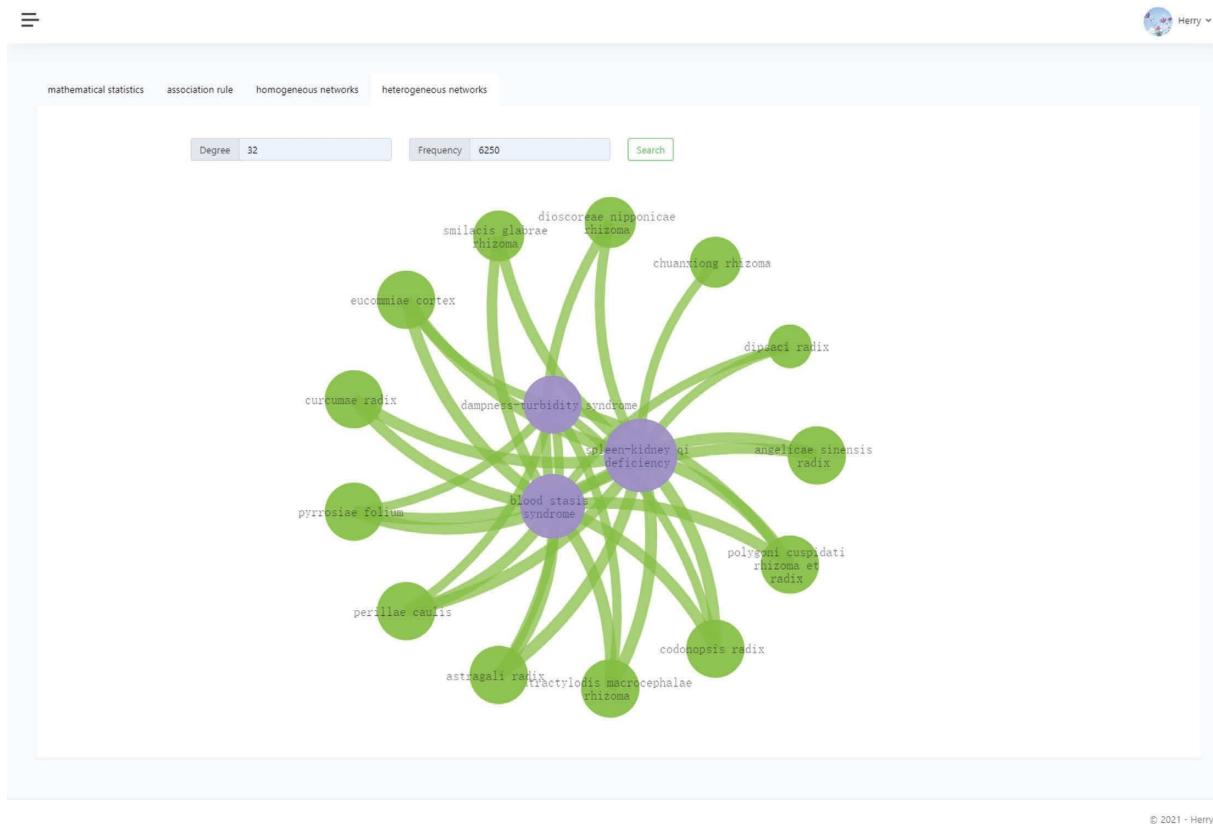


FIGURE 11: Knowledge mining-homogeneous network (syndrome). Build a syndrome complex network, set the node degree to 10, and frequency to 500.

(Cangzhu), and Hedyotis diffusa (Baihuasheshecao).” The efficacy of these CHM is mainly to tonify deficiency and also to promote water and moisture infiltration. In Chinese

nephrology, kidney deficiency is the basis of kidney disease, and the treatment of the tonifying kidney is crucial [27]. The core syndrome “spleen and kidney qi deficiency, blood stasis,



© 2021 - Henry

FIGURE 12: Knowledge mining-heterogeneous network (syndrome-CHM). Build a syndrome-CHM complex network, set node degree to 32, and frequency to 6,250.

damp turbidity, and damp heat syndrome,” are consistent with Professor Wei Sun’s treatment of CKD. He believes that the core pathogenesis of chronic kidney disease is “kidney deficiency and dampness and blood stasis.” Kidney deficiency is mainly due to kidney qi deficiency, which is the internal basis of the disease, and dampness and blood stasis are the basic links to aggravate the disease [28]. The complex network is used to analyze the core CHM corresponding to different syndromes. For example, for the “dampness turbidity syndrome,” Astragali Radix (Huangqi), Atractylodis Macrocephalae Rhizoma (Baizhu), Eucommiae Cortex (Duzhong), and Codonopsis Radix (Dangshen) are used to replenish qi and invigorate the spleen and dry dampness; Perillae Caulis (Zisugeng) is used to regulate qi and relieve pain; Pyrrosiae Folium (Shiwei) and Polygoni Cuspidati Rhizoma Et Radix (Huzhang) are used to clear away heat and promote diuresis. Xiaojuan Huang et al. [29] created the method of tonifying the kidney, clearing up, and promoting blood circulation. He often used Eucommiae Cortex (Duzhong), Dipsaci Radix (Xuduan), Visci Herba (Hujisheng), and Cuscutae Semen (Tusizi) to level and strengthen the kidney and spleen, and when used in combination with Astragali Radix (Huangqi), Codonopsis Radix (Dangshen), Atractylodis Macrocephalae Rhizoma (Baizhu), and Dioscoreae Rhizoma (Shanyao) for tonifying the kidney and spleen. Comparing the mining results of this paper with the CHM used by the two professors, we found that Astragali

Radix (Huangqi) and Atractylodis Macrocephalae Rhizoma (Baizhu) are the core CHM for tonifying the kidney.

According to the TCM development path “from clinical to clinical,” this study designed and developed a TCM diagnosis and treatment of CKD knowledge service platform based on the TCM diagnosis and treatment CKD knowledge graph, providing knowledge query and knowledge mining services for scientific researchers and medical workers. It helps users discover and refine the characteristics of TCM diagnosis and treatment of CKD as well as the correlation among “syndrome, prescription, and medicine,” clarifies the diagnosis and treatment rules of TCM treatment of CKD, and provides a scientific basis for improving treatment based on syndrome differentiation and clinical efficacy.

Our work focuses on traditional Chinese medicine data but does not integrate modern medical drug treatment data. The text information is processed in the knowledge extraction part, but the diagnosis and treatment information contained in the image is not extracted. Chinese medicine pays attention to seeing, hearing, and inquiring in the process of diagnosis and treatment, and the image data corresponding to the patient’s face-to-face diagnosis and tongue diagnosis are particularly important. In the future, we will gradually improve the knowledge graph in several aspects: (1) collect and sort out more TCM diagnosis and treatment of CKD based on the existing research, gradually form mentoring heritage context, summarize the various

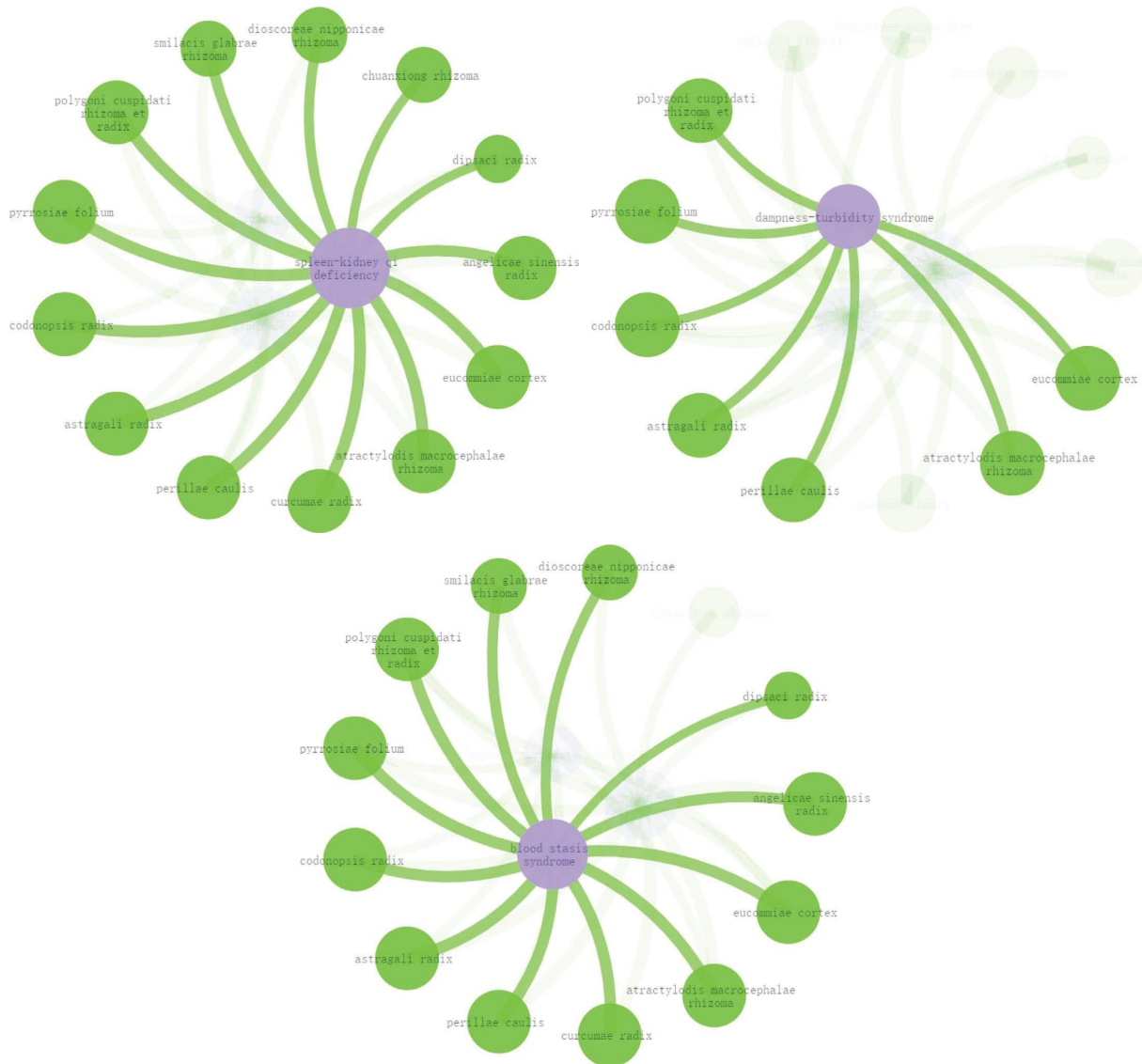


FIGURE 13: Relationship between syndrome and CHM. It shows the complex network composed of “spleen and kidney qi deficiency, damp turbidity syndrome, and blood stasis syndrome” and corresponding CHM.

stages of academic thought, and enrich the graph of TCM connotation. (2) Integrate the existing Chinese medicine and modern medical drug therapy and its molecular mechanism database such as SymMap [30] and Herb [31] to explore the relationship between the active ingredients of core Chinese medicine, genes, signaling pathways, and provide new ideas for explaining the mechanism of Chinese medicine clearly. (3) Introduce deep learning to improve the accuracy of knowledge extraction. (4) Study the methods of image acquisition, image recognition, and knowledge mining. (5) Improve the four diagnostic information and increase the image data of CKD patients such as their facial image and tongue image.

### 5. Conclusions

This study constructed a knowledge graph of CKD diagnosis and treatment with TCM, which contains relevant terms and clinical explanations such as diseases, syndromes, clinical common treatment principles, treatment methods and therapies, CHM knowledge, real clinical medical records, and other professional knowledge related to TCM diagnosis and treatment of CKD, providing strong data support for clinical diagnosis and treatment of CKD. In addition, we integrated data mining algorithms such as frequent pattern and complex network to provide knowledge mining applications for CKD diagnosis and treatment with TCM. It

contributes to inherit the clinical experience of famous veteran TCM doctors in the diagnosis and treatment of CKD and assist in the clinical diagnosis and treatment of CKD.

## Data Availability

The data used to support the findings of this study are included within the article, which are available from the corresponding authors upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Authors' Contributions

Jiadong Xie and Jiayi He contributed equally to this study. Kongfa Hu and Jiadong Xie designed the research. Jiayi He, Min Huang, and Jiadong Xie collected and processed the data. Jiayi He and Ping Xia analysed the data. Weiming He, Peipei Fang, and Chenjun Hu participated in intellectual discussions. Jiadong Xie and Jiayi He wrote the paper. All authors approved the final edited version of the manuscript.

## Acknowledgments

The authors kindly express our appreciation to open source community for the Python packages and graph databases related to the knowledge graph. The authors gratefully acknowledge the assistance of associate professor Tao Yang, his colleague at the Nanjing University of Chinese Medicine. This study was financially supported by the National Natural Science Foundation of China (81804219 to Jiadong Xie and 82074580 to Kongfa Hu), National Key Research and Development Program of China (2022YFC3502302), the Special Plan for the Development of Traditional Chinese Medicine Technology in Jiangsu Province (2020ZX13 to Weiming He), and the Natural Science Foundation of Nanjing University of Traditional Chinese Medicine (NZY81804219 to Jiadong Xie).

## References

- [1] A. C. Webster, E. V. Nagler, R. L. Morton, and P. Masson, "Chronic kidney disease," *Lancet*, vol. 389, no. 1252, pp. 1238–1252, 2017.
- [2] V. Jha, G. Garcia-Garcia, K. Iseki et al., "Chronic kidney disease: global dimension and perspectives," *The Lancet*, vol. 382, pp. 260–272, 2013.
- [3] L. Zhang, F. Wang, L. Wang et al., "Prevalence of chronic kidney disease in China: a cross-sectional survey," *Lancet*, vol. 379, pp. 815–822, 2012.
- [4] W. Chen, Y. Deng, X. Zhang et al., "Jianpi bushen tongluo decoction on spleen kidney Yang deficiency and blood stasis type, progression of IgA nephropathy," *Chinese Journal of Integrated Traditional and Western Nephrology*, vol. 13, no. 10, pp. 876–878, 2012.
- [5] W. Mao, N. Yang, L. Zhang et al., "Bupi Yishen formula versus losartan for non-diabetic stage 4 chronic kidney disease: a randomized controlled trial," *Frontiers in Pharmacology*, vol. 11, no. 2021, Article ID 627185, 2020.
- [6] Y. Wu, C. Li, L. Zhang et al., "Effectiveness of Chinese herbal medicine combined with Western medicine on deferring dialysis initiation for nondialysis chronic kidney disease stage 5 patients: a multicenter prospective nonrandomized controlled study," *Annals of Translational Medicine*, vol. 9, p. 490, 2021.
- [7] Y. Zou and G. Wang, "Zou Yunxiang, a clinical master of menghe medical school, on the experience of treating nephropathy," *Jiangsu Journal of Traditional Chinese Medicine*, vol. 48, no. 6, pp. 1–5, 2016.
- [8] M. Bai, A. Zeng, Y. Lan, and Y. Zou, "Professor Zou yanqin's experience in treating chronic kidney disease with He Luo method," *Journal of Changchun University of Chinese Medicine*, vol. 29, no. 1, pp. 70–77, 2013.
- [9] W. Sun, "On essence of kidney reinforcement and blood activation in treatment of chronic nephropathy," *Jiangsu Journal of Traditional Chinese Medicine*, vol. 43, no. 2, pp. 1–4, 2011.
- [10] X. Zhou, Z. Wu, A. Yin, L. Wu, W. Fan, and R. Zhang, "Ontology development for unified traditional Chinese medical language system," *Artificial Intelligence in Medicine*, vol. 32, no. 1, pp. 15–27, 2004.
- [11] T. Yu, J. Li, Q. Yu et al., "Knowledge graph for TCM health preservation: design, construction, and applications," *Artificial Intelligence in Medicine*, vol. 77, pp. 48–52, 2017.
- [12] X. Jia, W. Song, W. Li et al., "Research and implementation of the construction of slow obstructive pulmonary knowledge map," *Journal of Chinese Computer Systems*, vol. 41, pp. 1371–1374, 2020.
- [13] X. Li, *Construction of knowledge graph of physicians' treatment based on syndrome differentiation for patients with insomnia*, China academy of Chinese medical sciences, PhD Diss, 2018.
- [14] National Administration of Traditional Chinese Medicine, "Notice of the State Administration of TCM and the State Health Commission on printing and distributing 'the classification and code of diseases and syndromes of traditional Chinese medicine' and 'the terminology of clinical diagnosis and treatment of traditional Chinese Medicine,'" 2020, <http://yzs.satcm.gov.cn/zhengcewenjian/2020-11-23/18461.html>.
- [15] A. Levin, P. E. Stevens, R. W. Bilous et al., "Kidney Disease: improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease," *Kidney International Supplements*, vol. 3, no. 1, pp. 1–150, 2013.
- [16] J. Xie, J. He, W. He, C. Hu, K. Hu, and R. Jiang, "Research on structured information extraction method of electronic medical records of traditional Chinese medicine," in *Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1613–1616, IEEE, Seoul, Korea, December, 2020.
- [17] J. Xie, P. Fang, K. Hu, T. Yang, C. Hu, and L. Liu, "Research on data extraction and cleaning methods of clinical diagnosis and treatment of modern famous veteran doctors of TCM," *Lishizhen Medicine and Materia Medica Research*, vol. 28, no. 11, pp. 2786–2788, 2017.
- [18] Z. Zhao, K. Liu, X. Zhang, and J. Niu, "Application of the theory of treating different diseases with same method and treating same disease with different methods in prevention and treatment of multiple organ fibrosis," *Journal of Traditional Chinese Medicine*, vol. 46, no. 10, pp. 727–729, 2005.
- [19] J. D. Kenton, M. W. Chang, and L. K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the NAACL-HLT*, pp. 4171–4186, Los Angeles, CA, USA, June, 2019.



- [20] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.
- [21] K. Lewis, C. Li, M. H. Perrin et al., "Identification of urocortin III, an additional member of the corticotropin-releasing factor (CRF) family with high affinity for the CRF2 receptor," *Proceedings of the National Academy of Sciences*, vol. 98, no. 13, pp. 7570–7575, 2001.
- [22] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [23] W. Li, S. Wang, S. Wu, Z. Gu, and Y. Tian, "Performance benchmark on semantic web repositories for spatially explicit knowledge graph applications," *Computers, Environment and Urban Systems*, vol. 98, no. 2022, Article ID 101884, 2022.
- [24] M. Li, "Study on the compatibility regularity of drugs for the diagnosing and treating chronic kidney disease in the real world," Master's Thesis, Nanjing University of Chinese Medicine, 2020.
- [25] Y. Yang, Y. Guo, H. Li, and B. Deng, "Survey of community detection in heterogeneous networks," *Application Research of Computers*, vol. 35, no. 10, pp. 2881–2887, 2018.
- [26] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2017.
- [27] Y. Zou and W. Kong, "The TCM thinking on differentiating and treating chronic renal failure," *Jiangsu Journal of Traditional Chinese Medicine*, vol. 50, no. 12, pp. 1–5, 2018.
- [28] J. Zhao and W. Sun, "Professor Sun wei's experience in treating chronic kidney diseases," *Chinese Journal of Integrated Traditional and Western Nephrology*, vol. 23, no. 4, pp. 286–287, 2022.
- [29] X. Huang, W. Sun, and K. Gao, "Professor SUN wei's clinical experience in treating nephritic hematuria," *Journal of Zhejiang Chinese Medical University*, vol. 42, no. 1, pp. 81–83+86, 2018.
- [30] Y. Wu, F. Zhang, K. Yang et al., "SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping," *Nucleic Acids Research*, vol. 47, pp. D1110–D1117, 2019.
- [31] S. S. Fang, L. Dong, L. Liu et al., "HERB: a high-throughput experiment-and reference-guided database of traditional Chinese medicine," *Nucleic Acids Research*, vol. 49, pp. D1197–D1206, 2021.