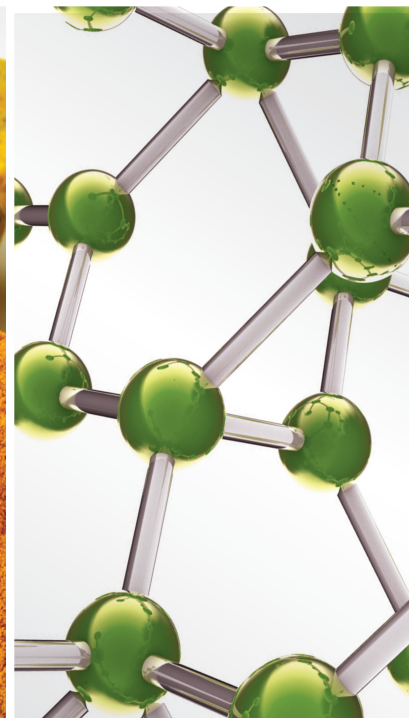# Real-World Big Data Processing and Analysis for Traditional Chinese Medicine 2022

Lead Guest Editor: Xuezhong Zhou
Guest Editors: Josiah Poon, Yonghong Peng, Tiancai Wen, and Jianan Xia

Real-World Big Data Processing and Analysis for Traditional Chinese Medicine 2022
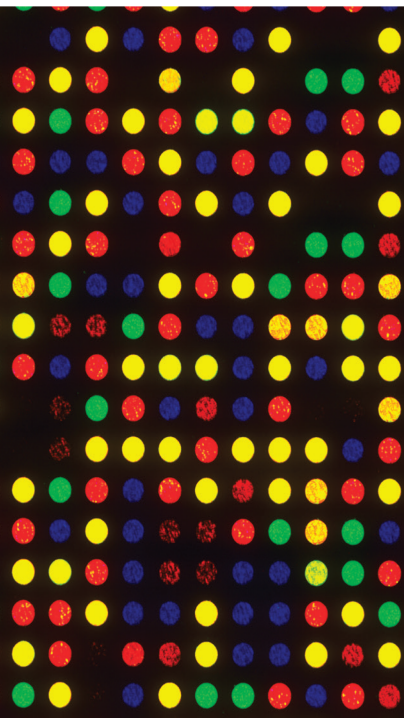
Real-World Big Data Processing and Analysis for Traditional Chinese Medicine 2022

Lead Guest Editor: Xuezhong Zhou
Guest Editors: Josiah Poon, Yonghong Peng, Tiancai Wen, and Jianan Xia

Arthur De Sá Ferreira ⬥, Brazil
Nunziatina De Tommasi ⬥, Italy
Marinella De leo ⬥, Italy
Gourav Dey ⬥, India
Dinesh Dhamecha, USA
Claudia Di Giacomo ⬥, Italy
Antonella Di Sotto ⬥, Italy
Mario Dioguardi, Italy
Jeng-Ren Duann ⬥, USA
Thomas Efferth ⬥, Germany
Abir El-Alfy, USA
Mohamed Ahmed El-Esawi ⬥, Egypt
Mohd Ramli Elvy Suhana, Malaysia
Talha Bin Emran, Japan
Roger Engel ⬥, Australia
Karim Ennouri ⬥, Tunisia
Giuseppe Esposito ⬥, Italy
Tahereh Eteraf-Oskouei, Iran
Robson Xavier Faria ⬥, Brazil
Mohammad Fattahi ⬥, Iran
Keturah R. Faurot ⬥, USA
Piergiorgio Fedeli ⬥, Italy
Laura Ferraro ⬥, Italy
Antonella Fioravanti ⬥, Italy
Carmen Formisano ⬥, Italy
Hua-Lin Fu ⬥, China
Liz G Müller ⬥, Brazil
Gabino Garrido ⬥, Chile
Safoora Gharibzadeh, Iran
Muhammad N. Ghayur ⬥, USA
Angelica Gomes ⬥, Brazil
Elena González-Burgos, Spain
Susana Gorzalczany ⬥, Argentina
Jiangyong Gu ⬥, China
Maruti Ram Gudavalli ⬥, USA
Jian-You Guo ⬥, China
Shanshan Guo, China
Narcís Gusi ⬥, Spain
Svein Haavik, Norway
Fernando Hallwass, Brazil
Gajin Han ⬥, Republic of Korea
Ihsan Ul Haq, Pakistan
Hicham Harhar ⬥, Morocco
Mohammad Hashem Hashempur ⬥, Iran
Muhammad Ali Hashmi ⬥, Pakistan

Waseem Hassan ⬥, Pakistan
Sandrina A. Heleno ⬥, Portugal
Pablo Herrero ⬥, Spain
Soon S. Hong ⬥, Republic of Korea
Md. Akil Hossain ⬥, Republic of Korea
Muhammad Jahangir Hossen ⬥, Bangladesh
Shih-Min Hsia ⬥, Taiwan
Changmin Hu ⬥, China
Tao Hu ⬥, China
Weicheng Hu ⬥, China
Wen-Long Hu, Taiwan
Xiao-Yang (Mio) Hu, United Kingdom
Sheng-Teng Huang ⬥, Taiwan
Ciara Hughes ⬥, Ireland
Attila Hunyadi ⬥, Hungary
Liaqat Hussain ⬥, Pakistan
Maria-Carmen Iglesias-Osma ⬥, Spain
Amjad Iqbal ⬥, Pakistan
Chie Ishikawa ⬥, Japan
Angelo A. Izzo, Italy
Satveer Jagwani ⬥, USA
Rana Jamous ⬥, Palestinian Authority
Muhammad Saeed Jan ⬥, Pakistan
G. K. Jayaprakasha, USA
Kyu Shik Jeong, Republic of Korea
Leopold Jirovetz ⬥, Austria
Jeeyoun Jung ⬥, Republic of Korea
Nurkhalida Kamal ⬥, Saint Vincent and the Grenadines
Atsushi Kameyama ⬥, Japan
Kyungsu Kang, Republic of Korea
Wenyi Kang ⬥, China
Shao-Hsuan Kao ⬥, Taiwan
Nasiara Karim ⬥, Pakistan
Morimasa Kato ⬥, Japan
Kumar Katragunta ⬥, USA
Deborah A. Kennedy ⬥, Canada
Washim Khan, USA
Bonglee Kim ⬥, Republic of Korea
Dong Hyun Kim ⬥, Republic of Korea
Junghyun Kim ⬥, Republic of Korea
Kyungho Kim, Republic of Korea
Yun Jin Kim ⬥, Malaysia
Yoshiyuki Kimura ⬥, Japan

Nebojša Kladar, Serbia
Mi Mi Ko, Republic of Korea
Toshiaki Kogure, Japan
Malcolm Koo, Taiwan
Yu-Hsiang Kuan, Taiwan
Robert Kubina, Poland
Chan-Yen Kuo, Taiwan
Kuang C. Lai, Taiwan
King Hei Stanley Lam, Hong Kong
Fanuel Lampiao, Malawi
Ilaria Lampronti, Italy
Mario Ledda, Italy
Harry Lee, China
Jeong-Sang Lee, Republic of Korea
Ju Ah Lee, Republic of Korea
Kyu Pil Lee, Republic of Korea
Namhun Lee, Republic of Korea
Sang Yeoup Lee, Republic of Korea
Ankita Leekha, USA
Christian Lehmann, Canada
George B. Lenon, Australia
Marco Leonti, Italy
Hua Li, China
Min Li, China
Xing Li, China
Xuqi Li, China
Yi-Rong Li, Taiwan
Vuanghao Lim, Malaysia
Bi-Fong Lin, Taiwan
Ho Lin, Taiwan
Shuibin Lin, China
Kuo-Tong Liou, Taiwan
I-Min Liu, Taiwan
Suhuan Liu, China
Xiaosong Liu, Australia
Yujun Liu, China
Emilio Lizarraga, Argentina
Monica Loizzo, Italy
Nguyen Phuoc Long, Republic of Korea
Zaira López, Mexico
Chunhua Lu, China
Ângelo Luís, Portugal
Anderson Luiz-Ferreira, Brazil
Ivan Luzardo Luzardo-Ocampo, Mexico

Michel Mansur Machado, Brazil
Filippo Maggi, Italy
Juraj Majtan, Slovakia
Toshiaki Makino, Japan
Nicola Malafronte, Italy
Giuseppe Malfa, Italy
Francesca Mancianti, Italy
Carmen Mannucci, Italy
Juan M. Manzaneque, Spain
Fatima Martel, Portugal
Carlos H. G. Martins, Brazil
Maulidiani Maulidiani, Malaysia
Andrea Maxia, Italy
Avijit Mazumder, India
Isac Medeiros, Brazil
Ahmed Mediani, Malaysia
Lewis Mehl-Madrona, USA
Ayikoé Guy Mensah-Nyagan, France
Oliver Micke, Germany
Maria G. Miguel, Portugal
Luigi Milella, Italy
Roberto Miniero, Italy
Letteria Minutoli, Italy
Prashant Modi, India
Daniel Kam-Wah Mok, Hong Kong
Changjong Moon, Republic of Korea
Albert Moraska, USA
Mark Moss, United Kingdom
Yoshiharu Motoo, Japan
Yoshiki Mukudai, Japan
Sakthivel Muniyan, USA
Saima Muzammil, Pakistan
Benoit Banga N'guessan, Ghana
Massimo Nabissi, Italy
Siddavaram Nagini, India
Takao Namiki, Japan
Srinivas Nammi, Australia
Krishnadas Nandakumar, India
Vitaly Napadow, USA
Edoardo Napoli, Italy
Jorddy Neves Cruz, Brazil
Marcello Nicoletti, Italy
Eliud Nyaga Mwaniki Njagi, Kenya
Cristina Nogueira, Brazil

Sakineh Kazemi Noureini (iD), Iran
Rômulo Dias Novaes, Brazil
Martin Offenbaecher (iD), Germany
Oluwafemi Adeleke Ojo (iD), Nigeria
Olufunmiso Olusola Olajuyigbe (iD), Nigeria
Luís Flávio Oliveira, Brazil
Mozaniel Oliveira (iD), Brazil
Atolani Olubunmi (iD), Nigeria
Abimbola Peter Oluyori (iD), Nigeria
Timothy Omara, Austria
Chiagoziem Anariochi Otuechere (iD), Nigeria
Sokcheon Pak (iD), Australia
Antônio Palumbo Jr, Brazil
Zongfu Pan (iD), China
Siyaram Pandey (iD), Canada
Niranjan Parajuli (iD), Nepal
Gunhyuk Park (iD), Republic of Korea
Wansu Park (iD), Republic of Korea
Rodolfo Parreira (iD), Brazil
Mohammad Mahdi Parvizi (iD), Iran
Luiz Felipe Passero (iD), Brazil
Mitesh Patel, India
Claudia Helena Pellizzon (iD), Brazil
Cheng Peng, Australia
Weijun Peng (iD), China
Sonia Piacente, Italy
Andrea Pieroni (iD), Italy
Haifa Qiao (iD), USA
Cláudia Quintino Rocha (iD), Brazil
DANIELA RUSSO (iD), Italy
Muralidharan Arumugam Ramachandran,
Singapore
Manzoor Rather (iD), India
Miguel Rebollo-Hernanz (iD), Spain
Gauhar Rehman, Pakistan
Daniela Rigano (iD), Italy
José L. Rios, Spain
Francisca Rius Diaz, Spain
Eliana Rodrigues (iD), Brazil
Maan Bahadur Rokaya (iD), Czech Republic
Mariangela Rondanelli (iD), Italy
Antonietta Rossi (iD), Italy
Mi Heon Ryu (iD), Republic of Korea
Bashar Saad (iD), Palestinian Authority
Sabiu Saheed, South Africa

Mohamed Z.M. Salem (iD), Egypt
Avni Sali, Australia
Andreas Sandner-Kiesling, Austria
Manel Santafe (iD), Spain
José Roberto Santin (iD), Brazil
Tadaaki Satou (iD), Japan
Roland Schoop, Switzerland
Sindy Seara-Paz, Spain
Veronique Seidel (iD), United Kingdom
Vijayakumar Sekar (iD), China
Terry Selfe (iD), USA
Arham Shabbir (iD), Pakistan
Suzana Shahar, Malaysia
Wen-Bin Shang (iD), China
Xiaofei Shang (iD), China
Ali Sharif (iD), Pakistan
Karen J. Sherman (iD), USA
San-Jun Shi (iD), China
Insop Shim (iD), Republic of Korea
Maria Im Hee Shin, China
Yukihiro Shoyama, Japan
Morry Silberstein (iD), Australia
Samuel Martins Silvestre (iD), Portugal
Preet Amol Singh, India
Rajeev K Singla (iD), China
Kuttulebbai N. S. Sirajudeen (iD), Malaysia
Slim Smaoui (iD), Tunisia
Eun Jung Sohn (iD), Republic of Korea
Maxim A. Solovchuk (iD), Taiwan
Young-Jin Son (iD), Republic of Korea
Chengwu Song (iD), China
Vanessa Steenkamp (iD), South Africa
Annarita Stringaro (iD), Italy
Keiichiro Sugimoto (iD), Japan
Valeria Sulsen (iD), Argentina
Zewei Sun (iD), China
Sharifah S. Syed Alwi (iD), United Kingdom
Orazio Taglialatela-Scafati (iD), Italy
Takashi Takeda (iD), Japan
Gianluca Tamagno (iD), Ireland
Hongxun Tao, China
Jun-Yan Tao (iD), China
Lay Kek Teh (iD), Malaysia
Norman Temple (iD), Canada

Kamani H. Tennekoon, Sri Lanka
Seong Lin Teoh, Malaysia
Menaka Thounaojam, USA
Jinhui Tian, China
Zipora Tietel, Israel
Loren Toussaint, USA
Riaz Ullah, Saudi Arabia
Philip F. Uzor, Nigeria
Luca Vanella, Italy
Antonio Vassallo, Italy
Cristian Vergallo, Italy
Miguel Vilas-Boas, Portugal
Aristo Vojdani, USA
Yun WANG, China
QIBIAO WU, Macau
Abraham Wall-Medrano, Mexico
Chong-Zhi Wang, USA
Guang-Jun Wang, China
Jinan Wang, China
Qi-Rui Wang, China
Ru-Feng Wang, China
Shu-Ming Wang, USA
Ting-Yu Wang, China
Xue-Rui Wang, China
Youhua Wang, China
Kenji Watanabe, Japan
Jintanaporn Wattanathorn, Thailand
Silvia Wein, Germany
Katarzyna Winska, Poland
Sok Kuan Wong, Malaysia
Christopher Worsnop, Australia
Jih-Huah Wu, Taiwan
Sijin Wu, China
Xian Wu, USA
Zuoqi Xiao, China
Rafael M. Ximenes, Brazil
Guoqiang Xing, USA
JiaTuo Xu, China
Mei Xue, China
Yong-Bo Xue, China
Haruki Yamada, Japan
Nobuo Yamaguchi, Japan
Junqing Yang, China
Longfei Yang, China

Mingxiao Yang, Hong Kong
Qin Yang, China
Wei-Hsiung Yang, USA
Swee Keong Yeap, Malaysia
Albert S. Yeung, USA
Ebrahim M. Yimer, Ethiopia
Yoke Keong Yong, Malaysia
Fadia S. Youssef, Egypt
Zhilong Yu, Canada
RONGJIE ZHAO, China
Sultan Zahiruddin, USA
Armando Zarrelli, Italy
Xiaobin Zeng, China
Y Zeng, China
Fangbo Zhang, China
Jianliang Zhang, China
Jiu-Liang Zhang, China
Mingbo Zhang, China
Jing Zhao, China
Zhangfeng Zhong, Macau
Guoqi Zhu, China
Yan Zhu, USA
Suzanna M. Zick, USA
Stephane Zingue, Cameroon

Contents

Research Article

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

Jiadong Xie [1,2] **Jiayi He** [1] **Ping Xia,**[3] **Weiming He** [4] **Min Huang,**[1] **Peipei Fang,**[1] **Chenjun Hu,**[1] **and Kongfa Hu** [1,5]

[1]*School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China*
[2]*Institute of Literature in Chinese Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China*
[3]*Department of Traditional Chinese Medicine, First Affiliated Hospital of Soochow University, Suzhou 215006, China*
[4]*Division of Nephrology, Affiliated Hospital of Nanjing University of Chinese Medicine,*
Jiangsu Province Hospital of Chinese Medicine, Nanjing 210029, China
[5]*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

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.
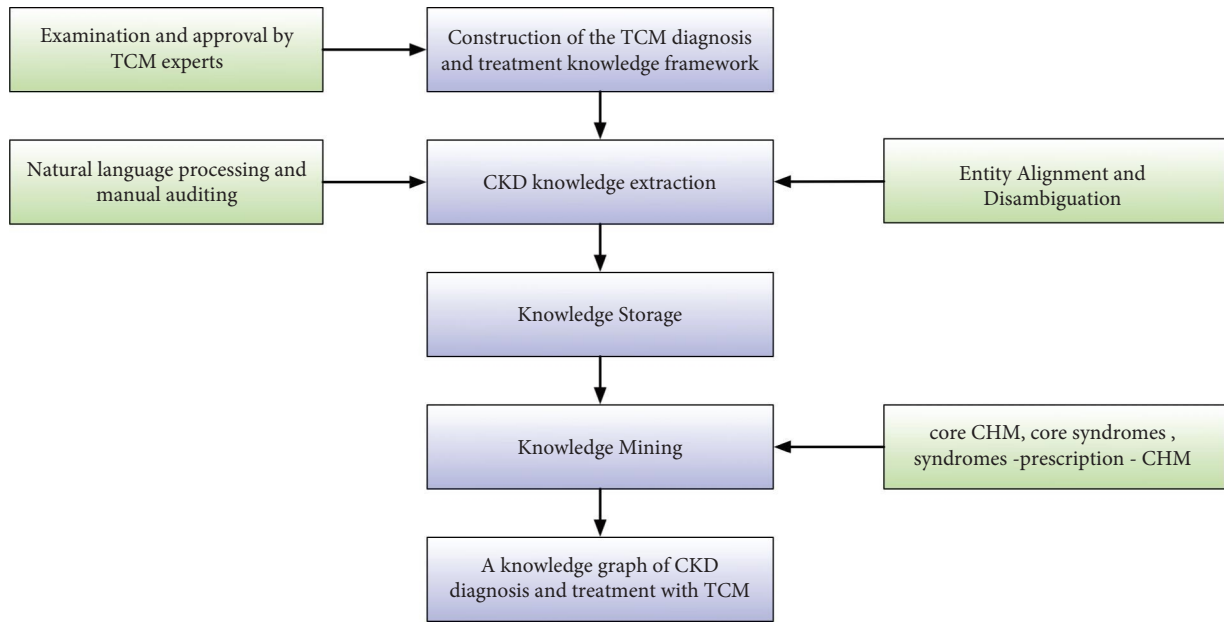
Figure 1: Technology roadmap for constructing the knowledge graph. It includes the construction of the TCM diagnosis and treatment knowledge framework, CKD knowledge extraction, knowledge storage, and knowledge mining. Finally, a knowledge graph of CKD diagnosis and treatment with TCM was constructed.
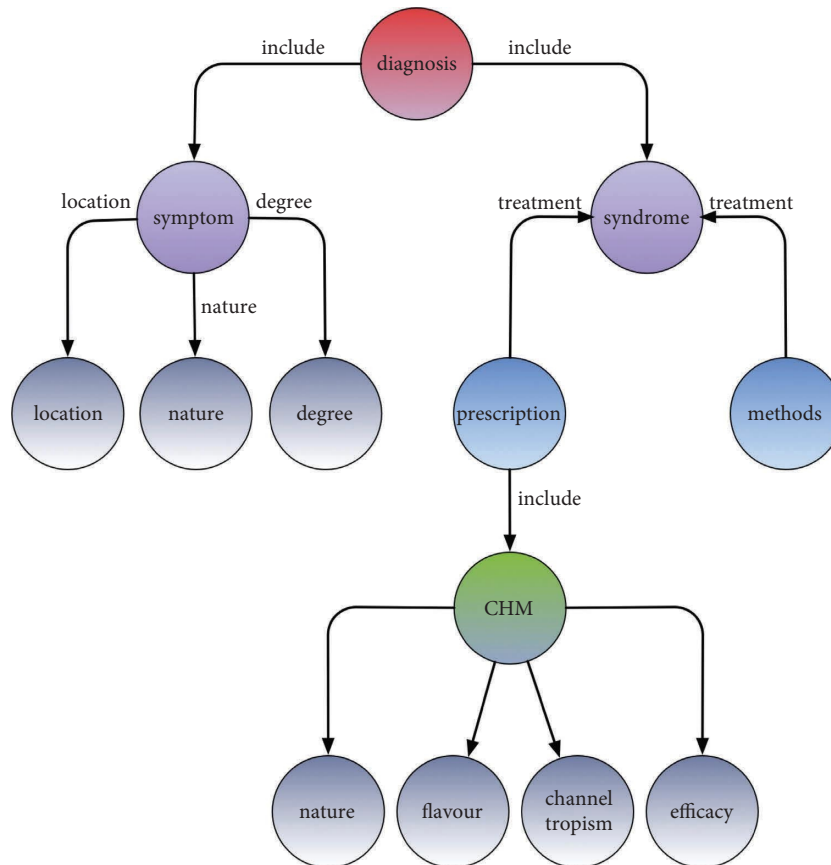
Figure 2: TCM diagnosis and treatment knowledge framework. The top-level ontology includes diagnosis, symptom, syndrome, treatment method, prescription, and CHM. Symptom has location, nature, and degree attributes. CHM has nature, flavour, channel tropism, and efficacy attributes.
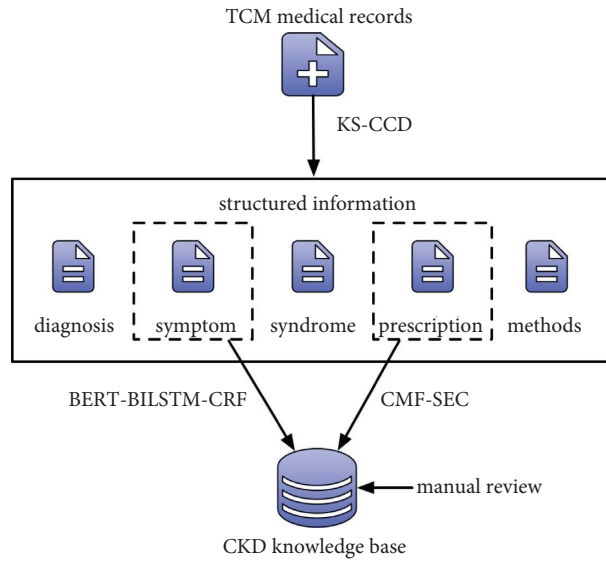
FIGURE 3: CKD knowledge extraction flowchart. We use KS-CCD to extract structured information from TCM inpatient electronic medical records, including diagnoses, symptoms, syndromes, prescriptions, and treatment methods. Then, BERT-BILSTM-CRF and CMF-SEC are used to clean the symptom and prescription, 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 (Baihuasheshecao) 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 BISLTM 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
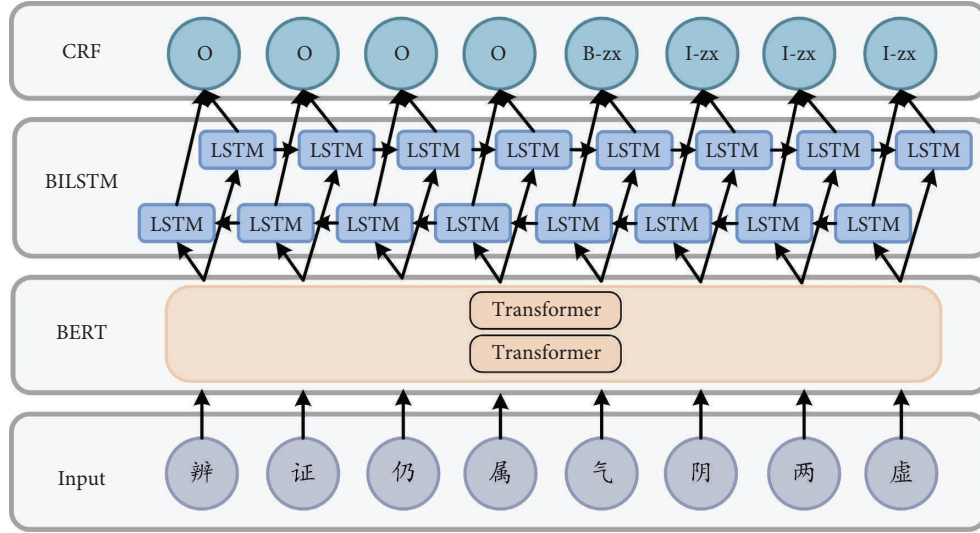
Figure 4: BERT-BILSTM-CRF. Use BERT-BILSTM-CRF for further CHM and syndrome extraction, including input layer, BERT, BILSTM, and CRF.

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 $X \longrightarrow Y$ 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.

$$\text{Support} (X \longrightarrow Y) = P(X, Y), \tag{1}$$

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

(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.

$$C_d (v_i) = k_i, \tag{3}$$

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

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),
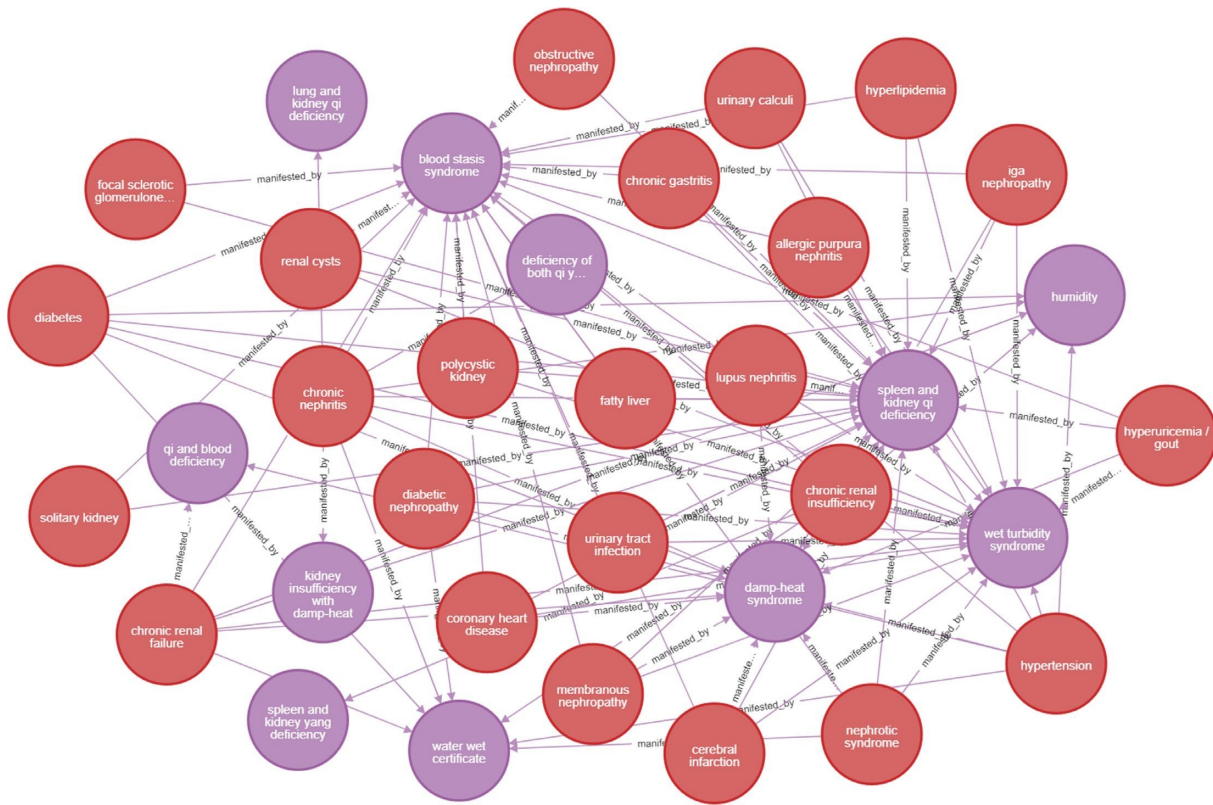
FIGURE 5: Knowledge graph (diagnosis-syndrome). The figure shows the disease diagnosis and corresponding syndrome of CKD. For example, chronic nephritis corresponds to the spleen and kidney qi deficiency, damp-heat syndrome, and other syndromes, vividly showing the relationship between diagnosis and syndrome.
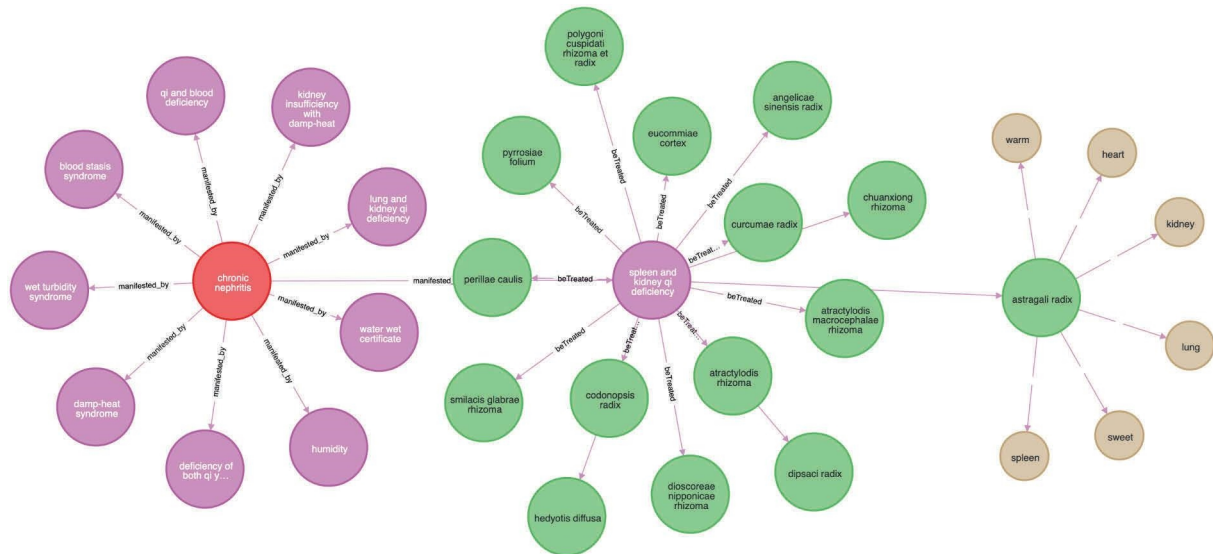
FIGURE 6: Knowledge graph (diagnosis-syndrome-CHM–CHM attributes). On the basis of Figure 5, the CHM and its attributes are further shown that are related to syndromes. Taking the spleen and kidney qi deficiency corresponding to chronic nephritis as an example, the associated CHM includes Astragali Radix (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.
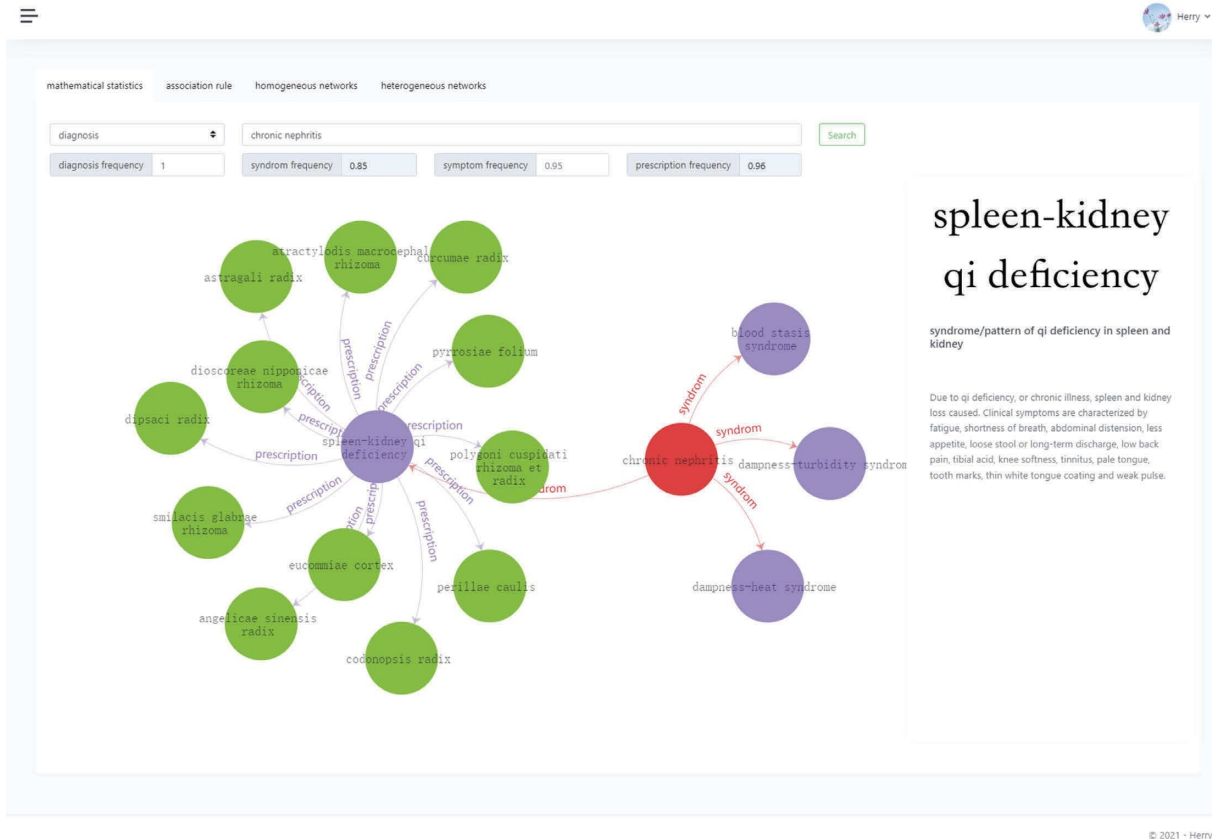
FIGURE 7: Knowledge query (spleen-kidney qi deficiency). Search the knowledge of syndrome and prescriptions related to chronic nephritis on the platform.

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 ProfessorWei 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), 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
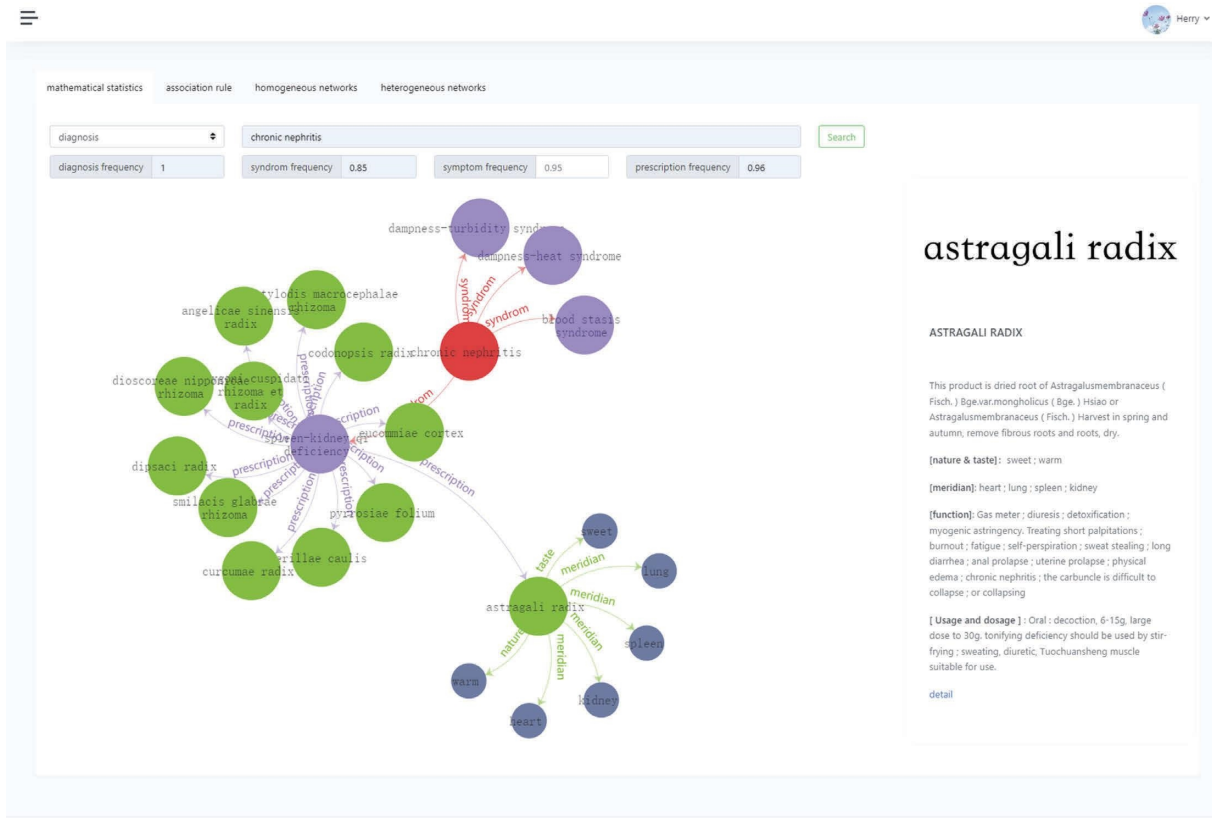
Figure 8: Knowledge query (Astragali Radix (Huangqi)). Show the attributes and knowledge card of Astragali Radix (Huangqi).
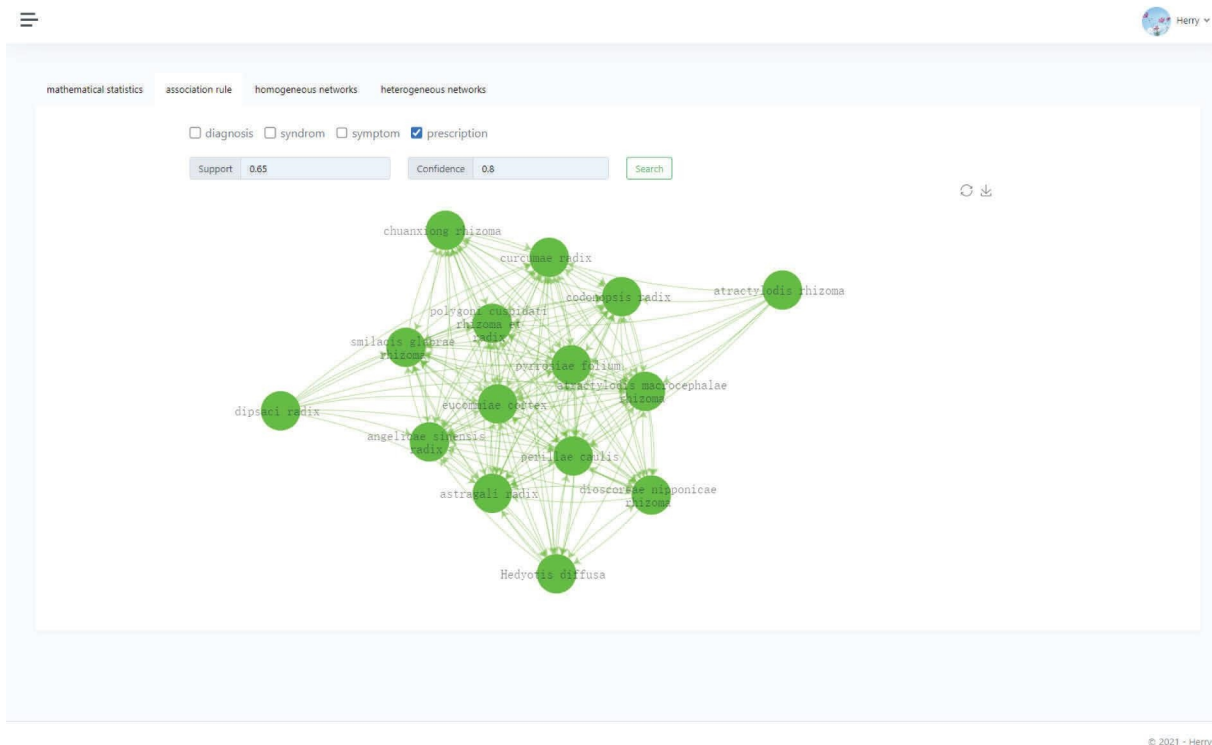
Figure 9: Knowledge mining-core CHM. Use association rules to mine core CHM, set the support of 0.65, and the confidence of 0.8.
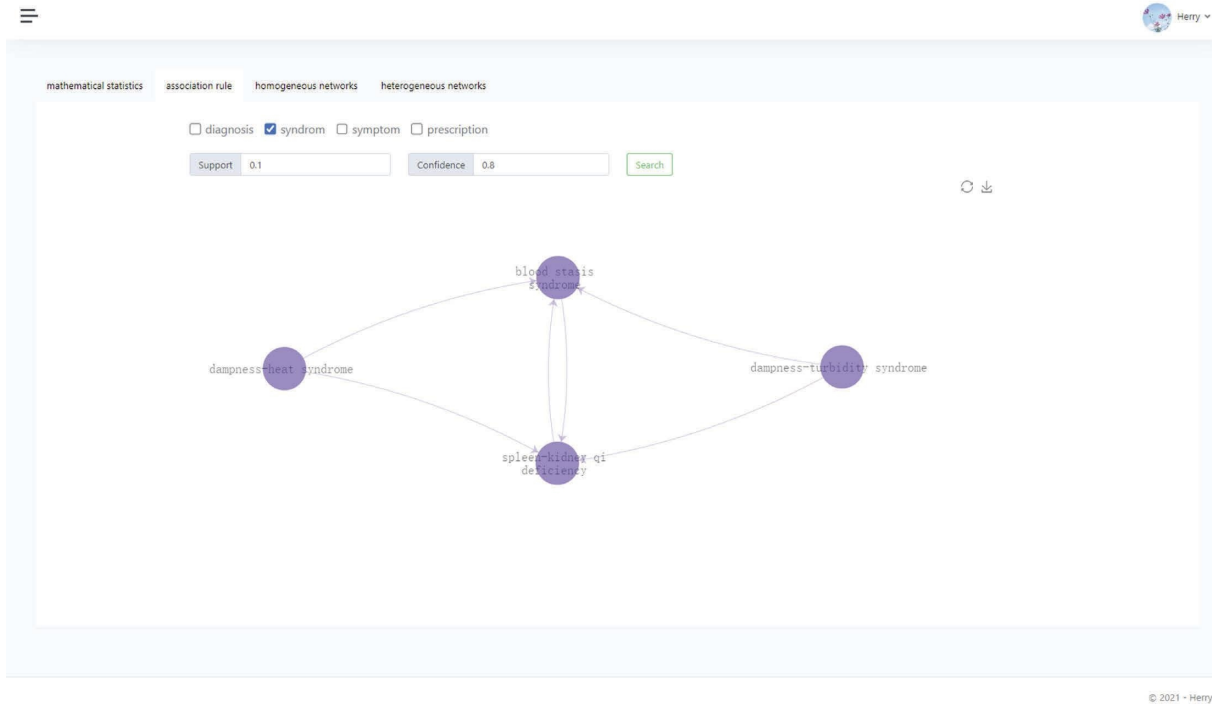
Figure 10: Knowledge mining-core syndromes. Use association rules to mine core certificate types, and set support of 0.1 and confidence of 0.8.
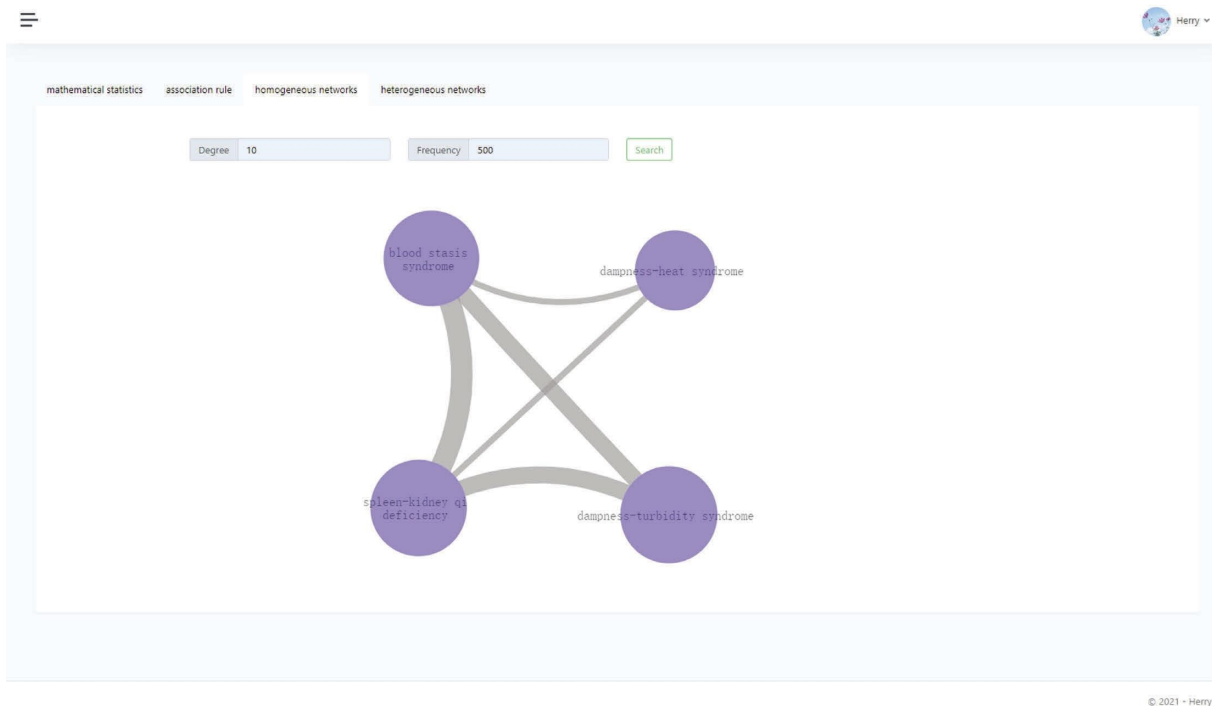
Figure 11: Knowledge mining-homogeneous network (syndrome). Build a syndrome complex network, set the node degree to 10, and frequency to 500.

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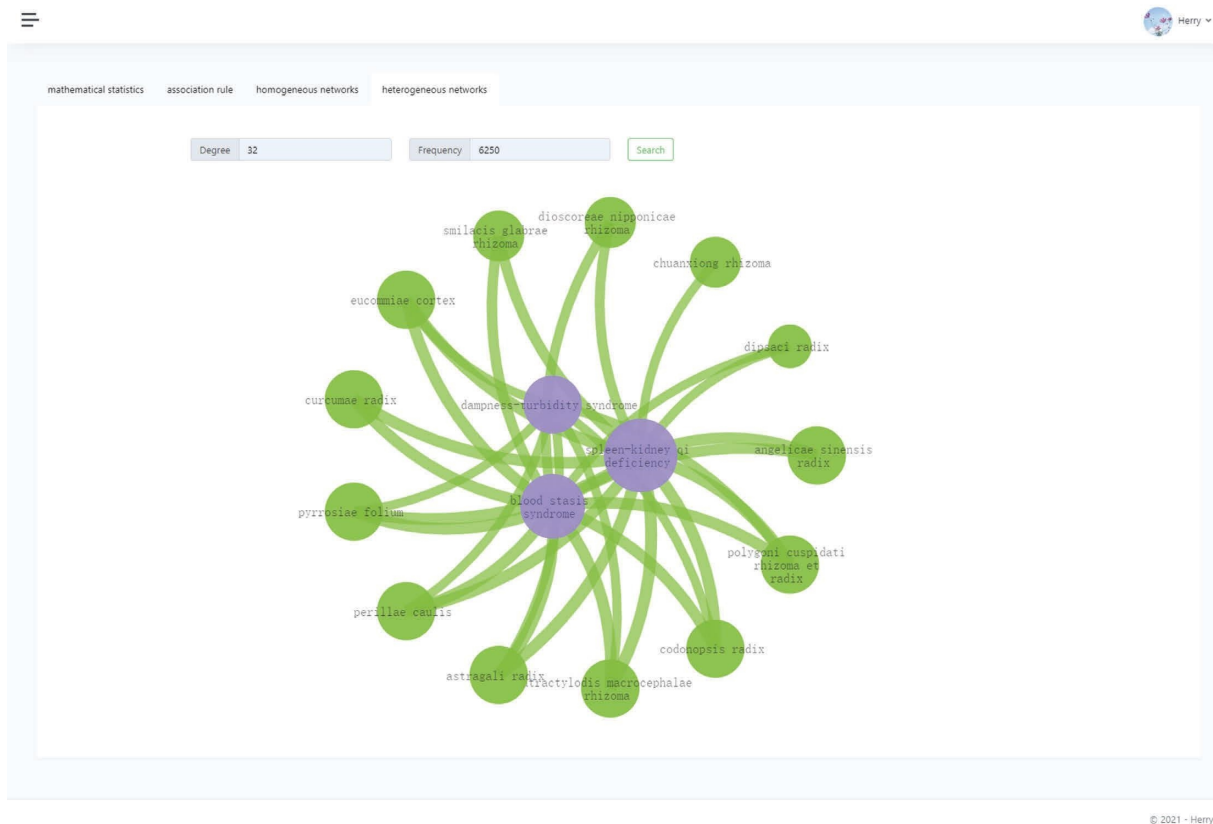
Figure 12: Knowledge mining-heterogeneous network (syndrome-CHM). Build a syndrome-CHM complex network, set node degree to 32, and frequency to 6,250.

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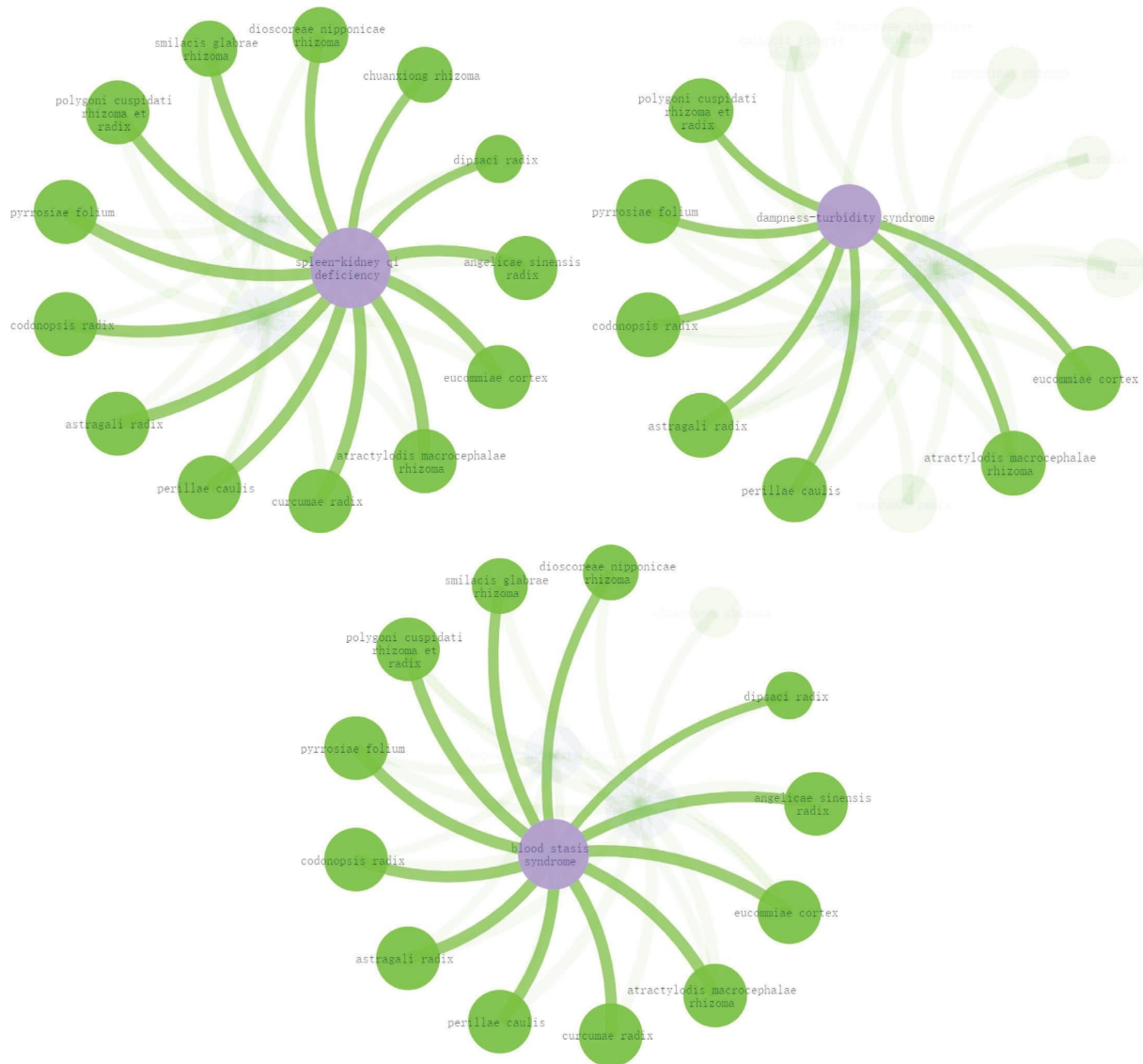
FIGURE 13: Relationship between syndrome and CHM. It shows the complex network composed of "spleen and kidney qi deficiency, damp turbidity syndrome, and blood stasis syndrome" and corresponding CHM.

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

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 tTreating 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: pretraining 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 diagnosising 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.

Research Article

Potential Therapeutic Mechanism of Radix Angelicae Biseratae and Dipsaci Radix Herb Pair against Osteoarthritis: Based on Network Pharmacology and Molecular Docking

Yujiang Xi [iD],[1] **Ting Zhao,**[1] **Mingqin Shi,**[2] **Xiaoyu Zhang,**[1] **Yanyuan Bao,**[1] **Jiamei Gao,**[1] **Jiayan Shen,**[1] **Hui Wang,**[1] **Zhaohu Xie,**[2] **Qi Wang** [iD],[2] **Zhaofu Li** [iD],[1,2] **and Dongdong Qin** [iD][2]

[1]*The First School of Clinical Medicine, Yunnan University of Chinese Medicine, Kunming, China*
[2]*School of Basic Medical Sciences, Yunnan University of Chinese Medicine, Kunming, China*

Correspondence should be addressed to Qi Wang; wangqinet@163.com, Zhaofu Li; lzf0817@126.com,
and Dongdong Qin; qindong108@163.com

Background. A major contributor to older disability is osteoarthritis. Radix Angelicae Biseratae (known as Duhuo in China, DH, the dried rhizome of Angelica pubescens) and Dipsaci Radix (known as Xuduan in China, XD, the dried rhizome of Dipsacus asper Wall) herb pair (DXHP) is widely used to treat osteoarthritis, but the underlying molecular mechanisms still have not been revealed. This research aimed to illustrate the therapeutic mechanism of DXHP against osteoarthritis through the techniques of network pharmacology and molecular docking. *Methods.* Gene targets for osteoarthritis and active ingredients for DXHP were screened based on the pharmacology public database and the gene-disease target database. The software program Cytoscape was used to visualize the active chemical target-disease gene network. The STRING biological information website was used to investigate protein interactions. On the Metascape bioinformatics website, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were carried out. The molecular docking of the important chemicals and primary targets identified by the aforementioned screening was performed using Autodock software. *Results.* Twenty-six active substances from the DXHP that had strong connections to 138 osteoarthritis-related targets were screened out. According to network analysis, TNF, GAPDH, IL-6, AKT-1, IL-1B, and VEGFA are prospective therapeutic targets, while osthole, cauloside A, ammidin, angelicone, beta-sitosterol, and asperosaponin VI may be significant active components. 1705 biological processes (BP), 155 molecular functions (MF), and 89 cellular components (CC) were identified by GO analysis. KEGG analysis indicated that IL-17, NF-kappa B, HIF-1, MAPK, and AGE-RAGE signaling pathways are potentially involved. Molecular docking showed that cauloside A, osthole, and β-sitosterol have excellent binding activity with main targets. *Conclusions.* This study comprehensively illuminated the active ingredients, potential targets, primary pharmacological effects, and relevant mechanisms of the DXHP in the treatment of OA. These findings provide fresh thoughts into the therapeutic mechanisms of the main active ingredients of DXHP and provide a reference for further exploration and clinical applications of DXHP.

1. Introduction

Destruction of cartilage, remodeling of the subchondral bone, the development of osteophytes, and synovial inflammation are all symptoms of osteoarthritis (OA), a degenerative joint disease [1–3]. OA is directly associated to heredity, age, obesity, injury, and chronic inflammation [4–7]. As populations of aging and obese individuals have grown in size, the prevalence of OA has also increased and is now considered a major public health problem worldwide [8]. According to recent studies, the global prevalence of OA grew by 48% between 1990 and 2019 [3, 9], and that OA affects more than 500 million persons, occurring most commonly in postmenopausal women over age 50 [10]. Some guidelines indicate that nonsteroidal anti-inflammatory drugs (NSAIDs) are the most crucial

medications for OA [11, 12]. This class of drugs achieves its anti-inflammatory and analgesic effects by inhibiting cyclooxygenase (COX), a key enzyme in the metabolism of arachidonic acid (AA) [13]. However, the use of COX inhibitors alone is not ideal. Studies have shown that AA can also be metabolized by lipoxygenase (LOX) to produce inflammatory substances such as leukotrienes [14]. Therefore, single introduction of COX inhibitors cannot limit production of inflammatory mediators but instead stimulates and increases the release of inflammatory mediators in the LOX pathway, resulting in adverse reactions [15]. The current mainstream treatment options for OA have limited effects on delaying the development of joint inflammation. In the long term, such drugs cause serious damage to the functions of the digestive tract, liver, and kidneys [16, 17]. In addition, long-term use accelerates the progression of arthritis [18]. Therefore, research into alternative therapies that can contribute to the development of new drugs could improve treatment options for OA.

Traditional Chinese medicine (TCM), as a supplementary therapy, has accumulated rich theoretical knowledge that can potentially help advance clinical prevention and treatment strategies for OA [19]. The earliest research on OA in TCM can be traced back thousands of years to the ancient text "Huangdi Neijing [20]." In recent years, TCM has been suggested to have anti-inflammatory, anti-apoptotic, antioxidative, anti-metabolic, and proliferative effects in the treatment of OA diseases [21]. Herb pair refers to the combination of drugs that can be used simultaneously in TCM clinical to enhance efficacy or reduce adverse reactions. Compared with TCM formulas, herb pairs have a more defined synergistic mechanism [22]. *Radix Angelicae Biseratae* (known as Duhuo (DH), the dried rhizome of *Angelica pubescens*) and *Dipsaci Radix* (known as Xuduan (XD), the dried rhizome of *Dipsacus asper* Wall) are a classic herbal drug combination that has been widely used in TCM. Clinical statistical studies have shown that DXHP is one of the most frequently used Chinese medicine pairs in the treatment of OA [23]. In modern pharmacological research, the active ingredients in DH are linked to antioxidant, anti-inflammatory, and analgesic effects [24], while the active components of XD purportedly protect against bone loss, promote cartilage formation, and improve bone metabolism [25]. In addition, the pharmacological compatibility mechanism of DXHP is similarly reflected in the TCM theory. DXHP originated from Duhuo Xuduan decoction in the ancient Chinese medical book "Waitai Miyao." Chinese Medicine Pharmacopoeia points out that DH helps relieve symptoms of rheumatism, particularly symptoms associated with dampness, swelling, and pain. XD supposedly tonifies the liver and kidneys, nourishes the muscles and bones, and repair fracture [26]. Hence, it is highly essential to research the pharmacological mechanism of DXHP therapy as a possible means of improving the treatment of OA. However, the molecular mechanism of DXHP in the treatment of OA remains unclear.

In recent years, network pharmacology, which combines multidisciplinary knowledge and methods, has the characteristics of being systematic and holistic. Molecular docking can calculate the binding energies between ligands and receptors and predict reasonable binding mode and has been used to explore the molecular process of drug active ingredients acting on the human body. Therefore, the aim of this research is to clarify the molecular underpinnings behind the effects of DXHP on OA using network pharmacology and molecular docking technologies. The specific technical route adopted in this study is shown in Figure 1.

2. Materials and Methods

2.1. Acquisition of Bioactive Compounds of DXHP. To obtain more comprehensive compositional data, the active ingredients of DH and XD were retrieved from three herbal pharmacological databases, including the TCMSP data platform (https://tcmsp-e.com/) [27], the ETCM data platform (https://www.tcmip.cn/ETCM/index.html) [28], and the SymMap data platform (https://www.symmap.org/) [29]. According to previous studies and recommended standards [30], oral bioavailability (OB) ≥30% is regarded as having an excellent absorption. As the selection criteria for "drug-like" compounds in traditional Chinese medicinal materials, DL ≥0.18 is thought to be appropriate for drug design. Therefore, OB and DL criteria were chosen to investigate compounds that meet basic pharmacokinetic criteria for effective drug development [31]. Finally, we searched PubMed and CNKI databases for compounds that did not meet the criteria but had been proved to have significant therapeutic effects on OA.

2.2. Target Prediction for Compounds. The active ingredients obtained after the aforementioned screening were queried by PubChem (database https://pubchem.ncbi.nlm.nih.gov/) for SMILES codes of their potential ingredients, which were entered into Swiss Target Prediction (https://www.swisstargetprediction.ch/) for further prediction of active ingredient targets, setting the screening parameter criteria to the probability value >0.1. To normalize gene nomenclature and species and avoid over-annotation of homologous proteins, these target genes were transformed into matching gene symbols that corresponded to the "*Homo sapiens*" species using the UniProt information system (https://www.uniprot.org/) [32].

2.3. Collecting Herb-Disease Genes. We retrieved OA-related genes from the GEO data platform (https://www.ncbi.nlm.nih.gov/geo/) [33] database using "osteoarthritis" as a keyword search. The datasets were processed using the robust multiarray average algorithm for background correction and matrix data normalization. The screening conditions for significantly differential genes were $P < 0.05$ and $|\log FC| > 1.5$. The differentially expressed genes for OA were obtained. Then, we search the OA-related genes in the following databases: GeneCards (https://www.genecards.org) [34], the TTD information system (https://db.idrblab.net/ttd) [35], and the OMIM information system (https://www.omim.org) [36]. For these searches, the species source was limited to "*Homo sapiens*." The above-obtained genes were
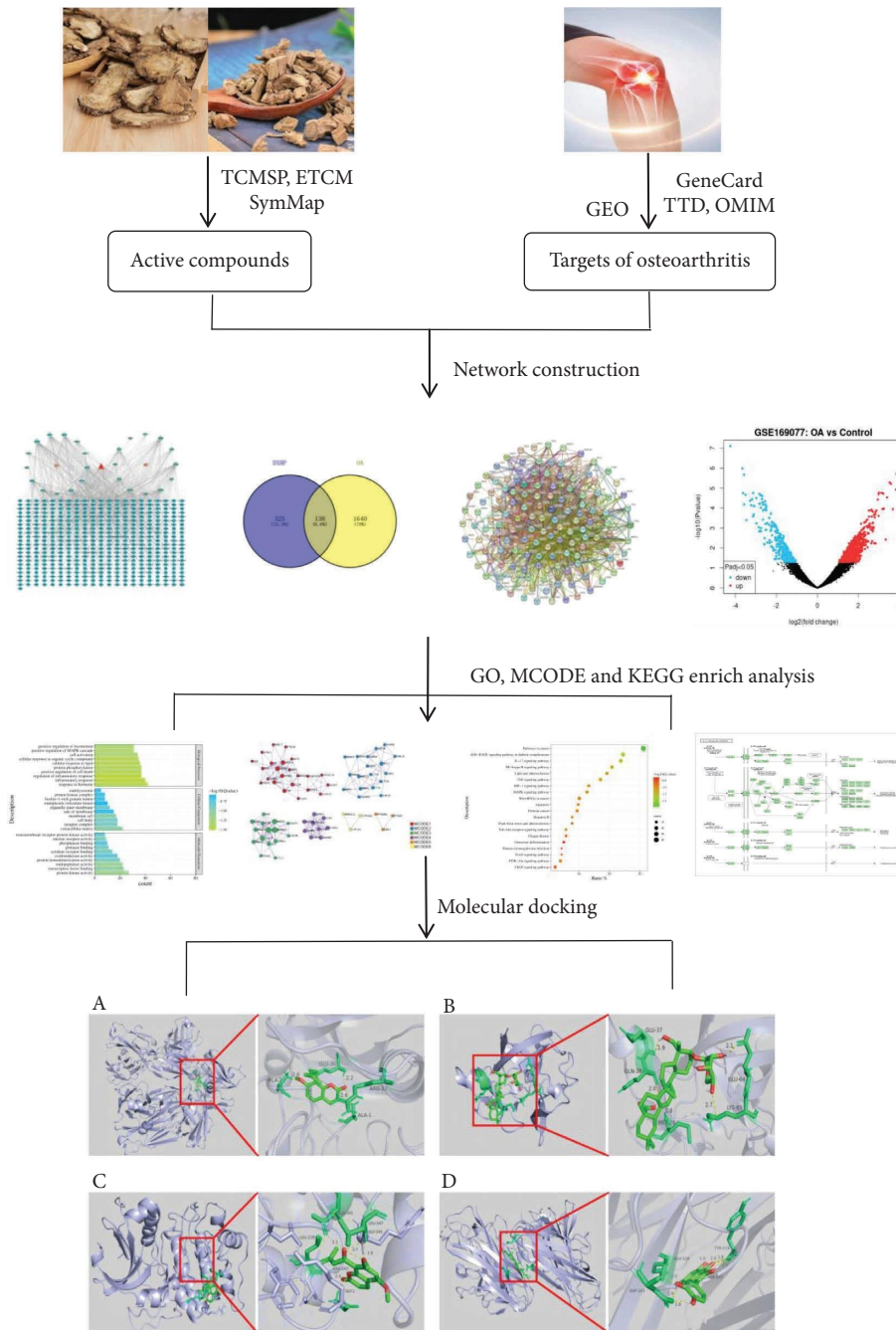
FIGURE 1: The flowchart of network pharmacology-based prediction and molecular docking technology.

merged, and the duplicated genes were deleted to obtain the defined disease targets of osteoarthritis. Subsequently, the OA-related targets were matched with the active compound targets of DXHP to obtain the target of DXHP in the treatment of OA. The Venn diagram was drawn using the Venny2.1 plotter (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

2.4. Protein-Protein Interaction (PPI) Network. Based on the STRING Bioinformatics system (https://www.string-db.org/), a protein-protein interaction (PPI) network was crafted for the common targets of compounds and diseases [37]. The choice of

species was human, and the confidence level was set to >0.4. Others parameters keep the default settings. To identify the primary target genes for OA therapy that have a strong correlation with DXHP, the topological properties of common targets were analyzed using the CytoNCA topology analyzer (A plugin for Cytoscape). The network analyzer tool was applied for topology analysis, referring to three parameters, namely the degree centrality value (DC), the betweenness centrality value (BC), and the closeness centrality value (CC). The greater the DC of the node, the higher the importance of the node in a PPI network. Targets with all three parameters above average were selected and ordered by the degree centrality value.

2.5. Protein-Protein Interaction Enrichment MCODE Analysis.

MCODE, as a clustering algorithm, helps to capture modules with high-quality biological processes in a large PPI spherical network and helps to discover a subset of targets that are closely related to that functional module. MCODE subcluster enrichment analysis was performed on the PPI network. PPI enrichment analysis used the following databases: STRING, InWebIM, OmniPath, and BioGRID. Only the STRING Bioinformatics system (Physical Score >0.132) and BioGRID data are used in the Physical Interaction module. A subset of the proteins in the resultant network physically interacts with at least one other item on the list. The MCODE methodology was used to find components of highly linked networks in subnetworks with 3 to 500 proteins.

2.6. Functional Annotation from Gene Ontology (GO) and Pathway Enrichment Analysis from the Kyoto Encyclopedia of Genes and Genomes (KEGG).

The Metascape database (https://metascape.org/) is a powerful broad-coverage and fast-updating gene function annotation analysis tool that can analyze a large number of gene or protein functions. In order to effectively research the biological ontology of DXHP in the regulation of OA and to clarify the biological process of each core target protein and its function in signaling pathway transduction, GO functional annotation and KEGG pathway enrichment analysis were performed based on the Metascape biological system. The common targets of DXHP and OA obtained in the above screening were imported into the target gene list, with confined to just human species and correction of all target genes to their recognized gene symbols. Multiple comparisons were performed using *Benjamini–Hochberg's* FDR correction to avoid false positives. *Bonferroni-corrected P* values <0.01 for GO and KEGG terms were considered significant. The minimal count for the KEGG analysis was 3, and the enrichment factor was more than 1.5. Finally, the top 20 items were picked, and the annotated chart was formed on the Bioinformatics (https://www.bioinformatics.com/) platform for visualization.

2.7. Molecular Docking Verification of Compound Target.

In network pharmacology, molecular docking is a critical tool for verifying compound-target interactions. It works by combining proteins of known targets obtained from network pharmacology with small compounds of active ingredients in natural drugs and then evaluating the strength and activity of the binding [38]. The mol2 format files of the main active components of DXHP were searched and obtained from the PubChem database in advance. Next, the selected protein's best resolution 3D structure was acquired in the PDB database (https://www.rcsb.org/) and their PDBID files were downloaded. For these 3D structures, the species source was limited to "*Homo sapiens*" and contains the crystal structure of the complete pocket. MGLtool 1.5.7 was used to process the protein by adding hydrogen, calculating the charge, merging the nonpolar hydrogen, and saved as a PDBQT format file as a docking ligand. Gridbox coordinates and docking box sizes were set, and molecular docking was performed using Autodock Vina1.1.2. The lowest binding energy score conformation was selected, with lower binding energy scores indicating better docking activity and strength. The molecular docking structures were demonstrated using Pymol 2.3.

3. Results

3.1. Screening of DXHP Active Compounds.

We obtained 106 active components of DH and 63 active components of XD. After screening by OB and DL standards, 20 active ingredients with good ADME properties were obtained for the subsequent study. In addition to this, DH and XD were found to contain some pharmacologically active ingredients, which were excluded because their OB and DL values were less than the screening criteria. Thus, according to previous literature reports, we included a total of six active ingredients such as osthole, columbianadin, umbelliferone, asperosaponin VI, ursolic acid, and loganin [39, 40]. A total of 26 DXHP active compounds and their corresponding 463 targets were screened (Table 1). The results show that a single compound can regulate multiple targets, indicating that DXHP has multicomponent and multitarget components. The active compounds of DXHP and corresponding genes are shown in Figure 2.

3.2. Drug-Disease Core Target Acquisition.

In the GEO database, we identified the dataset that met the criteria as GSE169077. The chip contained six osteoarthritis cartilage tissue samples as the experimental group and five normal knee cartilage tissue samples as the control group. The results of differential expression gene analysis of osteoarthritis showed that 792 genes were found to be differentially expressed. The heatmap and Limma package in the R language were used to draw the dataset volcano map of differential genes (Figure 3). The GeneCards database yielded 3647 OA-related genes in total, and 885 OA-related genes were obtained using a relevance score >1 as the screening criterion. The TTD database attained 33 OA-related genes in total, whereas the OMIM database attained 30 OA-related genes. The above gene sets were combined, and duplicate values were removed to obtain a total of 1778 OA-related genes. The screened drug targets and OA-related genes were input into Venny 2.1.0 plotter, a Venn figure was constructed, and a total of 138 common targets were obtained (Figure 4).

3.3. PPI Network Construction and Centiscape Analysis.

The herb-disease common targets were imported into the STRING Bioinformatics system. Species option was selected as "*Homo sapiens*," the PPI network was constructed (Figure 5), and the CSV file was entered into Cytoscape software 3.9.1 for network topology analysis using the network Stats tool and the centiscape plugin for the construction of the three parameter correlations such as DC, BC, and CC. 138 nodes and 1599 edges made up the network, which had an average connectedness of 23.70. A total of 25 core targets were selected through the centiscape plugin by selecting

Table 1: Active compounds of DXHP.

Drug	Serial no.	MOL ID	Active compound	OB	DL
DH	DH1	MOL001941	Ammidin	34.55	0.22
	DH2	MOL001942	Isoimperatorin	45.46	0.23
	A1	MOL000358	Beta-sitosterol	36.91	0.75
	DH3	MOL003608	O-acetylcolumbianetin	60.04	0.26
	DH4	MOL004777	Angelol D	34.85	0.34
	DH5	MOL004778	[(1R,2R)-2,3-dihydroxy-1-(7-methoxy-2-oxochromen-6-yl)-3-methylbutyl] (Z)-2-methylbut-2-enoate	46.03	0.34
	DH6	MOL004780	Angelicone	30.99	0.19
	DH7	MOL004792	Nodakenin	57.12	0.69
	DH8	MOL000614	Osthole	38.75	0.13
	DH9	MOL002905	Columbianadin	14.82	0.36
	DH10	MOL001950	Psoralen	33.06	"0.10
	DH11	MOL003621	Meranzin hydrate	43.61	0.17
	DH12	MOL005785	Bergaptol	24.22	0.12
	DH13	MOL002558	Umbelliferone	27.37	0.05
XD	XD1	MOL003152	Gentisin	64.06	0.21
	A1	MOL000358	Beta-sitosterol	36.91	0.75
	XD2	MOL009312	(E,E)-3,5-di-O-caffeoylquinic acid	48.14	0.68
	XD3	MOL009316	Cauloside A	43.32	0.81
	XD4	MOL008188	Japonine	44.11	0.25
	XD5	MOL009323	Sylvestroside III	48.02	0.53
	XD6	MOL003106	Asperosaponin VI	1.67	0.07
	XD7	MOL000651	Sweroside aglycone	68.68	0.08
	XD8	MOL000652	Venoterpine	68.97	0.04
	XD9	MOL009313	Mesitol	65.67	0.03
	XD10	MOL000511	Ursolic acid	16.77	0.75
	XD11	MOL001680	Loganin	5.90	0.44
	XD12	MOL000296	Hederagenin	36.91	0.75
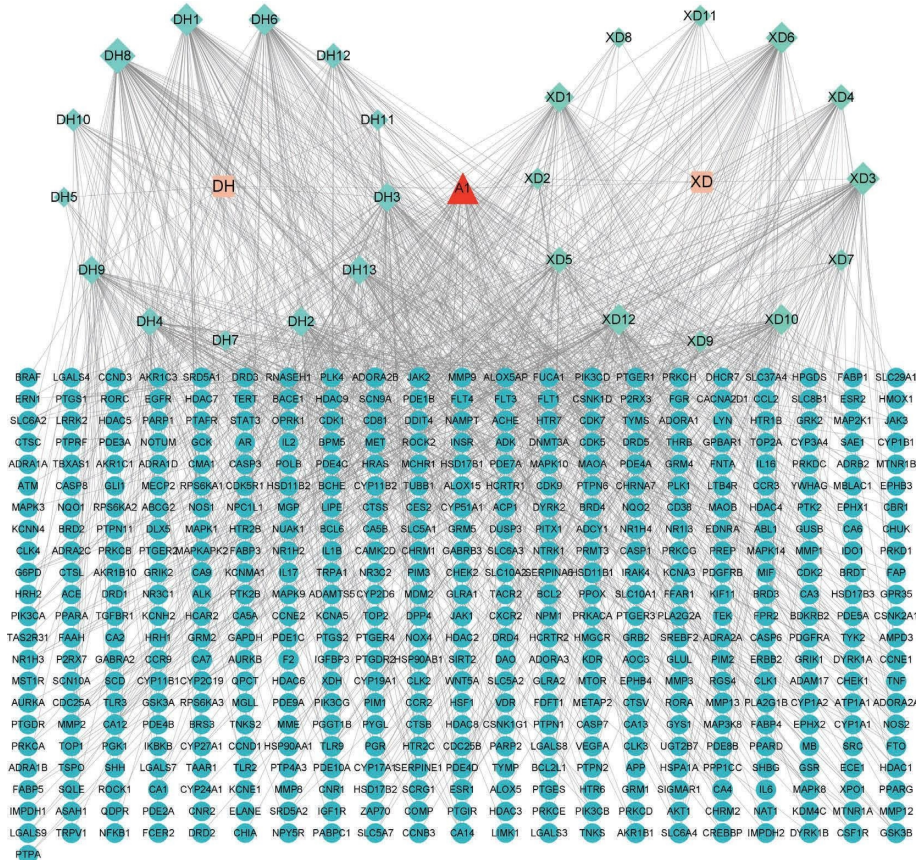
Figure 2: The compound-target network for DXHP. Ultramarine circles represent genes, greenish blue circles represent active compounds of DXHP, orange circles represent natural medicines of DXHP, and the red circle represents the common compounds of the DXHP.
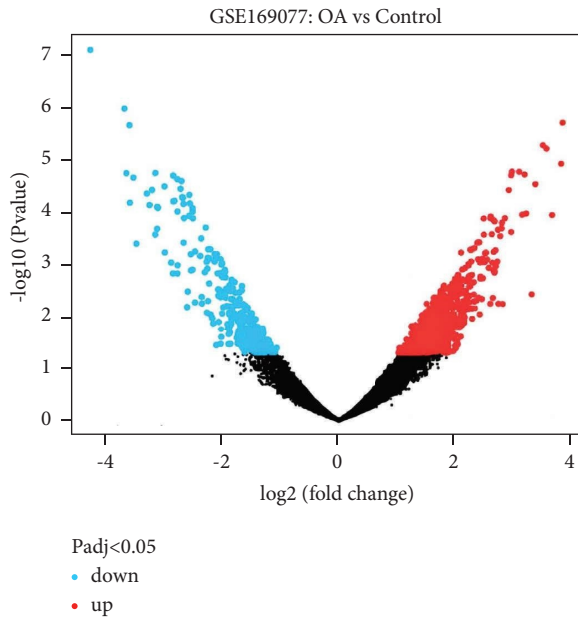
FIGURE 3: The volcano map of differentially expressed genes associated with osteoarthritis. Blue represents decrease and red represents increase.
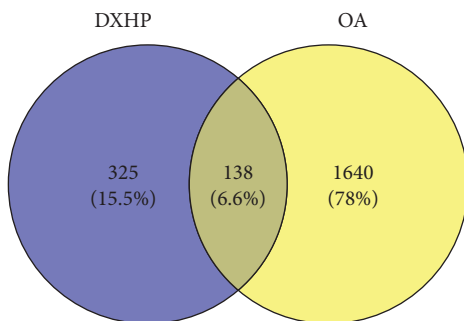
FIGURE 4: The Venn diagram showing 1778 OA-related targets and 463 DXHP-related targets. The intersection section indicates the 138 targets of DXHP in the treatment of OA.
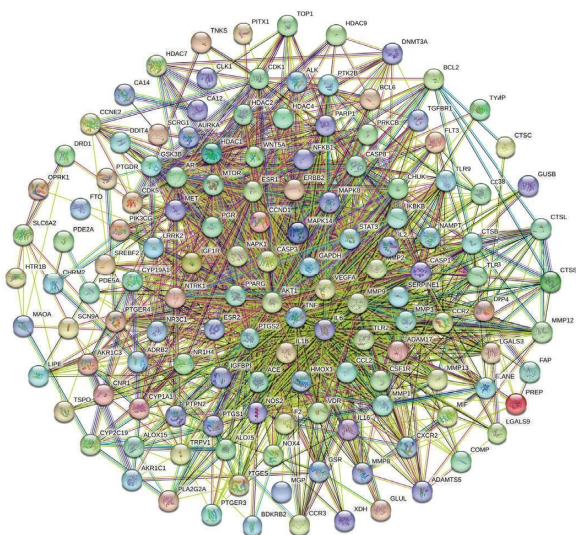
FIGURE 5: The PPI network of common targets of DXHP and OA.

genes with greater than average DC, BC, and CC parameters as key targets and ranking them by the degree centrality value. The 10 highest ranked core targets are listed in Table 2. DH and XD have 155 overlapping targets, 69 of which are associated with OA (Figure 6).

3.4. GO Functional Annotation. GO functional annotation analysis showed that the enrichment results included 1705 biological processes (BPs), 155 molecular functions (MFs), and 89 cellular components (CCs). Most GO annotations of BP were contained to regulation of inflammatory response, response to hormone, inflammatory response, positive regulation of locomotion, cellular response to lipid, cellular response to organic cyclic compound, positive regulation of cell death, protein phosphorylation, positive regulation of MAPK cascade, and cell activation. MF annotation was mainly involved in protein kinase activity, endopeptidase activity, transcription factor binding, nuclear receptor activity, cytokine receptor binding, protease binding, protein homodimerization activity, transmembrane receptor protein kinase activity, oxidoreductase activity, and phosphatase binding. The majority of CC annotations were included to the extracellular matrix, membrane raft, receptor complex, cell body, organelle outer membrane, ficolin-1-rich granule lumen, endolysosome, protein kinase complex, side of the membrane, and endoplasmic reticulum lumen. Based on their Q-value (the Q-value was used for multiple testing, it was calculated using the Benjamini–Hochberg procedure, and higher Q-values showed the greater GO term enrichment), the top 10 terms of BP, CC, and MF were rated. The terms are presented in Figure 7.

3.5. KEGG Pathway Analysis. KEGG pathway analysis enriched 179 signaling pathways after FDR correction. The top 20 pathways included pathways in cancer, the AGE-RAGE signaling pathway in diabetic complications, the IL-17 signaling pathway, the NF-kappa B signaling pathway, lipid and atherosclerosis, the TNF signaling pathway, the MAPK signaling pathway, hepatitis B, the toll-like receptor signaling pathway, prostate cancer, the HIF-1 signaling pathway, the PI3K-Akt signaling pathway, microRNAs in cancer, apoptosis, fluid shear stress and atherosclerosis, chagas disease, the VEGF signaling pathway, human cytomegalovirus infection, the FoxO signaling pathway, and osteoclast differentiation. Based on their Q-value, the top 20 critical signaling pathways are presented in Figure 8. The figures of the two inflammation-related signaling pathways with the highest number of core genes enriched are shown in Figures 9 and 10.

3.6. PPI Enrichment MCODE Analysis. Core genes were identified through network construction and MCODE analysis using the complete datasets for the independent enrichment analysis of gene clusters. The enrichment analysis of biological processes was used for each MCODE component (Figure 11 and Table 3). MCODE analysis showed that the core genes of AKT1, IL1B, TLR9, VEGFA, MMP2, PTGS1,

TABLE 2: Top 10 core targets of PPI analysis.

Targets	DC	BC	CC
TNF	96	0.076310133	0.740540541
GAPDH	95	0.088430017	0.740540541
IL6	91	0.05922158	0.721052632
AKT1	90	0.069715376	0.724867725
IL1B	85	0.052242761	0.698979592
VEGFA	82	0.067869259	0.695431472
CASP3	75	0.042146765	0.671568627
STAT3	70	0.024000163	0.643192488
MMP9	69	0.028131984	0.637209302
PTGS2	68	0.028605206	0.634259259
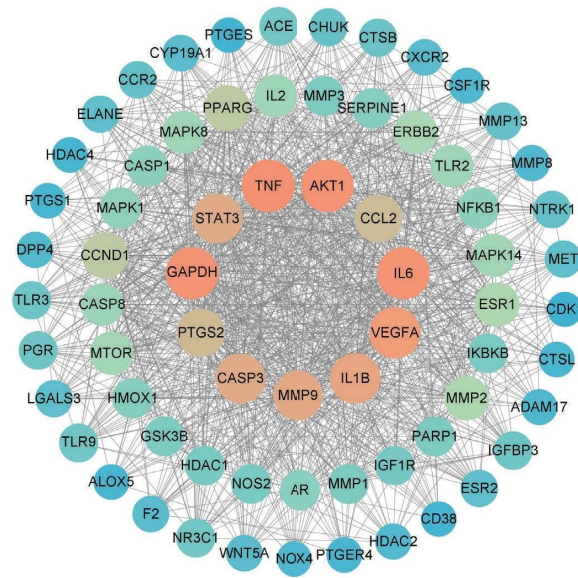
FIGURE 6: Overlapping targets of DH and XD acting on OA.
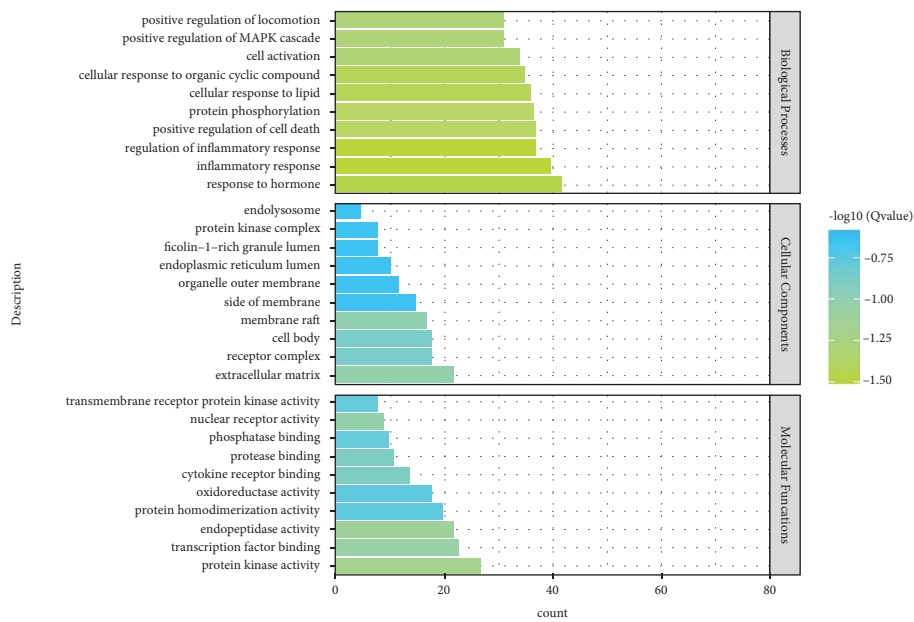
FIGURE 7: BP, CC, and MF of GO enrichment analysis.
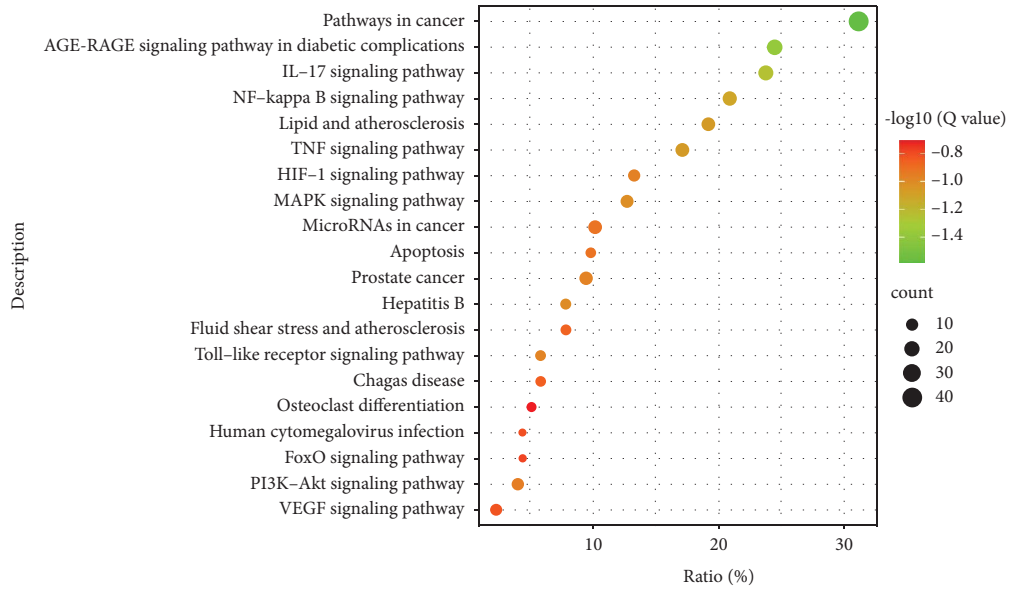
Figure 8: The top 20 enriched KEGG pathway.
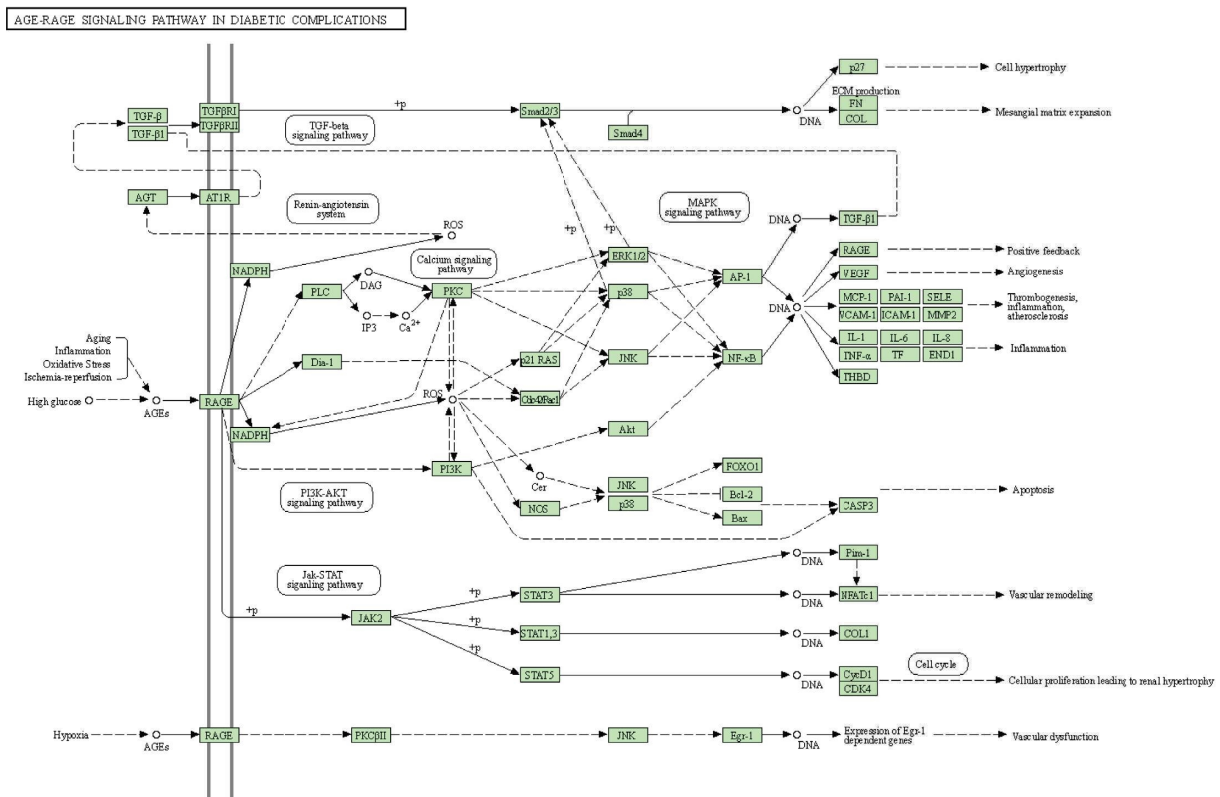
Figure 9: The AGE-RAGE signaling pathway. The green squares represent the core targets of this study.

NOS2, IL6, etc., are mainly involved in the biological process of regulating the inflammatory response in OA. The core genes of ESR1, MTOR, MAPK1, CASP3, CCL2, PGR, etc., are mainly involved in the biological process of regulating defense response in OA. The three terms with the highest scores based on the P values were used as the functional description of the corresponding components (Table 4).

3.7. Molecular Docking Verification. To further validate the reliability of the binding of key targets and components screened by the network analysis, molecular docking verification of some key targets and important active components was carried out using Autodock Vina1.2.0 software. It is largely accepted that the steadier the binding structure, the smaller the binding energy of the receptor-ligand docking
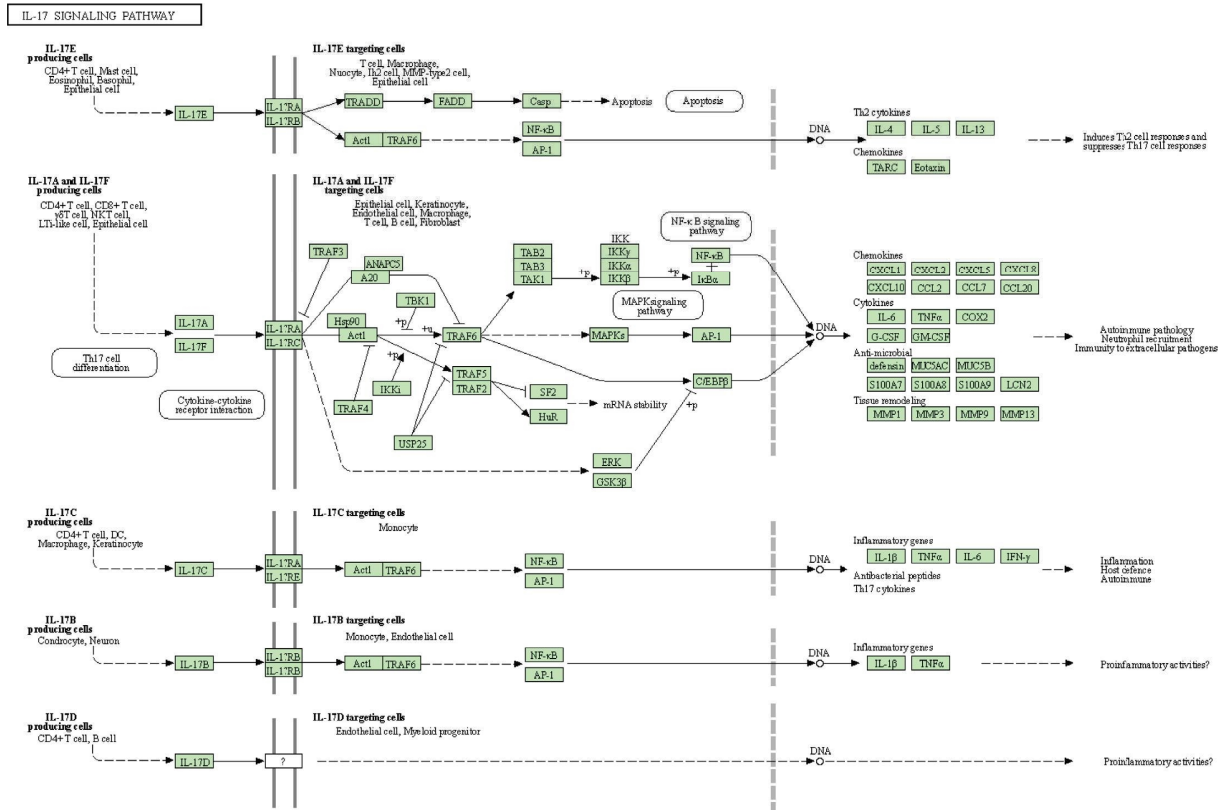
Figure 10: The IL-17 signaling pathway. The green squares represent the core targets of this study.

■ MODULE1	■ MODULE4	■ MODULE1	■ MODULE4
■ MODULE2	■ MODULE5	■ MODULE2	■ MODULE5
■ MODULE3	■ MODULE6	■ MODULE3	■ MODULE6
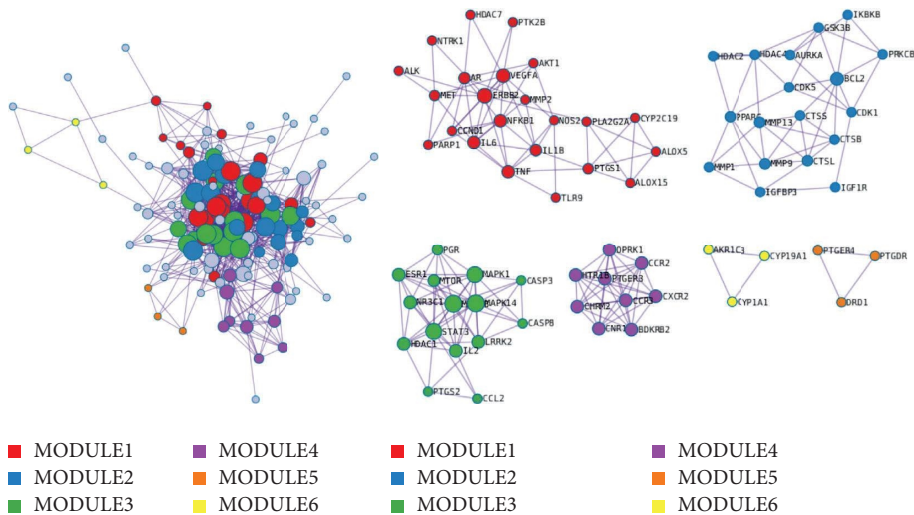
Figure 11: The top 6 gene clusters in the enrichment MCODE analysis.

and therefore the higher the likelihood of interaction between the two, with a binding energy < −5.0 kcal/mol as the screening criterion. Binding energies < −5.0 kcal/mol indicate potential activity, and docking with binding energies < −7.0 kcal/mol is extremely stable [41]. The top six active ingredients in the component-target network were selected for molecular docking with the top six targets in the PPI network. The results showed that about 83% of targets

and active components exhibited binding ability and 16% exhibited extremely strong binding ability. These molecular docking findings align with earlier network screening conclusions, which indirectly validate the treatment ability of DXHP on OA and demonstrate the reliability of network pharmacology applied to this study. The docking results of binding affinity and detailed compound-target interactions are presented in Table 5 and Figure 12.

TABLE 3: Biological processes of protein-protein interaction in the enrichment.

Module	GO	Description	Log10 (P)
MODULE1	GO: 0043410	Positive regulation of MAPK cascade	−15.4
	GO: 0043408	Regulation of MAPK cascade	−13.7
	GO: 0033674	Positive regulation of kinase activity	−13.6
MODULE2	GO: 0022411	Cellular component disassembly	−13.3
	GO: 0030574	Collagen catabolic process	−13.1
	GO: 0032963	Collagen metabolic process	−12.0
MODULE3	GO: 0071407	Cellular response to organic cyclic compound	−12.3
	GO: 0032496	Response to lipopolysaccharide	−10.0
	GO: 0002237	Response to molecule of bacterial origin	−9.8
MODULE4	GO: 0007204	Positive regulation of cytosolic calcium ion concentration	−11.2
	GO: 0007188	Adenylate cyclase-modulating G protein-coupled receptor signaling pathway	−10.8
	GO: 0051048	Negative regulation of secretion	−9.1
MODULE5	GO: 0007204	Positive regulation of cytosolic calcium ion concentration	−6.5
MODULE6	GO: 0006694	Steroid biosynthetic process	−7.3
	GO: 0034754	Cellular hormone metabolic process	−7.1
	GO: 0120254	Olefinic compound metabolic process	−6.9

Table 4: The top three biological processes in enrichment MCODE analysis.

GO	Description	Log10 (P)
GO: 0050727	Regulation of inflammatory response	−37.1
GO: 0006954	Inflammatory response	−35.7
GO: 0031347	Regulation of defense response	−35.7

Table 5: The binding ability of active compounds to core targets.

Object		Binding energy (kcal/mol)					
Target	PDB ID	DH8	XD3	DH1	DH6	A1	XD6
TNF	2AZ5	−6.63	−6.67	−5.62	−6.52	−9.55	−5.01
GAPDH	1IHY	−5.82	−6.81	−5.58	−6.35	−5.76	−3.16
IL6	1ALU	−5.47	−6.73	−5.84	−6.41	−8.03	−4.13
AKT1	3MV5	−5.61	−6.51	−6.03	−6.69	−7.79	−3.12
IL1B	6Y8M	−6.07	−8.01	−4.93	−6.26	−7.98	−3.91
VEGFA	6T9D	−6.68	−5.99	−5.92	−6.05	−8.06	−2.29

(a)

(b)

(c)

(d)
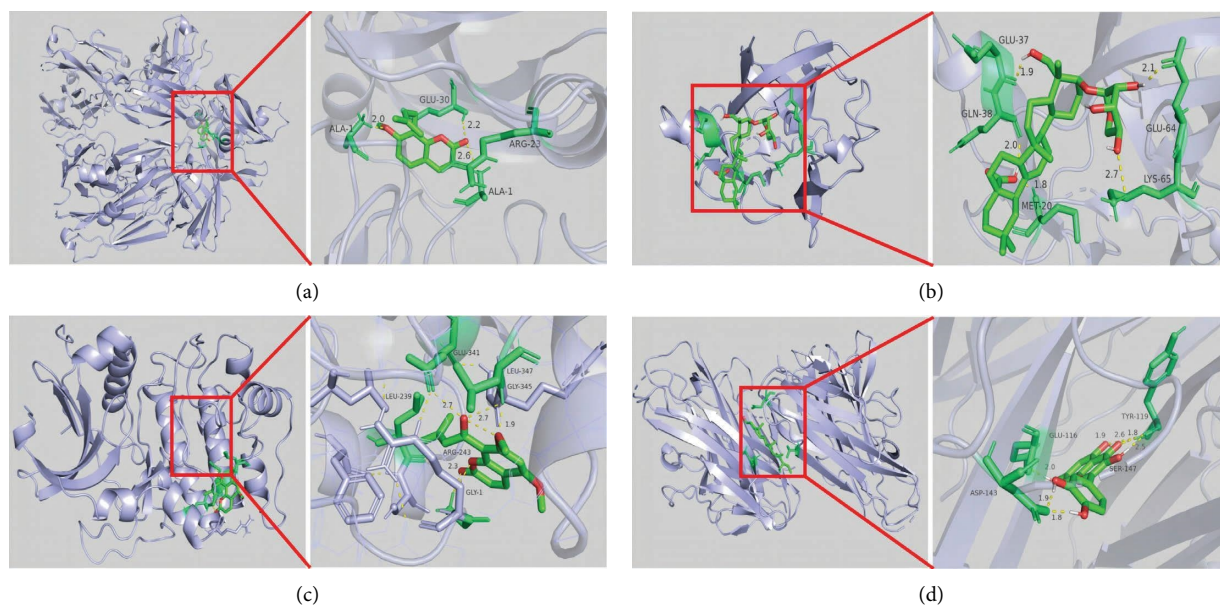
Figure 12: The model of molecular docking simulation results. (a) VEGFA and osthole. (b) IL1B and cauloside A. (c) AKT1 and angelicone. (d) TNF and beta-sitosterol.

4. Discussion

OA has a high prevalence in the elderly population, and modern medical treatment options are primarily symptomatic, with no options available to curb disease progression [42]. The advantages of DXHP in relieving OA symptoms as well as delaying OA disease progression have been demonstrated in previous studies [43], but its mechanism of action remains incompletely elucidated. Therefore, we investigated the mechanism of action of DSHP on OA in a more systematic way using a network pharmacology and the molecular docking approach.

According to this investigation, DXHP's primary therapeutic ingredients for OA include substances such as osthole, asperosaponin VI, angelicone, beta-sitosterol, ammidin, and cauloside A, among which β-sitosterol is the

overlapping component of DH and XD. In recent years, the pharmacological effects of asperosaponin VI in anti-inflammatory analgesia, prevention of osteoporosis, neuroprotection, and anti-apoptosis have attracted the attention of many scholars [44]. Many in vivo and cell experiments have shown that asperosaponin VI can phosphorylate ERK/2 protein to promote the expression of osteogenic genes such as ALP, OCN, COL1 and RUNX2, which is closely related to the PI3K/Akt signaling pathway and the MAPK signaling pathway [45]. In addition, asperosaponin VI could significantly reduce MDA, TNF-α, IL6, and IL10 by significantly inhibiting oxidative stress and inflammatory response in tissues [46]. Osthole, one of the main active components of DH, is thought to improve bone metabolism and promote osteoblast activation [47]. Recent studies on the anti-inflammatory mechanism of osthole have found that

osthole can inhibit the formation and resorption activity of osteoclasts by inhibiting the activation of NF-κB and NFATc1 and reducing the expression of osteoclast-specific genes such as CTSK, MMP-9, TRAP, integrin β3, C-SRC, and NFATc1 [48]. β-sitosterol has a strong down-regulate effect on proinflammatory factors, such as IL-1β, IL-6, and TNF-α and quenches ROS produced by the human body through antioxidant activity. It can also alleviate inflammation through eosinophil percolation [49]. Chen et al. assessed the knee joints of OA rabbits by morphological and histological methods and found that β-sitosterol could significantly inhibit the secretion of matrix metalloproteinases (MMPs) and inhibit the degradation of cartilage [50].

The observations of the coincide target PPI analysis demonstrated that the core targets of DXHP in the healing of OA encompassed multiple targets such as TNF, GAPDH, IL6, AKT1, IL1B, VEGFA, CASP3, STAT3, MMP9, and PTGS2. TNF-α has been demonstrated in many diseases, including osteoarthritis, autoimmune diseases, ankylosing spondylitis, insulin resistance, psoriasis, nephropathy, and cancer [51]. TNF-α is associated with many cytokines, and it has been found that TNF-α implicated in angiogenesis by synergistic induction of VEGF production with IL-1β and IL-6 [52]. Nie et al. found that TNF-α could regulate the harmony between Treg cells and TH17 and TH1 in the joint synovium through FOXP3 dephosphorylation [51]. When cells are stimulated with NO, GAPDH will be nitrosylated, bound to E3 ubiquitin ligase Siah1, and undergo nuclear translocation and apoptosis [53]. The results of molecular docking suggest that DXHP may achieve the purpose of treating OA by binding to GAPDH. Studies have shown that [54] knockdown of the IL-6 gene in an OA rat model can lead to inhibition of MMP13 expression and secretion, which may be related to the inhibition of c-fos/ap-1-mediated inflammatory stimulation in OA chondrocytes. IL-6 can also cause an increase in the expression of MMP9, prevents the formation of type II collagen and proteoglycans, accelerates the degradation of extracellular matrix, and influences bone resorption by activating osteoclasts [55]. Shahine and Elhadidi found that the expression amount of the IL-1β gene in OA organisms was positively correlated with the pain index [56]. Correspondingly, high expression of IL-1β was detected in the synovial membrane and fluid of OA organisms and was positively correlated with OA disease [57]. Apoptotic chondrocytes are essential for the progression of OA. AKT1 is a key down-regulate gene kinase involved in the PI3K pathway. Researchers have demonstrated that the PI3K/AKT pathway that involves AKT1 prevents chondrocyte apoptosis [58]. TP53 inhibits DNA replication, induces apoptosis, and accelerates cartilage degradation [59]. In addition, researchers have demonstrated that the VEGFA is closely associated with many pathological responses, such as osteophyte formation and cartilage degeneration in OA [60]. The expression of VEGFA stimulates the division of vascular endothelial cells to promote angiogenesis, accelerate the exchange of nutrients in the knee joint and the metabolism of inflammatory products, and promote the growth of cartilage synovial cells.

The KEGG enrichment analysis showed that the main enriched inflammation-related signaling pathways included AGE-RAGE, IL-17, NF-kappa B, MAPK, TLR, HIF-1, PI3K-Akt, and the VEGF signaling pathway. The other signaling pathways enriched included pathways in cancer, lipid and atherosclerosis, hepatitis B, prostate cancer, microRNAs in cancer, apoptosis, fluid shear stress and atherosclerosis, chagas disease, human cytomegalovirus infection, FoxO signaling pathway, and osteoclast differentiation. The MCODE analysis revealed that the regulatory response to inflammation was the most significant biological process in treating OA by DXHP. The NF-κb pathway is closely related to cartilage destruction in OA and targeted therapy of OA [61]. Previous research has showed that the secretion level of NF-κb in joint synovial fluid and peripheral blood of OA patients is increased, and the degree of up-regulation of the NF-κb pathway is positively correlated with the degree of cartilage erosion and destruction [62]. After phosphorylation, IκBα dissociates from NF-κB, and NF-κB is activated and enters the nucleus, promoting the synthesis and secretion of TNF-α and IL-6 [63, 64] and eventually leading to the degeneration of articular cartilage [65, 66]. In addition, IL-17 can trigger the release of chemokines, cytokines, antimicrobial peptides, and matrix metalloproteinases from mesenchymal and bone marrow cells [67]. Bai et al. [68] found that the level of IL-17 in the serum of OA patients increased and was positively correlated with the severity of OA. Due to the lack of capillaries in articular cartilage, the cartilage microenvironment is essentially a hypoxic environment. The main hypoxia-inducible factor (HIF) of articular chondrocytes is a transcription factor, which is also the main mediator of homeostatic response that enables cells to survive under hypoxic conditions [69]. In OA, the expression of HIF-1α decreases [70], and the loss of HIF-1α can up-regulate the expression of MMP13, degrade col2A1 and ACAN, promote chondrocyte degradation, and promote the development of OA [71]. The advanced glycation end product receptor (RAGE) is secreted in a variety of cells, including macrophages and mast cells [72]. AGE is a ligand of RAGE and is the end product of glycosylation of proteins and sugar [73]. RAGE is essential for the induction of several inflammatory genes as well as important signaling pathways linked to proinflammatory responses. Although AGE induces inflammation by exciting NF-κB and MAPK in a variety of cells including osteocytes [74], it can also up-regulate the expression of PGE2 and NO via the MAPK pathway and induces the chondrocytes inflammatory [75].

As an herb pair, DH and XD have many overlapping targets and active ingredients, and the compatibility mechanism of DXHP in treating OA may be related to this. β-sitosterol, as its overlapping active ingredient, is associated with anti-inflammatory and antioxidant effects, immune regulation, and bone metabolic balance [76]. Furthermore, the aforementioned KEGG and MCODE analyses revealed that the overlapping targets such as IL6, IL-1β, AKT1, VEGFA, MMP13, and STAT3 are vital links in inflammatory signalling pathways such as AGE-RAGE, IL-17, MAPK, and NF-kappa B and are extensively involved in the biological processes of regulating the inflammatory response in OA.

This finding may help us to better understand the compatibility mechanism of DXHP from a molecular perspective.

Molecular docking tests were carried to further confirm the molecular mechanism of DXHP in the treatment of OA. The results revealed that osthole, angelicone, cauloside A, and β-sitosterol have excellent binding activity with multiple key targets, such as TNF, GAPDH, IL-6, AKT1, IL-1β, and VEGFA. This also explains the therapeutic mechanism of DXHP from another perspective. These results offer insight for the subsequent application of network pharmacology methods to improve the efficiency of natural medicine ingredient development, which could facilitate the development of new high-efficiency, low-toxicity, multi-target OA drugs capable of improving symptoms and delaying disease progression.

Compared to recently published network pharmacology studies of similar diseases [77], we have expanded the source of the drug and disease database and introduced differential genes from the GEO microarray, making the source more comprehensive. In order to capture modules with high-quality biological processes in large PPI networks, we apply the MCODE clustering algorithm, which allows for a further in-depth interpretation of the target network. In addition, we have investigated the overlapping active ingredients and targets of this herb pair to elucidate the compatibility mechanisms of DXHP from a new perspective.

Despite the study's advantages, several limitations should be mentioned. Firstly, the sources of component and target data are scattered and vary widely between databases, resulting in insufficiently accurate data sources [78]. Secondly, the validation of the findings is only based on computer simulations of molecular docking, which has limitations in terms of convincingness [79]. Therefore, the key pathways and targets identified in this study need to be validated in more in-depth in vivo and in vitro experiments.

5. Conclusion

In this study, we found that DXHP is a valuable TCM herb pair for treating patients with OA through multiple components, targets, and pathways. Its pharmacological mechanism could be through the AGE-RAGE, IL-17, NF-kappa B, MAPK, TLR, HIF-1, PI3K-Akt, and VEGF signaling pathways to alleviate OA. Hopefully, our research may provide a scientific basis for the prodrug discovery of its natural ingredients and the identification of therapeutic targets in the future.

Data Availability

The original contributions presented in the study are included within the article/supplementary material. The data supporting the findings of the current study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yujiang Xi, Ting Zhao, and Mingqin Shi contributed equally to this work.

Acknowledgments

References

[1] L. J. Wang, N. Zeng, Z. P. Yan, J. T. Li, and G. X. Ni, "Post-traumatic osteoarthritis following ACL injury," *Arthritis Research and Therapy*, vol. 22, no. 1, p. 57, 2020.

[2] D. Bryazka, M. B. Reitsma, M. G. Griswold et al., "Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020," *The Lancet*, vol. 400, no. 10347, pp. 185–235, 2022.

[3] D. J. Hunter and S. Bierma-Zeinstra, *The Lancet*, vol. 393, no. 10182, pp. 1745–1759, 2019.

[4] E. Macfarlane, M. J. Seibel, and H. Zhou, "Arthritis and the role of endogenous glucocorticoids," *Bone Res*, vol. 8, no. 1, p. 33, 2020.

[5] D. A. Young, M. J. Barter, and J. Soul, "Osteoarthritis year in review: genetics, genomics, epigenetics," *Osteoarthritis and Cartilage*, vol. 30, no. 2, pp. 216–225, 2022.

[6] P. Kulkarni, A. Martson, R. Vidya, S. Chitnavis, and A. Harsulkar, "Pathophysiological landscape of osteoarthritis," *Advances in Clinical Chemistry*, vol. 100, pp. 37–90, 2021.

[7] T. A. Perry, M. J. Parkes, R. J. Hodgson, D. T. Felson, N. K. Arden, and T. W. O'Neill, "Association between Bone marrow lesions & synovitis and symptoms in symptomatic knee osteoarthritis," *Osteoarthritis and Cartilage*, vol. 28, no. 3, pp. 316–323, 2020.

[8] G. Aubourg, S. J. Rice, P. Bruce-Wootton, and J. Loughlin, "Genetics of osteoarthritis," *Osteoarthritis and Cartilage*, vol. 30, no. 5, pp. 636–649, 2022.

[9] D. Veret, C. Jorgensen, and J. M. Brondello, "Osteoarthritis in time for senotherapeutics," *Joint Bone Spine*, vol. 88, no. 2, Article ID 105084, 2021.

[10] I. A. Szilagyi, J. H. Waarsing, D. Schiphof, J. B. J. van Meurs, and S. M. A. Bierma-Zeinstra, "Towards sex-specific osteoarthritis risk models: evaluation of risk factors for knee osteoarthritis in males and females," *Rheumatology*, vol. 61, no. 2, pp. 648–657, 2022.

[11] Q. Y. Shi, X. Y. Tan, and W. Y. Li, "Joint surgery group of orthopedic branch of Chinese medical as- sociation, orthopedic Expert committee of Wu jieping medical foundation. Consensus of four-stepladder program of knee osteoarthritis (2018)," *Chin J Joint Surg (Electronic Edition)*, vol. 13, no. 1, pp. 124–130, 2019.

[12] D. Huang, "Professional Committee of Orthopedics and Traumatology of Chi- nese Society of Integrative Medicine. Guidelines for diagnosis and treatment of knee osteoarthritis with integrated traditional Chinese and western medicine," *Nature Medicine J*, vol. 98, no. 45, pp. 3653–3658, 2018.

[13] A. A. El-Malah, M. M. Gineinah, P. K. Deb et al., "Selective COX-2 inhibitors: road from success to controversy and the quest for repurposing," *Pharmaceuticals*, vol. 15, no. 7, p. 827, 2022.

[14] R. Dey, S. Dey, A. Samadder, A. Saxena, and S. Nandi, "Natural inhibitors against potential targets of cyclooxygenase, lipoxygenase and leukotrienes," *Comb Chem High Throughput Screen*, vol. 25, no. 14, pp. 2341–2357, 2021.

[15] M. Sisa, M. Dvorakova, V. Temml, V. Jarosova, T. Vanek, and P. Landa, "Synthesis, inhibitory activity and in silico docking of dual COX/5-LOX inhibitors with quinone and resorcinol core," *European Journal of Medicinal Chemistry*, vol. 204, Article ID 112620, 2020.

[16] L. Brosseau, G. A. Wells, G. P. Kenny et al., "The implementation of a community-based aerobic walking program for mild to moderate knee osteoarthritis (OA): a knowledge translation (KT) randomized controlled trial (RCT): Part I: the Uptake of the Ottawa Panel clinical practice guidelines (CPGs)," *BMC Public Health*, vol. 12, no. 1, p. 871, 2012.

[17] J. Y. Reginster, A. Neuprez, M. P. Lecart, N. Sarlet, and O. Bruyere, "Role of glucosamine in the treatment for osteoarthritis," *Rheumatology International*, vol. 32, no. 10, pp. 2959–2967, 2012.

[18] D. Dragos, M. Gilca, L. Gaman et al., "Phytomedicine in joint disorders," *Nutrients*, vol. 9, no. 1, p. 70, 2017.

[19] C. Wang, Y. Gao, Z. Zhang et al., "Safflower yellow alleviates osteoarthritis and prevents inflammation by inhibiting PGE2 release and regulating NF-κB/SIRT1/AMPK signaling pathways," *Phytomedicine*, vol. 78, Article ID 153305, 2020.

[20] Q. L. Yuan, P. Wang, L. Liu et al., "Acupuncture for musculoskeletal pain: a meta-analysis and meta-regression of sham-controlled randomized clinical trials," *Scientific Reports*, vol. 6, no. 1, Article ID 30675, 2016.

[21] L. Li, H. Liu, W. Shi et al., "Insights into the action mechanisms of traditional Chinese medicine in osteoarthritis," *Evidence-based Complementary and Alternative Medicine*, vol. 2017, Article ID 5190986, 13 pages, 2017.

[22] S. Feng, T. Wang, L. Fan et al., "Exploring the potential therapeutic effect of Eucommia ulmoides-Dipsaci Radix herbal pair on osteoporosis based on network pharmacology and molecular docking technology," *RSC Advances*, vol. 12, no. 4, pp. 2181–2195, 2022.

[23] C. Li, *Research on Medication Rules and Systematic Evaluation of Osteoarthritis Based on Data Mining*, Shandong University of Chinese Medicine, Shandong, China, 2020.

[24] L. Yang, A. Hou, S. Wang et al., "A review of the botany, traditional use, phytochemistry, analytical methods, pharmacological effects, and toxicity of angelicae pubescentis Radix," *Evidence-based Complementary and Alternative Medicine*, vol. 2020, Article ID 7460781, 28 pages, 2020.

[25] J. Zhao, G. Liang, and Y. Han, "Research progress of dipsoponin VI from Dipsophidium dipsophidium in the prevention and treatment of osteoporosis," *Chinese Journal of Osteoporosis*, vol. 26, no. 05, pp. 755–759, 2020.

[26] Pharmacopoeia Commission, *Pharmacopoeia of the People's Republic of China*, China Medical Science Press, Beijing, China, 2015.

[27] J. Ru, P. Li, J. Wang et al., "TCMSP: a database of systems pharmacology for drug discovery from herbal medicines," *Journal of Cheminformatics*, vol. 6, no. 1, p. 13, 2014.

[28] H. Y. Xu, Y. Q. Zhang, Z. M. Liu et al., "ETCM: an encyclopaedia of traditional Chinese medicine," *Nucleic Acids Research*, vol. 47, no. D1, pp. D976–D982, 2019.

[29] 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, no. 1, pp. D1110–D1117, 2019.

[30] Y. Wen, L. He, R. Peng et al., "A novel strategy to evaluate the quality of herbal products based on the chemical profiling, efficacy evaluation and pharmacokinetics," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 161, pp. 326–335, 2018.

[31] S. K. Yan, J. Zhao, S. Dou, P. Jiang, R. H. Liu, and W. D. Zhang, "Methodology of modernization research in traditional Chinese medicine based on systems biology and network biology," *Chinese Journal of Natural Medicines*, vol. 7, no. 4, pp. 249–259, 2009.

[32] H. Wang, H. Wang, J. Zhang et al., "Molecular mechanism of crataegi folium and alisma rhizoma in the treatment of dyslipidemia based on network pharmacology and molecular docking," *Evidence-based Complementary and Alternative Medicine*, vol. 2022, Article ID 4891370, 13 pages, 2022.

[33] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets--update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, 2012.

[34] G. Stelzer, N. Rosen, I. Plaschkes et al., "The GeneCards suite: from gene data mining to disease genome sequence analyses," *Current Protocols in Bioinformatics*, vol. 54, pp. 1.30.1–1.30.33, 2016.

[35] Y. Zhou, Y. Zhang, X. Lian et al., "Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents," *Nucleic Acids Research*, vol. 50, no. D1, pp. D1398–D1407, 2022.

[36] J. S. Amberger and A. Hamosh, "Searching online mendelian inheritance in man (OMIM): a knowledgebase of human genes and genetic phenotypes," *Current Protocols in Bioinformatics*, vol. 58, pp. 1.2.1–1.2.12, 2017.

[37] D. Szklarczyk, A. L. Gable, D. Lyon et al., "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Research*, vol. 47, no. D1, pp. D607–D613, 2019.

[38] J. K. Maier and P. Labute, "Assessment of fully automated antibody homology modeling protocols in molecular operating environment," *Proteins*, vol. 82, no. 8, pp. 1599–1610, 2014.

[39] C. Zhang, C. Yang, and C. Xiao, "Studies on chemical constituents of Dipsaci Radix," *Shandong Science*, vol. 35, no. 05, pp. 1–9, 2022.

[40] L. Zhou and J. Zen, "Research progress on chemical composition and pharmacological activity of Radix Angelicae Biseratae," *Chinese Modern Traditional Medicine*, vol. 21, no. 12, pp. 1739–1748, 2019.

[41] K. Y. Hsin, S. Ghosh, and H. Kitano, "Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology," *PLoS One*, vol. 8, no. 12, Article ID e83922, 2013.

[42] S. T. Skou and E. M. Roos, "Physical therapy for patients with knee and hip osteoarthritis: supervised, active treatment is current best practice," *Clinical & Experimental Rheumatology*, vol. 120, no. 5, pp. 112–117, 2019.

[43] H. Zhu, *A Retrospective Study of Osteoarthritis Treated by Master of Traditional Chinese Medicine Zhou Zhongying and Discussion of Clinical Experience*, Nanjing University of Traditional Chinese Medicine, Nanjing, China, 2022.

[44] H. Tian, F. Zhao, and Y. Li, "Research progress of asperosaponin VI," *Chinese Journal of Experimental Medicine*, vol. 24, no. 05, pp. 226–234, 2018.

[45] K. Ke, Q. Li, X. Yang et al., "Asperosaponin VI promotes bone marrow stromal cell osteogenic differentiation through the PI3K/AKT signaling pathway in an osteoporosis model," *Scientific Reports*, vol. 6, no. 1, Article ID 35233, 2016.

[46] C. Li, Y. Gao, J. Tian, Y. Xing, H. Zhu, and J. Shen, "Long-term oral Asperosaponin VI attenuates cardiac dysfunction, myocardial fibrosis in a rat model of chronic myocardial infarction," *Food and Chemical Toxicology*, vol. 50, no. 5, pp. 1432–1438, 2012.

[47] J. Zhang, X. Xie, and J. Li, "Research overview of osthole extract in prevention and treatment of primary osteoporosis," *Chinese Journal of Osteoporosis*, vol. 26, no. 10, pp. 1546–1549, 2020.

[48] Y. Ma, L. Wang, S. Zheng et al., "Osthole inhibits osteoclasts formation and bone resorption by regulating NF-κB signaling and NFATc1 activations stimulated by RANKL," *Journal of Cellular Biochemistry*, vol. 120, no. 9, pp. 16052–16061, 2019.

[49] J. N. Choi, Y. H. Choi, J. M. Lee et al., "Anti-inflammatory effects of β-sitosterol-β-D-glucoside from Trachelospermum jasminoides (Apocynaceae) in lipopolysaccharide-stimulated RAW 264.7 murine macrophages," *Natural Product Research*, vol. 26, no. 24, pp. 2340–2343, 2012.

[50] W. P. Chen, C. Yu, P. F. Hu, J. P. Bao, J. L. Tang, and L. D. Wu, "Stigmasterol blocks cartilage degradation in rabbit model of osteoarthritis," *Acta Biochimica Polonica*, vol. 59, no. 4, pp. 537–541, 2012.

[51] H. Nie, Y. Zheng, R. Li et al., "Phosphorylation of FOXP3 controls regulatory T cell function and is inhibited by TNF-α in rheumatoid arthritis," *Nature Medicine*, vol. 19, no. 3, pp. 322–328, 2013.

[52] H. Nakahara, J. Song, M. Sugimoto et al., "Anti-interleukin-6 receptor antibody therapy reduces vascular endothelial growth factor production in rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 48, no. 6, pp. 1521–1529, 2003.

[53] K. Li, M. Huang, P. Xu et al., "Microcystins-LR induced apoptosis via S-nitrosylation of GAPDH in colorectal cancer cells," *Ecotoxicology and Environmental Safety*, vol. 190, Article ID 110096, 2020.

[54] A. Haseeb, M. Y. Ansari, and T. M. Haqqi, "Harpagoside suppresses IL-6 expression in primary human osteoarthritis chondrocytes," *Journal of Orthopaedic Research*, vol. 35, no. 2, pp. 311–320, 2017.

[55] T. Wang and C. He, "Pro-inflammatory cytokines: the link between obesity and osteoarthritis," *Cytokine & Growth Factor Reviews*, vol. 44, pp. 38–50, 2018.

[56] E. Shahine and A. Elhadidi, "AB0776 efficacy of glucosamine sulphate in lowering serum level of interleukin-1β in symptomatic primary knee osteoarthritis: clinical and laboratory study," *Annals of the Rheumatic Diseases*, vol. 73, no. 2, pp. 1061.1–1061, 2014.

[57] T. Mabey and S. Honsawek, "Cytokines as biochemical markers for knee osteoarthritis," *World Journal of Orthopedics*, vol. 6, no. 1, pp. 95–105, 2015.

[58] K. M. Tong, D. C. Shieh, C. P. Chen et al., "Leptin induces IL-8 expression via leptin receptor, IRS-1, PI3K, Akt cascade and promotion of NF-κB/p300 binding in human synovial fibroblasts," *Cellular Signalling*, vol. 20, no. 8, pp. 1478–1488, 2008.

[59] M. Lin, Y. Lin, X. Li et al., "Warm sparse-dense wave inhibits cartilage degradation in papain-induced osteoarthritis through the mitogen-activated protein kinase signaling pathway," *Experimental and Therapeutic Medicine*, vol. 14, no. 4, pp. 3674–3680, 2017.

[60] N. Saetan, S. Honsawek, A. Tanavalee et al., "Relationship of plasma and synovial fluid vascular endothelial growth factor with radiographic severity in primary knee osteoarthritis," *International Orthopaedics*, vol. 38, no. 5, pp. 1099–1104, 2014.

[61] E. Jimi, F. Huang, and C. Nakatomi, "NF-κB signaling regulates physiological and pathological chondrogenesis," *International Journal of Molecular Sciences*, vol. 20, no. 24, p. 6275, 2019.

[62] M. Ostojic, A. Zevrnja, K. Vukojevic, and V. Soljic, "Immunofluorescence analysis of NF-kB and iNOS expression in different cell populations during early and advanced knee osteoarthritis," *International Journal of Molecular Sciences*, vol. 22, no. 12, p. 6461, 2021.

[63] M. C. Choi, J. Jo, J. Park, H. K. Kang, and Y. Park, "NF-κB signaling pathways in osteoarthritic cartilage destruction," *Cells*, vol. 8, no. 7, p. 734, 2019.

[64] T. Tanaka, M. Narazaki, and T. Kishimoto, "IL-6 in inflammation, immunity, and disease," *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 10, Article ID a016295, 2014.

[65] B. B. Aggarwal, S. C. Gupta, and J. H. Kim, "Historical perspectives on tumor necrosis factor and its superfamily: 25 years later, a golden journey," *Blood*, vol. 119, no. 3, pp. 651–665, 2012.

[66] P. Wang, P. P. Guan, C. Guo, F. Zhu, K. Konstantopoulos, and Z. Y. Wang, "Fluid shear stress-induced osteoarthritis: roles of cyclooxygenase-2 and its metabolic products in inducing the expression of proinflammatory cytokines and matrix metalloproteinases," *The FASEB Journal*, vol. 27, no. 12, pp. 4664–4677, 2013.

[67] Y. Quan, B. Zhou, Y. Wang et al., "Association between IL17 polymorphisms and risk of cervical cancer in Chinese women," *Clinical and Developmental Immunology*, vol. 2012, Article ID 258293, 6 pages, 2012.

[68] Y. Bai, S. Gao, Y. Liu, S. Jin, H. Zhang, and K. Su, "Correlation between Interleukin-17 gene polymorphism and osteoarthritis susceptibility in Han Chinese population," *BMC Medical Genetics*, vol. 20, no. 1, p. 20, 2019.

[69] C. Maes, G. Carmeliet, and E. Schipani, "Hypoxia-driven pathways in bone development, regeneration and disease," *Nature Reviews Rheumatology*, vol. 8, no. 6, pp. 358–366, 2012.

[70] J. Fernández-Torres, Y. Zamudio-Cuevas, G. A. Martínez-Nava, and A. G. López-Reyes, "Hypoxia-Inducible Factors (HIFs) in the articular cartilage: a systematic review," *European Review for Medical and Pharmacological Sciences*, vol. 21, no. 12, pp. 2800–2810, 2017.

[71] W. Bouaziz, J. Sigaux, D. Modrowski et al., "Interaction of HIF1α and β-catenin inhibits matrix metalloproteinase 13 expression and prevents cartilage damage in mice," *Proceedings of the National Academy of Sciences of the U S A*, vol. 113, no. 19, pp. 5453–5458, 2016.

[72] J. H. Rosenberg, V. Rai, M. F. Dilisio, and D. K. Agrawal, "Damage-associated molecular patterns in the pathogenesis of osteoarthritis: potentially novel therapeutic targets," *Molecular and Cellular Biochemistry*, vol. 434, no. 1-2, pp. 171–179, 2017.

[73] A. Suzuki, A. Yabu, and H. Nakamura, "Advanced glycation end products in musculoskeletal system and disorders," *Methods*, vol. 203, pp. 179–186, 2022.

[74] J. Xie, J. D. Méndez, V. Méndez-Valenzuela, and M. M. Aguilar-Hernández, "Cellular signalling of the receptor for advanced glycation end products (RAGE)," *Cellular Signalling*, vol. 25, no. 11, pp. 2185–2197, 2013.

[75] Z. Rasheed and T. M. Haqqi, "Endoplasmic reticulum stress induces the expression of COX-2 through activation of eIF2α, p38-MAPK and NF-κB in advanced glycation end products stimulated human chondrocytes," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1823, no. 12, pp. 2179–2189, 2012.

[76] S. Babu and S. Jayaraman, "An update on β-sitosterol: a potential herbal nutraceutical for diabetic management," *Biomedicine & Pharmacotherapy*, vol. 131, Article ID 110702, 2020.

[77] X. Q. Hao, Y. Q. Kou, J. W. Xie Xj, J. B. Lv, and J. Su, "Network pharmacology-based study of the anti-oxidative mechanism of san miao wan in treatment of arthritis," *World J Tradit Chin Med*, vol. 8, pp. 100–109, 2022.

[78] H. Wang, J. Liu, J. He et al., "Potential mechanisms underlying the therapeutic roles of sinisan formula in depression: based on network pharmacology and molecular docking study," *Frontiers in Psychiatry*, vol. 13, Article ID 1063489, 2022.

[79] S. Li, M. Niu, K. Yang, B. Zhang, H. Xu, and M. Yang, "Network pharmacology evaluation method guidance-draft," *World J Tradit Chin Med*, vol. 7, pp. 146–154, 2021.

Research Article

Molecular Modeling Identification of Key Secondary Metabolites from *Xylopia aethiopica* as Promising Therapeutics Targeting Essential Measles Viral Proteins

Jeremiah John Oloche [1,2] **Bolaji Bosede Oluremi,**[3] **Christiana Eleojo Aruwa** [4] **and Saheed Sabiu** [4]

[1]*Department of Pharmacology and Therapeutics, College of Health Sciences, Benue State University of Makurdi, Makurdi, Nigeria*
[2]*Department of Pharmacology and Therapeutics, College of Medicine, University of Ibadan, Oyo State, Ibadan, Nigeria*
[3]*Department of Pharmaceutical Microbiology, University of Ibadan, Oyo State, Ibadan, Nigeria*
[4]*Department of Biotechnology and Food Science, Faculty of Applied Sciences, Durban University of Technology, Durban, South Africa*

Correspondence should be addressed to Saheed Sabiu; sabius@dut.ac.za

This study computationally screened three key compounds (vanillin (VAN), oxophoebine (OPB), and dihydrochalcone (DHC)) derived from *Xylopia aethiopica* (Guinea pepper), a medicinal plant with known antiviral activity, against key druggable measles virus (MV) proteins (fusion protein (FUP), haemagglutinin protein (HMG), and phosphoprotein (PSP)). Each molecular species was subjected to a 100 ns molecular dynamics (MD) simulation following docking, and a range of postdynamic parameters including free binding energy and pharmacokinetic properties were determined. The docking scores of the resulting OPB-FUP (-5.4 kcal/mol), OPB-HMG (-8.1 kcal/mol), and OPB-PSP (-8.0 kcal/mol) complexes were consistent with their respective binding energy values (-25.37, -28.74, and -40.68 kcal/mol), and higher than that of the reference standard, ribavirin (RBV) in each case. Furthermore, all the investigated compounds were thermodynamically compact and stable, especially HMG of MV, and this observation could be attributed to the resulting intermolecular interactions in each system. Overall, OPB may possess inhibitory properties against MV glycoproteins (FUP and HMG) and PSP that play important roles in the replication of MV and measles pathogenesis. While OPB could serve as a scaffold for the development of novel MV fusion and entry inhibitors, further *in vitro* and *in vivo* evaluation is highly recommended.

1. Introduction

Measles is a communicable disease caused by a member of the genus *Morbillivirus*, and species *Measles morbillivirus*, also referred to as the measles virus (MV) [1, 2]. The measles virus infects the central nervous system leading to serious neurological disorders. Despite the availability of clinically effective live attenuated vaccines currently in use, the disease is one the major cause of morbidity and mortality largely among unvaccinated children. The disease accounted for approximately 140,000 global deaths in children below

5 years in 2018, with most of the reported cases occurring in Africa and the Eastern Mediterranean regions [3]. Recently, over a thousand children were reportedly infected in a measles outbreak in North-East Nigeria [4].

The measles virus is an enveloped antigenically monotypic negative-sensesingle-stranded RNA virus [5]. The measles virus core consists of an RNA genome covered by a nucleocapsid protein that is surrounded by an envelope composed of glycoproteins such as haemagglutinin (*H*) and fusion (*F*) proteins [6].

The pathogenesis of measles virus infection is highly coordinated and involves an initial binding of H protein

(Figure 1) to host cell surface receptor targets, namely, the signaling lymphocyte activation molecule (CD150 or SLAM), CD 46, and the adherent junction protein PVRL4 (nectin-4) receptors [7]. The measles virus H protein head domain distinctively binds to a β-sheet of the membrane distal ectodomain of SLAM by its β-propeller fold and then triggers a conformational change on the F protein and extension into the target membrane [8, 9]. The H protein-host surface receptor (SLAM) interaction provides an excellent antimeasles virus drug target. The measles virus fusion (F) protein (Figure 1) plays an all-important function in the fusion of the virus and entry into the host cell. Membrane fusion machinery generated by the conformational change of F protein on the host cell leads to viral-cell membrane merging and cell-cell viral spread [9]. A second F protein conformational change results in the formation of fusion pores and subsequent entry of the genetic material (nucleocapsid) into the cytoplasm of the infected cell [5, 9].

While measles vaccines have been clinically administered over the years, the two recommended doses do not confer immunity in up to 10% of individuals due to inadequate development of antibodies [10]. Global eradication of the disease would require a 95% herd immunity that is achievable by the synergistic use of antimeasles virus drugs coupled with vaccination [5]. Unfortunately, there are presently no specific drugs approved for use in acute complications or persistent measles virus infections [11].

The resurgence of measles virus (MV) infection, a common cause of childhood morbidity and mortality especially in developing countries despite the availability of clinically effective measles vaccines [3, 4], requires a concerted effort in the discovery of antimeasles therapeutics to reduce its global human health impact. Previous research studies by Krumm and collaborators [12] reported a potential oral measles virus polymerase inhibitor. However, there is a dearth of information on antimeasles virus drugs. Therefore, this research study was designed to screen compounds isolated from *Xylopia aethiopica* that have been reported to demonstrate notable antimeasles virus activity *in vitro* [13] by employing molecular dynamics (MD) simulation techniques against essential druggable viral proteins of MV. Furthermore, in silico pharmacokinetic evaluation of the most promising compound was also undertaken and the entire protocols adopted in this study are presented in Figure 2.

2. Methodology

2.1. Acquisition and Preparation of Target Proteins and Compounds. The X-ray crystal structures of measles virus fusion protein (FUP), haemagglutination (HMG) protein, and phosphoprotein (PSP) with PDB codes 5YZC, 2ZB5, and 5E4V, respectively, were sourced from the RSCB Protein Data Bank (https://www.rcsb.org/). Water molecules and residue connectivity were removed from the protein structures, and the missing side chains were added to the structures using the UCSF Chimera software V1.14 [14, 15]. The 2D structures of phytocompounds, oxophoebine (OPB),

dihydrochalcone (DHC), and vanillin (VAN) and the ribavirin (RBV) standard were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov/). Optimization of the 3-D structures of the compounds that were used for the molecular docking was achieved by adding Gasteiger charges and nonpolar hydrogen atoms using Avogadro software [16, 17].

2.2. Molecular Docking. The prepared compounds (OPB, DHC, VAN, and RBV) were docked into the binding sites of measles virus fusion protein (FUP), haemagglutination protein (HMG), and phosphoprotein (PSP), using the Autodock Vina Plugin on Chimera. The grid box was defined with a spacing of 1 Å and appropriate sizes pointing in x, y, and z directions in each case. The docked complexes that showed the finest pose based on the docking scores were further subjected to molecular dynamics (MD) simulation.

2.3. Molecular Dynamics Simulation. The molecular dynamics (MD) simulation was performed by the method described by Idowu et al. [18]. The MD simulation was performed utilizing the GPU version provided with the ., in which the FF18SB variant of the AMBER force field was used to describe the systems [19]. The restrained electrostatic potential and the general amber force field procedures were used to generate atomic partial charges for the compounds, and the addition of hydrogen atoms, sodium, and chloride countered ions for neutralization of all systems in an ANTECHAMBER of the leap module of AMBER 18 [20]. The amino acid residues were numbered 1–81 for FUP, 1–330 for HMG, and 1–315 for PSP. The system was then suspended implicitly within an orthorhombic box of TIP3P water molecules such that all atoms were within 8 Å of any box edge.

First, 2000 steps of minimization were performed for the solutes in the initial 1,000 steps following the method of steepest descent, and a subsequent 1000 steps of conjugate gradients with an applied restraint potential of 500 kcal/mol. A further 1000 steps of full minimizations were performed by employing the conjugate gradient algorithm with no restraint. Stepwise heating from 0 to 300 K MD simulation was carried out for 50 ps in a manner that a fixed number of atoms and at a fixed volume were maintained and the systems equilibrated at 500 ps. A potential harmonic restraint of 10 kcal/mol and a collision frequency of 1.0 ps were imposed on the systems containing the solutes. In order to mimic an isobaric-isothermal ensemble (NPT), the operating temperature was fixed at 300 K using the Langevin thermostat, and the pressure was kept constant at 1 bar using the Berendsen barostat [21, 22], with randomized seeding, and a pressure-coupling constant of 2 ps [23]. The SHAKE algorithm was employed to constrict the bonds of hydrogen atoms during the MD simulations and was conducted for a period of 100 ns at a step size of 2 fs using the SPFP precision model [24, 25].
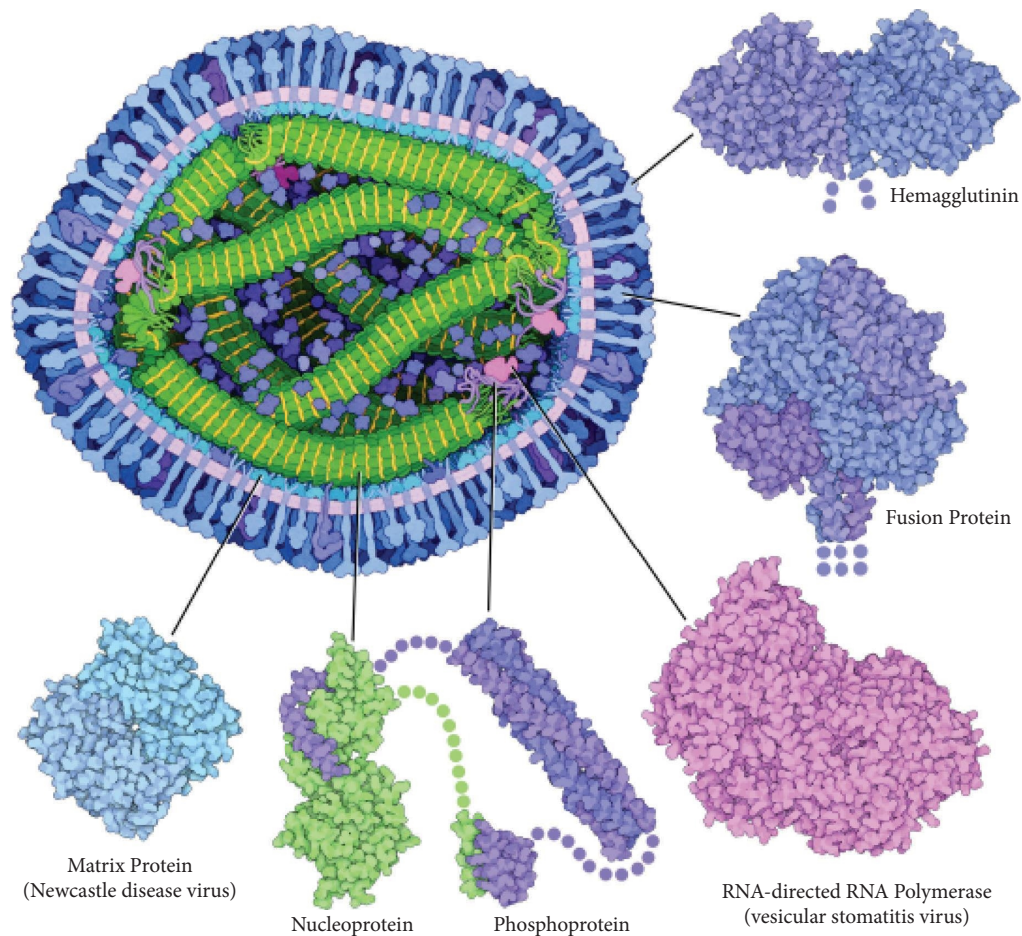
FIGURE 1: A cross-section of measles viral particle showing the significant proteins in the pathogenesis of measles virus infection (adapted from https://pdb101.rcsb.org/motm/231).
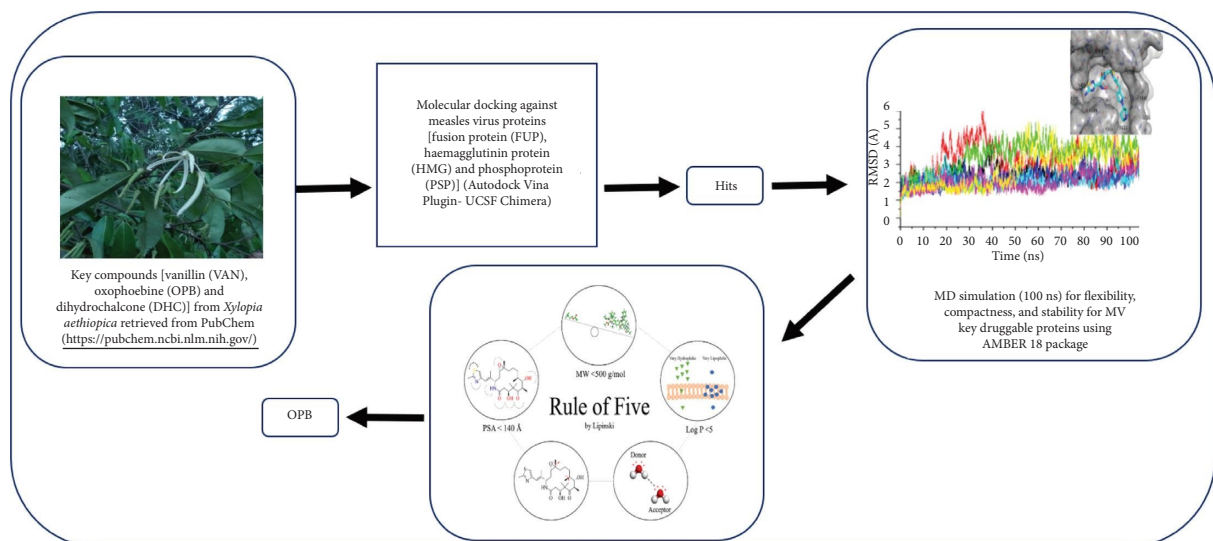
FIGURE 2: Workflow of the strategy adopted.

2.4. Postdynamic Analysis. Root mean square deviation (RMSD), root means square fluctuation (RMSF), solvent accessible surface area (SASA), and radius of gyration (RoG) were analyzed using the CPPTRAJ module employed in the AMBER 18 suite, and the raw data plots were subsequently generated by employing the Origin data analysis software [26].

2.5. Binding Free Energy Calculations. The free binding energy for each molecular species was calculated using the molecular mechanics with the generalized born surface area method (MM/GBSA) [27], and the binding affinity of the systems was compared. The mean binding free energy estimate was determined for over 100000 snapshots extracted from the 100 ns trajectory. Mathematically, the computation of the free binding energy (ΔG_{bind}) by employing this method is represented in the following equations:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand},$$
$$\Delta G_{bind} = E_{gas} + G_{sol} - TS,$$
$$E_{gas} = E_{int} + E_{vdw} + E_{ele}, \quad (1)$$
$$G_{sol} = G_{GB} + G_{SA},$$
$$G_{SA} = \gamma SASA,$$

where E_{gas} = gas-phase energy, E_{int} = internal energy, E_{elec} = Coulomb energy, E_{vdw} = van der Waals energies, G_{sol} = solvation free energy, GSA = nonpolar solvation energy, SASA = solvent-accessible surface area, GGB = polar solvation, S = total entropy of the solute, and T = temperature.

2.6. Analysis of Pharmacokinetic Properties of Test Compounds. The pharmacokinetic properties of the potential lead compound(s) were predicted using SwissADME online platform (http://swissadme.ch/index.php) as described by Daina et al. [28].

2.7. Data Analysis. Descriptive statistics and the OriginPro software (V9.10) were used to present ΔG_{bind} and all raw data plots, respectively [26].

3. Results

The molecular docking analysis for all the research compounds against the possible MV targets is presented in Table 1. The compounds VAN, OPB, and DHC had docking scores of −3.7, −5.4, and −4.5 kcal/mol, respectively, on FUP relative to −4.6 kcal/mol observed for the standard antiviral agent, RBV. The docking score of OPB on HMG was −8.1 kcal/mol and it was significantly higher when compared to the docking scores of VAN, −5.5 kcal/mol, and DHC, −6.6 kcal/mol. However, the docking score of OPB on HMG was comparable to the docking score of RBV at −7.0 kcal/mol. Findings from this study showed that the docking scores of OPB and RBV against PSP were −5.9 and −4.4 kcal/mol, respectively (Table 1).

Table 1 further shows the energy component profiles, including the binding free energy (ΔG_{bind}) values obtained through the 100 ns exposure period of the investigated systems. The binding affinities of OPB on measles virus FUP, HMG, and PSP were −25.37, −28.74, and −40.68 kcal/mol, respectively, and were significantly higher compared to the ΔG_{bind} of VAN, DHC, and RBV on the same targets.

The data on the predicted structural and conformational changes with the potential to modify the functioning of the investigated targets presented as RMSD, RoG, and RMSF of alpha carbon ($C\alpha$) atoms are shown in Table 2, while the plots are presented in Figures 3(a)–3(f).

Figure 3(a) shows the RMSD value of the OPB-FUP complex (18.96 Å) that was higher than the RMSD values of the RBV-FUP complex (15.92 Å) and unbound FUP (14.30 Å). Observed RMSD values of the molecules OPB (3.45 Å) and RBV (3.05 Å) exhibited relatively similar deviations when in an HMG system. However, the unbound system produced an RMSD value of 2.50 Å (Table 2 and Figure 3(c)). The RMSD values for the unbound PSP and the complexes formed were higher than the values derived from the unbound HMG and its complexes (Table 2 and Figure 3(e)). The RoG plot of FUP-OPB and FUP-RBV systems produced average RoG values of 17.19 and 17.71 Å, respectively. These values were lower than 24.48 and 19.11 Å produced by DHC and the unbound FUP, respectively (Table 2 and Figure 3(b)). However, average RoG values of HMG and PSP systems with the experimental drug molecules and the respective unbound proteins were similar (Table 2 and Figures 3(d) and 3(f)).

The root mean square fluctuation (RMSF) values of alpha carbon atoms after molecular dynamics simulation of the target proteins-ligands interactions are presented in Figure 4. Oxophoebine gave the highest observed RMSF value of 8.78 Å and was comparable with 8.92 Å for the unbound FUP system. The standard antiviral, RBV and DHC recorded RMSF values of 6.04 Å and 7.43 Å, respectively, and these were lower than that of the unbound FUP (Table 2 and Figure 4(a)). The RMSF values of OPB (1.43 Å) and RBV (1.32 Å) were similar to 1.44 Å recorded for the unbound HMG system. In PSP systems, the RSMF values of OPB, RBV, and the unbound PSP system were calculated to be 3.08, 2.17, and 2.68 Å, respectively (Table 2 and Figures 4(b) and 4(c)). The SASA investigations of the FUP complex systems showed respective average values of 7080.96 Å and 8007.82 Å for OPB and DHC against RBVs average of 6963.91 Å; while the average values were closely comparable for unbound HMG protein and the complex system it formed with OPB and RBV standards (Table 2 and Figures 4(d) and 4(e)). A similar observation was drawn for the unbound PSP protein and PSP-OPB complex, but the PSP-OPB complex had the highest average SASA value at 23578.47 Å (Table 2 and Figure 4(f)).

The results from post-MD simulation interactions in FUP bound systems with OPB showed 18 interactions, compared to 9 and 7 interactions for DHC and RBV, respectively (Figures 5(a)–5(f)). Oxophoebine interacted with 14 amino acid residues of the HMG protein forming van der Waal's bonds, as well as conventional H (Ser328), 2 alkyl (Pro277 and Leu327), 3 C-H (Met279, Pro278, and Ser329), and 1 pi-pi$T$-shaped (His257) interactions (Figures 6(a) and 6(b)). On the other hand, a total of 9 interactions were observed for the HMG-RBV system including 3 C-H (Lys280, Ala283, and Pro277), van der Waal's, and conventional hydrogen interactions (Figures 6(c) and 6(d)). The PSP-bound systems with OPB and RBV showed 19 and 16 total interactions, respectively (Figures 7(a)–7(d)). The PSP-OPB interactions were conventional H (Trp294), 3 C-H

TABLE 1: Docking score and thermodynamic binding free energy of the compounds with MV druggable targets.

Target/compound	Docking score (kcal/mol)	Energy components (kcal/mol)				
		ΔE_{vdW}	ΔE_{elec}	ΔG_{gas}	ΔG_{solv}	ΔG_{bind}
FUP						
VAN	−3.7	—	—	—	—	—
RBV	−4.6	−20.42 ± 4.1	−14.53 ± 7.7	−34.95 ± 9.6	17.72 ± 6.3	−17.23 ± 4.5
OPB	−5.4	−33.30 ± 6.0	−8.09 ± 3.9	−41.40 ± 8.4	16.03 ± 4.4	−25.37 ± 5.1
DHC	−4.5	−16.51 ± 11.2	−2.42 ± 3.5	−18.93 ± 13.0	5.79 ± 4.7	−13.14 ± 9.1
HMG						
VAN	−5.5	—	—	—	—	—
RBV	−7.0	−25.16 ± 4.6	−31.50 ± 13.2	−56.66 ± 14.6	37.63 ± 10.2	−19.03 ± 6.7
OPB	−8.1	−41.85 ± 4.4	−16.84 ± 5.3	−58.70 ± 8.0	29.95 ± 4.2	−28.74 ± 5.1
DHC	−6.6	—	—	—	—	—
PSP						
VAN	−5.0	—	—	—	—	—
RBV	−4.4	−28.77 ± 4.9	−18.76 ± 15.2	−47.53 ± 15.1	28.09 ± 10.9	−19.43 ± 6.4
OPB	−8.0	−51.46 ± 2.9	−17.43 ± 3.9	−68.89 ± 5.4	28.20 ± 3.3	−40.68 ± 3.7

VAN = vanillin; OPB = oxophoebine; DHC = dihydrochalcone; RBV = ribavirin; FUP = fusion protein; HMG = haemagglutinin protein; PSP = phosphoprotein; ΔE_{elec} = electrostatic energy; ΔE_{vdW} = van der Waals energy; ΔG_{bind} = total binding free energy; ΔG_{solv} = solvation free energy; ΔE_{gas} = gas phase free energy. Data presented are ± standard deviation of average values, — = not determined.

TABLE 2: Postmolecular dynamics simulation analysis of interactions of compounds of *Xylopia aethiopica* with measles virus druggable targets.

Target/compound	Average RMSD (Å)	Average RMSF (Å)	Average RoG (Å)	Average SASA (Å)
FUP	14.30 ± 3.19	8.92 ± 2.70	19.11 ± 2.96	7433.19 ± 531.60
FUP-RBV	15.92 ± 2.43	6.04 ± 2.20	17.71 ± 2.57	6963.91 ± 650.40
FUP-OPB	18.96 ± 3.55	8.78 ± 2.57	17.19 ± 2.64	7080.96 ± 435.40
FUP-DHC	12.58 ± 1.85	7.43 ± 3.09	24.48 ± 1.61	8007.82 ± 291.40
HMG	2.50 ± 0.32	1.44 ± 0.10	21.59 ± 0.095	19429.50 ± 399.60
HMG-RBV	3.05 ± 0.36	1.32 ± 0.11	21.62 ± 0.087	19288.04 ± 440.85
HMG-OPB	3.45 ± 0.33	1.43 ± 0.14	21.48 ± 0.061	19217.43 ± 348.90
PSP	5.92 ± 1.02	2.68 ± 2.03	26.07 ± 0.89	22837.87 ± 682.54
PSP-RBV	5.66 ± 1.07	2.17 ± 1.29	26.23 ± 0.89	22795.77 ± 682.54
PSP-OPB	5.62 ± 1.42	3.08 ± 2.64	27.79 ± 0.94	23578.47 ± 505.89

FUP = fusion protein; HMG = haemagglutination protein; PSP = phosphoprotein; RBV = ribavirin; OPB = oxophoebine; DHC = dihydrochalcone; RMSD = root mean square deviation; RMSF = root means square fluctuation; RoG = radius of gyration; SASA = solvent accessible surface area. Data presented are ± standard deviation of average values.

(Ser286, Asn312, and Gly226), 4 alkyl, 1 pi-Cation (Arg315), 1 pi-anion (Asp147), 1 pi-sigma (Ala228), and van der Waal's bonds (Figures 7(a) and 7(b)), while the PSP-RBV system had conventional H, 1 C-H (Gly310), 3 pi-alkyl, 1 attractive charge (Glu153), and van der Waal's bonds interactions (Figures 7(c) and 7(d)).

The pharmacokinetic and physicochemical properties of the promising compound, oxophoebine, and ribavirin are presented in Table 3. The molecular weights and $\log P$ of oxophoebine (365.34 g/mol and 2.97) and ribavirin (244.20 g/mol and 0.13) were below 500 g/mol and 5, respectively. Oxophoebine was predicted to have high gastrointestinal absorption, while ribavirin was predicted to have low gastrointestinal absorption. The experimental compound (oxophoebine) and the standard antiviral drug share the same bioavailability score of 0.55 and the same number of hydrogen acceptors which is 7, but below 10.

4. Discussion

In this study, three compounds isolated from *Xylopia aethiopica* were computationally evaluated against key structural proteins of MV and the higher docking scores observed with OPB and DHC relative to the low score of VAN could be an indication of better fitness and orientation of OPB and DHC at the binding sites of FUP, HMG, and PSP. The scoring provides the necessary information to predict the binding affinity of the prospective drug to the target [29]. Binding energy calculations and MD simulation indicated that OPB and DHC showed better poses or best fit on the MV target proteins. As such, these were preferentially selected and subsequently subjected to MDS analysis. The MM/GBSA technique estimates the total binding free energy (ΔG_{bind}) of an inhibitor that could cause structural and conformational changes that may alter the biological activities of the target protein [18, 30, 31]. The higher negative

(a)

(b)

(c)

(d)

(e)

(f)
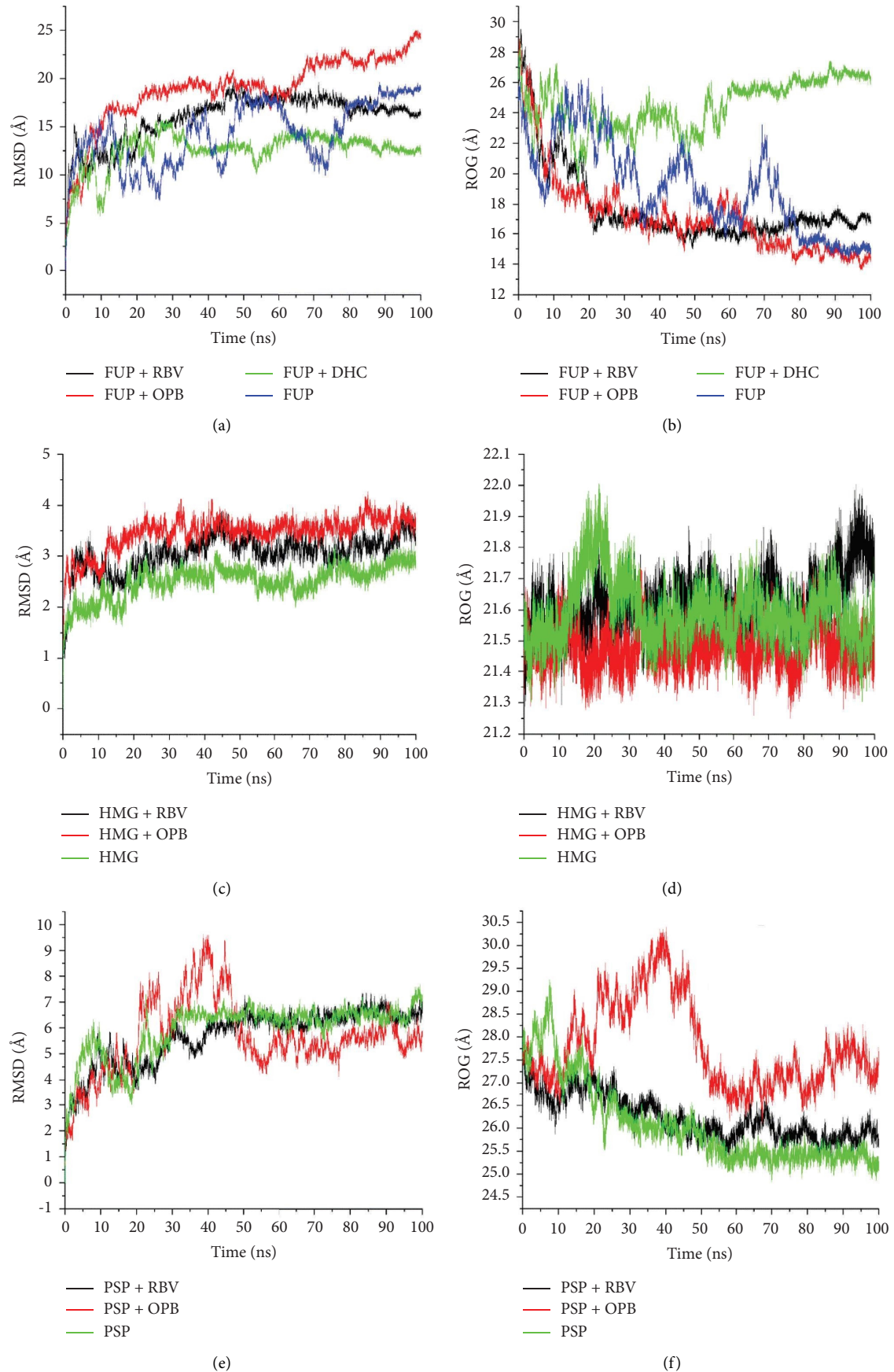
Figure 3: Comparative plots of the FUP, HMG, and PSP complexes with OPB, DHC, and RBV presented as RMSD (a), (c), and (e) and RoG (b), (d), and (f). FUP = fusion protein, PSP = phosphoprotein, and HMG = haemagglutinin protein.
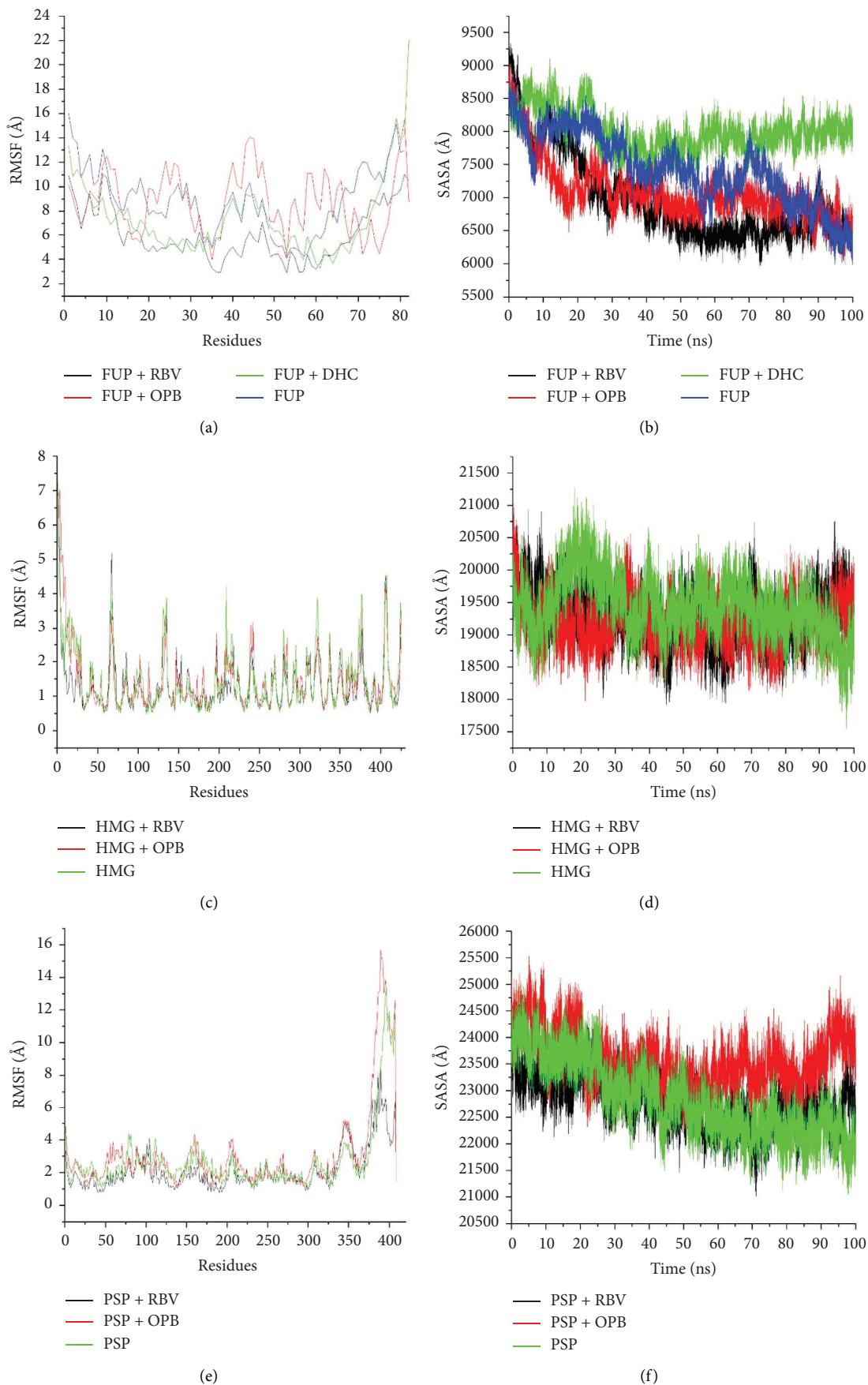
Figure 4: Comparative plots of FUP, HMG, and PSP with molecules OPB, DHC, and RBV presented as RMSF (a), (c), and (e) and SASA (b), (d), and (f). FUP = fusion protein, HMG = haemagglutinin protein, and PSP = phosphoprotein.

(a)

(b)

(c)

(d)

(e)

(f)
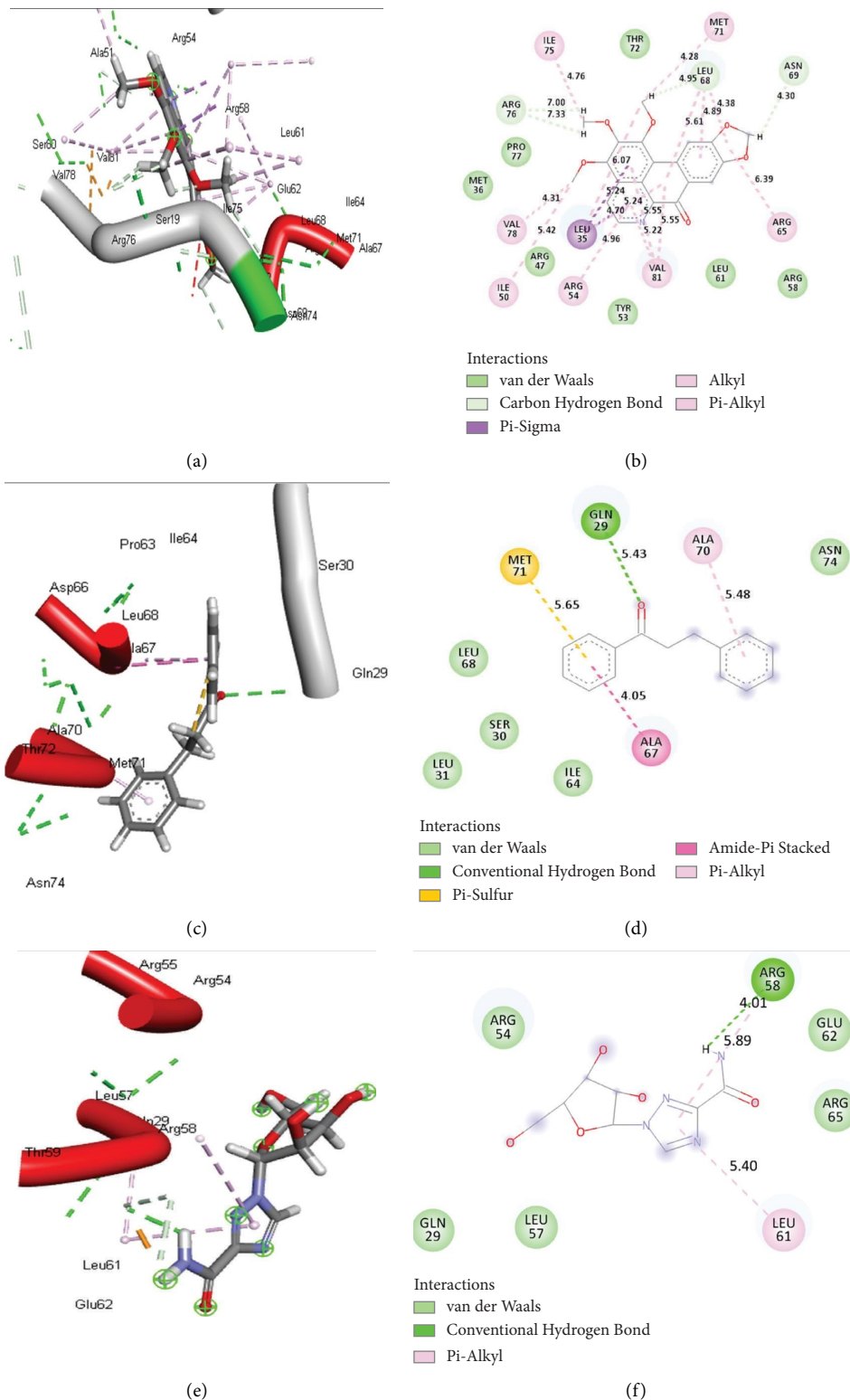
FIGURE 5: The 3D (a), (c), and (e) and 2D (b), (d), and (f) interaction plots of FUP-OPB, FUP-DHC, and FUP-RBV, respectively, with 18, 9, and 7 bonds. FUP = measles virus fusion protein, OPB = oxophoebine, DHC = dihydrochalcone, and RBV = ribavirin.

binding affinity for the MV proteins, FUP, HMG, and PSP observed with OPB in this study may indicate better interactions that could lead to alterations in the normal functioning of these proteins. Although, the binding affinity observed for DHC on FUP was significantly lower compared to OPB on FUP, and it was rather comparable to ribavirin

(a)

(b)

(c)

Interactions
- van der Waals
- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Pi T-shaped
- Alkyl
- Pi-Alkyl

Interactions
- van der Waals
- Conventional Hydrogen Bond
- Carbon Hydrogen Bond

(d)
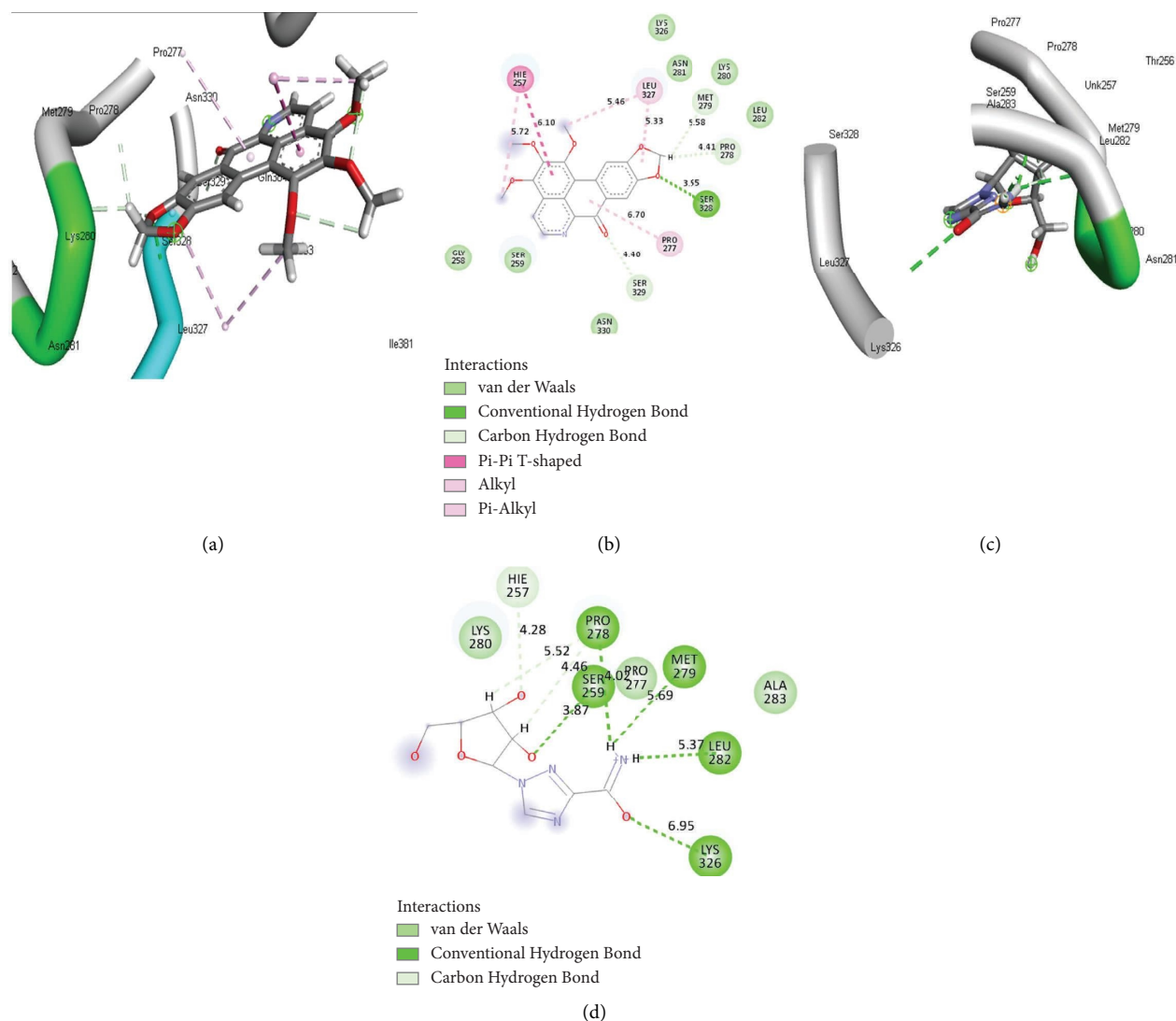
FIGURE 6: The 3D (a) and (c) and 2D (b) and (d) interaction plots of HMG-OPB and HMG-RBV systems with 14 and 9 bonds, respectively. HMG = measles virus haemagglutinin protein, OPB = oxophoebine, and RBV = ribavirin.

(RBV), the standard antiviral agent. This observation suggests that DHC may not significantly alter the normal biological activities of the MV fusion protein.

The RMSD measures the degree of convergence, stability, or deviations produced by a protein in a simulation system [32]. In this study, though the lower mean RMSD of the FUP-DHC system relative to the unbound FUP could imply that DHC stabilized FUP than OPB and RBV, but this was, however, not consistent with the thermodynamic binding energy value and may represent a pseudostability of the complex. The trend for the simulated OPB-HMG was not different from that observed with the FUP system. In contrast to this observation, the marginally lower mean RMSD value of the PSP-OPB complex than that of the apo PSP compared with RBV could be suggestive of OPB's potential as a prospective lead compound, and perhaps a novel inhibitor of MV PSP. Interestingly, this result agrees with the thermodynamic profiles where the lowest binding free energy correlates with greater compactness and

structural stability of the complex. The compound, OPB lowered the mean RoG value of FUP resulting in more structural compactness and stability compared to the unbound FUP which further suggests that OPB is a potential inhibitor of FUP. However, in the case of HMG, the RoG value of the protein-bound molecules were like those of the unbound protein and OPB seemed not to have changed the compactness of the target protein indicating that HMG may not be an important target for the inhibitory activity of OPB. The flexibility and structural compactness of the bound target proteins is a function of the radius of gyration (RoG), and binding of the drug molecule to the protein may lead to a change in biological function. A low RoG value for a drug molecule is an indication of a more stable system relative to the unbound protein [31].

The RMSF value is used to determine the fluctuations at the active site of a protein when bound to a drug molecule, and higher or lower α-carbon fluctuations imply more or less flexible movements, respectively [33]. This justifies the

(a)

(b)

(c)

(d)
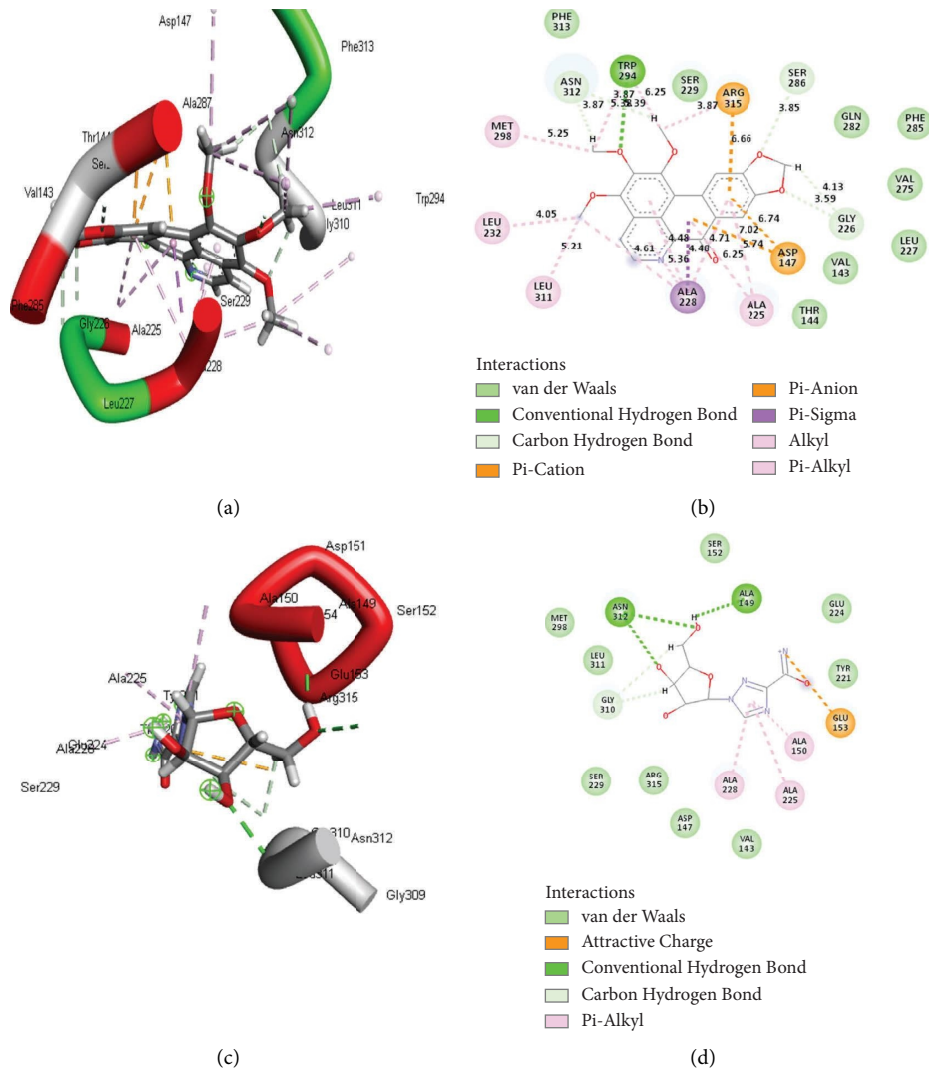
FIGURE 7: The 3D (a) and (c) and 2D (b) and (d) interaction plots of PSP-OPB and PSP-RBV systems with 19 and 16 bonds, respectively. PSP = measles virus phosphoprotein, OPB = oxophoebine, and RBV = ribavirin.

determination of RMSF value in this research study. In this study, similar fluctuation patterns were seen in the residues, 0–10, 75–200, and 100–350 in all the FUP, HMG, and PSP systems, respectively. The estimated RMSF values of OPB-FUP or OPB-HMG systems were not different from those of the unbound proteins which indicates that the compound did not exact any reasonable dynamic alterations at the active sites of the target proteins residues and is predictive of the inhibitory activity of OPB. The SASA is useful to determine the surface area of the drug-protein complex accessibility to solvents, and the impact on the drug molecule on SASA [34]. When solvent accessibility surface area values are high, it indicates reduced structural/system stability and increased surface area, and vice versa [35]. The respective lower SASA values of the OPB-HMG and OPB-FUP systems relative to the standard indicated reduced protein residue-solvent molecule interactions and targets' surface area. These results suggested increased systems stability and emphasized the compounds' potential as

treatment or drug options against such targets. However, OPB caused the most significant structural entropy with all target proteins studied compared to other test compounds. This observation is in agreement with its lower free binding energy values against all targets compared to the standard.

Compounds with inhibitory activity usually bind to amino acids at the active sites of a protein to inactivate it. However, binding is dependent on thermodynamic parameters linked to protein interaction types, protein compactness, and the stability and flexibility of the amino acid residues [36, 37]. The number, nature, and length of bonds of interactions of the test drug candidates and standard antiviral drugs at the active sites of FUP, HMG, and PSP proteins varied, and it impacted the reported binding free energy values. Among observed bond, interactions were the C-H bonds, conventional hydrogen bonds, pi-piT-shaped, van der Waals, amide pi-stacked, pi-cation, pi-anion, pi-sigma, pi-pi-stacked, and attractive charges. The OPB-bound systems generally had the highest number of interactions with

TABLE 3: Comparison of predicted ADMET properties of oxophoebine and ribavirin.

Property	Molecule (ligand)	
	Oxophoebine (OPB)	Ribavirin (RBV)
Mol. formula	$C_{20}H_{15}NO_6$	$C_8H_{12}N_4O_5$
Mol. weight (g/mol)	365.34	244.20
Lipophilicity (iLog P)	2.97	0.13
Water solubility	Moderate	Low
GIT absorption	High	Low
BBB permeability	Yes	No
Hydrogen bond acceptors	7	7
Hydrogen bond donors	0	4
Bioavailability score	0.55	0.55
Drug likeness (Lipinski)	Yes	Yes

Mol. = molecular; Log P = partition coefficient; GIT = gastrointestinal tract; BBB = blood-brain barrier.

all protein targets, which was in direct agreement with the high ΔG_{bind} scores for OPB-bound systems compared to DHC and RBV-bound systems. The number of hydrogen bonds was also highest in OPB compound-protein systems with exception of HMG-RBV systems. Hydrogen bonds have unusually strong interactions between molecules and constitute one of the most essential bonds in drug discovery [38]. Nonetheless, the HMG-RBV system interactions had higher average bond lengths (5.02 Å) compared to the HMG-OPB (4.48 Å) complex system and accounted for the low free binding energy of the former. Bond length (BL) is important in determining the affinity/tightness of atoms, and shorter BLs contribute to stronger intramolecular or intermolecular pull, greater atomic hold, and affinity [39, 40]. Again, among the test compounds simulated, only oxophoebine (OPB) interacted with a hydrogen bond linked to an essential amino acid residue, Trp294, at the active site of PSP (Figure 6). This gives an insight into the compound's affinity for the PSP active site and could be responsible for the observed higher binding free energy.

In this study, OPB showed a potential inhibitory effect on the biological activity of the HMG and fusion proteins of the MV. The effect of OPB as a potential MV fusion inhibitor could be similar to its blockage of the human immunodeficiency virus (HIV)-1 which envelops glycopeptides fusion with the cell membrane of host CD4 cells [41], as well as the anti-MV effects of the small compound, AS-48 via binding to hydrophobic pockets in the region of measles virus fusion protein [42]. The PSP, much like the nucleoprotein, also binds to the RNA polymerase during viral replication [43, 44]. Interference with the process of PSP and RNP complex formation leads to altered transcription and efficiency in viral particle assembly, viral replication inhibition, or death. The inhibition of PSP by OPB may be closely linked to this mechanism of action, that is, via PSP-RNP complex formation. This effect, like that of small molecules previously identified and characterized by Krumm et al. [12] could ultimately result in the inhibition of viral replication and death of MV. Measles virus HMG and fusion proteins mediate the binding of the viral envelope to host cell surface proteins and the process of virus-cell fusion, respectively. Inhibition of the HMG protein-host would interfere with the interaction between MV surface receptor, signaling lymphocyte activation molecule (SLAM) that might result in the inhibition of the required initial conformational changes of the fusion protein, a rate-limiting step in the pathogenesis of MV infection. In addition, compounds with an antifusion activity that interfere with conformational changes of FUP would exhibit potent measles virus fusion inhibition [6, 9]. The measles virus HMG and fusion proteins are therefore promising targets for the discovery and development of MV fusion inhibitors [11].

Pharmacokinetics study showed log P values of the OPB compound and the standard drug, ribavirin to be less than 5, suggesting good lipid membrane permeability and absorption, and the potential of OPB formulation for oral administration. In-silico pharmacokinetics values give an overview of *in vivo* drug interactions and decrease the chances of failure during drug development [45]. Lipinski's rule of five (Ro5) gives an indication of the physicochemical properties, drug-likeness, and safety properties of prospective bioactive compounds intended for oral administration [46]. The Ro5 states that compounds that have log P values >5 may poorly permeate the lipid membrane and not pass through the gut walls [47, 48]. A high rate of absorption in the gastrointestinal tract (GIT) is a major consideration during oral drugs design [49]. Again, both OPB and RBV showed a good bioavailability score of 0.55. A drug's bioavailability score (the proportion of a drug dose that remains unchanged and reaches the systemic circulation) is also key in drug dose calculations [50]. A high bioavailability score indicates a high drug concentration needed to elicit maximum therapeutic effects at the site of action. Still, bioactive molecules with low oral bioavailability could be formulated for administration via nonoral routes [50]. The compound OPB was also predicted to permeate the blood-brain barrier (BBB), but not the standard drug. It hence possesses an added advantage if used in the management of MV-induced encephalomyelitis. The BBB prevents the passage of toxic compounds of molecular weight greater than 400 g/mol [51]. Our results suggest that both OPB and RBV molecules are good drug candidates for oral administration against MV. Lastly, this study's findings also support the previous *in vitro* antimeasles virus report on *Xylopia aethiopica* [13], with OPB as the most promising compound responsible for the elicited activity.

5. Conclusion

The glycoproteins (haemagglutinin and fusion proteins) and phosphoprotein play integral functions in the replication of the measles virus and pathogenesis of measles, thus, finding potent and clinically useful inhibitors for these proteins is undoubtedly important in the treatment of measles disease. The drug candidate, oxophoebine (OPB), that showed similar or better physicochemical and pharmacokinetic properties compared to the FDA-approved antiviral drug, ribavirin might possess multiple-target inhibitory activities against the measles virus. Hence, this compound could serve

as a scaffold for the development of inhibitors of MV fusion, haemagglutinin proteins, and phosphoproteins.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. E. Griffin and W. J. Bellini, "Measles virus," in *Virology*, B. N. Fields, D. M. Knipe, and P. M. Howley, Eds., pp. 1267–1312, Lippincott-Raven, Philadelphia, PA, USA, 1996.

[2] Centre for Disease Control, *Measles*, US Department of Health and Human Services, Washington, DC, USA, 2020.

[3] World Health Organization, *Measles*, World Health Organization, Geneva, Switzerland, 2019.

[4] O. Kola, "Nigeria: 1,158 displaced children infected by measles," 2021, https://www.aa.com.tr/en/africa/nigeria-1-158-displaced-children-infected-by-measles/2203817.

[5] P. Plattet, L. Alves, M. Herren, and H. C. Aguilar, "Measles virus fusion protein: structure, function and inhibition," *Viruses*, vol. 8, no. 4, p. 112, 2016.

[6] J. C. Welsch, A. Talekar, C. Mathieu et al., "Fatal measles virus infection prevented by brain-penetrant fusion inhibitors," *Journal of Virology*, vol. 87, no. 24, pp. 13785–13794, 2013.

[7] J. B. Patterson, D. Thomas, H. Lewicki, M. A. Billeter, and M. B. Oldstone, "V and C Proteins of measles virus function as virulence factors in vivo," *Virology*, vol. 267, no. 1, pp. 80–89, 2000.

[8] T. Hashiguchi, T. Ose, M. Kubota et al., "Structure of the measles virus hemagglutinin bound to its cellular receptor SLAM," *Nature Structural & Molecular Biology*, vol. 18, no. 2, pp. 135–141, 2011.

[9] C. Mathieu, D. Huey, E. Jurgens et al., "Prevention of measles virus infection by intranasal delivery of fusion inhibitor peptides," *Journal of Virology*, vol. 89, no. 2, pp. 1143–1155, 2015.

[10] G. A. Poland and R. M. Jacobson, "The re-emergence of measles in developed countries: time to develop the next-generation measles vaccines?" *Vaccine*, vol. 30, no. 2, pp. 103-104, 2012.

[11] J. M. White, S. E. Delos, M. Brecher, and K. Schornberg, "Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 43, no. 3, pp. 189–219, 2008.

[12] S. A. Krumm, D. Yan, E. S. Hovingh et al., "An orally available, small-molecule polymerase inhibitor shows efficacy against a lethal morbillivirus infection in a large animal model," *Science Translational Medicine*, vol. 6, Article ID 232ra52, 2014.

[13] B. B. Oluremi and J. A. Adeniji, "Anti-viral activity evaluation of selected medicinal plants of Nigeria against Measles virus," *British Microbiology Research Journal*, vol. 7, no. 5, pp. 218–225, 2015.

[14] Z. Yang, K. Lasker, D. Schneidman-Duhovny et al., "UCSF Chimera, MODELLER, and IMP: an integrated modeling system," *Journal of Structural Biology*, vol. 179, no. 3, pp. 269–278, 2012.

[15] G. Branden, T. Sjogren, V. Schnecke, and Y. Xue, "Structure-based ligand design to overcome CYP inhibition in drug discovery projects," *Drug Discovery Today*, vol. 19, no. 7, pp. 905–911, 2014.

[16] S. Kim, P. A. Thiessen, E. E. Bolton et al., "PubChem substance and compound databases," *Nucleic Acids Research*, vol. 44, no. D1, Article ID D1202, 2016.

[17] J. O. Aribisala, R. A. Abdulsalam, Y. Dweba, K. Madonsela, and S. Sabiu, "Identification of secondary metabolites from Crescentia cujete as promising antibacterial therapeutics targeting type 2A topoisomerases through molecular dynamics simulation," *Computers in Biology and Medicine*, vol. 145, Article ID e105432, 2022.

[18] K. Idowu, P. Ramharack, M. Nlooto, and M. Gordon, "Molecular dynamic mechanism(s) of inhibition of bioactive antiviral phytochemical compounds targeting cytochrome P450 3A4 and P-glycoprotein," *Journal of Biomolecular Structure and Dynamics*, vol. 3, pp. 1221–1233, 2020.

[19] P. C. Nair and J. O. Miners, "Molecular dynamics simulations: from structure function relationships to drug discovery," *In Silico Pharmacology*, vol. 2, no. 1, p. 4, 2014.

[20] J. O Uhomoibhi, K. A. Idowu, F. O. Shode, and S. Sabiu, "Molecular modelling identification of potential drug candidates from selected African plants against SARS-CoV-2 key druggable proteins," *Scientific African*, vol. 17, Article ID e01279, 2022.

[21] J. E. Basconi and M. R. Shirts, "Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 9, no. 7, pp. 2887–2899, 2013.

[22] P. Gonnet, "P-SHAKE: a quadratically convergent SHAKE in," *Journal of Computational Physics*, vol. 220, no. 2, pp. 740–750, 2007.

[23] J. A. Izaguirre, D. P. Catarello, J. M. Wozniak, and R. D. Skeel, "Langevin stabilization of molecular dynamics," *The Journal of Chemical Physics*, vol. 114, no. 5, pp. 2090–2098, 2001.

[24] J. P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

[25] J. O. O. Uhomoibhi, F. O. Shode, K. A. Idowu, and S. Sabiu, "Molecular modelling identification of phytocompounds from selected African botanicals as promising therapeutics against druggable human host cell targets of SARS-CoV-2," *Journal of Molecular Graphics and Modelling*, vol. 114, Article ID 108185, 2022.

[26] E. Seifert, "OriginPro 9.1: scientific data analysis and graphing software-software review," *Journal of Chemical Information and Modeling*, vol. 54, no. 5, p. 1552, 2014.

[27] M. Ylilauri and O. T. Pentikainen, "MMGBSA as a tool to understand the binding affinities of filamin-peptide

interactions," *Journal of Chemical Information and Modeling*, vol. 53, no. 10, pp. 2626–2633, 2013.

[28] A. Daina, O. Michielin, and V. Zoete, "SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Scientific Reports*, vol. 7, no. 1, Article ID 42717, 2017.

[29] M. Abdullahi, F. A. Olotu, and M. E. Soliman, "Allosteric inhibition abrogates dysregulated LFA-1 activation: structural insight into mechanisms of diminished immunologic disease," *Computational Biology and Chemistry*, vol. 73, pp. 49–56, 2018.

[30] R. O. Kumi, O. S. Soremekun, A. R. Issahaku, C. Agoni, F. A. Olotu, and M. E. S. Soliman, "Exploring the ring potential of 2,4-diaminopyrimidine derivatives towards the identification of novel caspase-1 inhibitors in Alzheimer's disease therapy," *Journal of Molecular Modeling*, vol. 26, no. 4, p. 68, 2020.

[31] T. Sindhu and P. Srinivasan, "Exploring the binding properties of agonists interacting with human TGR5 using structural modeling, molecular docking and dynamics simulations," *RSC Advances*, vol. 5, no. 19, Article ID 14202, 2015.

[32] B. Hess, "Convergence of sampling in protein simulations," *Physical Review*, vol. 65, no. 3, Article ID 31910, 2002.

[33] F. O. Shode, A. S. K. Idowu, O. J. Uhomoibhi, and S. Sabiu, "Repurposing drugs and identification of inhibitors of integral proteins (spike protein and main protease) of SARS-CoV-2," *Journal of Biomolecular Structure and Dynamics*, vol. 40, no. 14, pp. 6587–6602, 2022.

[34] S. Gupta, A. K. Singh, P. P. Kushwaha et al., "Identification of potential natural inhibitors of SARS-CoV2 main protease by molecular docking and simulation studies," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 12, pp. 4334–4345, 2021.

[35] S. S. Mousavi, A. Karami, T. M. Haghighi et al., "In silico evaluation of Iranian medicinal plant phytoconstituents as inhibitors against main protease and the receptor-binding domain of SARS-CoV-2," *Molecules*, vol. 26, no. 18, p. 5724, 2021.

[36] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nature Reviews Drug Discovery*, vol. 3, no. 11, pp. 935–949, 2004.

[37] M. A. F. Nasution, E. P. Toepak, A. H. Alkaff, and U. S. F. Tambunan, "Flexible docking-based molecular dynamics simulation of natural product compounds and Ebola virus Nucleocapsid (EBOV NP): a computational approach to discover new drug for combating Ebola," *BMC Bioinformatics*, vol. 19, no. S14, pp. 419–149, 2018.

[38] H. Izadi, K. M. E. Stewart, and A. Penlidis, "Role of contact electrification and electrostatic interactions in gecko adhesion," *Journal of The Royal Society Interface*, vol. 11, no. 98, Article ID 20140371, 2014.

[39] X. Du, Y. Li, Y. L. Xia et al., "Insights into protein–ligand interactions: mechanisms, models, and methods," *International Journal of Molecular Sciences*, vol. 17, no. 2, p. 144, 2016.

[40] J. O. Aribisala and S. Sabiu, "Cheminformatics identification of phenolics as modulators of penicillin-binding protein 2a of Staphylococcus aureus: a structure activity-relationship-based study," *Pharmaceutics*, vol. 14, no. 9, p. 1818, 2022.

[41] X. Zhang, Y. Zhu, H. Hu et al., "Structural insights into the mechanisms of action of short-peptide HIV-1 fusion inhibitors targeting the Gp41 pocket," *Frontiers in Cellular and Infection Microbiology*, vol. 8, p. 51, 2018.

[42] T. Hashiguchi, Y. Fukuda, R. Matsuoka et al., "Structures of the prefusion form of measles virus fusion protein in complex with inhibitors," *Proceedings of the National Academy of Sciences*, vol. 115, no. 10, pp. 2496–2501, 2018.

[43] J. Curran, J.-B. Marq, and D. Kolakofsky, "An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication," *Journal of Virology*, vol. 69, no. 2, pp. 849–855, 1995.

[44] D. Spehner, R. Drillien, and P. M. Howley, "The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein," *Virology*, vol. 232, no. 2, pp. 260–268, 1997.

[45] F. Yamashita and M. Hashida, "In silico approaches for predicting ADME properties of drugs," *Drug Metabolism and Pharmacokinetics*, vol. 19, no. 5, pp. 327–338, 2004.

[46] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews*, vol. 64, pp. 4–17, 2012.

[47] M. Remko, A. Bohac, and L. Kovacikova, "Molecular structure, pKa, lipophilicity, solubility, absorption, polar surface area, and blood brain barrier penetration of some antiangiogenic agents," *Structural Chemistry*, vol. 22, no. 3, pp. 635–648, 2011.

[48] L. A. Velez, Y. Delgado, Y. Ferrer-Acosta, I. J. Suarez-Arroyo, P. Rodriguez, and D. Perez, "Theoretical prediction of gastrointestinal absorption of phytochemicals," *International Journal of Plant Biology*, vol. 13, no. 2, pp. 163–179, 2022.

[49] P. Kremers, "In vitro tests for predicting drug-drug interactions: the need for validated procedures," *Pharmacology & Toxicology*, vol. 91, no. 5, pp. 209–217, 2002.

[50] R. P. Heaney, "Factors influencing the measurement of bioavailability, taking calcium as a model," *Journal of Nutrition*, vol. 131, no. 4, p. 1344S, 2001.

[51] D. J. Begley and M. W. Brightman, "Structural and functional aspects of the blood-brain barrier," *Peptide Transport and Delivery into the Central Nervous System*, vol. 61, no. 5, pp. 39–78, 2011.

Research Article

Combining the External Medical Knowledge Graph Embedding to Improve the Performance of Syndrome Differentiation Model

Qing Ye [ID], **Rui Yang** [ID], **Chun-lei Cheng** [ID], **Lin Peng, and Yong Lan**

School of Computer, Jiangxi University of Chinese Medicine, Nanchang, Jiangxi, China

Correspondence should be addressed to Chun-lei Cheng; chunlei_cheng@163.com

The electronic medical records (EMRs) of traditional Chinese medicine (TCM) include a wealth of TCM knowledge and syndrome diagnosis information, which is crucial for improving the quality of TCM auxiliary decision-making. In practical diagnosis, one disease corresponds to one syndrome, posing considerable hurdles for the informatization of TCM. The purpose of this work was to create an end-to-end TCM diagnostic model, and the knowledge graph (KG) created in this article is used to improve the model's information and realize auxiliary decision-making for TCM disorders. We approached auxiliary decision-making for syndrome differentiation in this article as a multilabel classification task and presented a knowledge-based decision support model for syndrome differentiation (KDSD). Specifically, we created a KG based on TCM features (TCMKG), supplementing the textual representation of medical data with embedded information. Finally, we proposed fusing medical text with KG entity representation (F-MT-KER) to get prediction results using a linear output layer. After obtaining the vector representation of the medical record text using the BERT model, the vector representation of various KG embedded models can provide additional hidden information to a certain extent. Experimental results show that our method improves by 1% (P@1) on the syndrome differentiation auxiliary decision task compared to the baseline model BERT. The usage of EMRs can aid TCM development more efficiently. With the help of entity level representation, character level representation, and model fusion, the multilabel classification method based on the pretraining model and KG can better simulate the TCM syndrome differentiation of the complex cases.

1. Introduction

With the fast advancement of the medical information technology in recent years, the number of electronic medical records (EMRs) accessible has increased. EMRs are mostly semistructured or unstructured text records of the diagnostic and treatment processes. EMRs serve as the central repository for big medical data, which contains a wealth of medical information and therapeutic concepts. The course record in medical records can be divided into four diagnostic information, physical examination, chief complaint, syndromes, treatment plan, and traditional Chinese medicine (TCM) prescription.

In general, there are few same syndromes in TCM diagnosis. Usually, a case corresponds to a syndrome, or a patient corresponds to a syndrome. The number of syndromes is far too many to classify. Therefore, we proposed the task of transforming the syndrome differentiation problem into identifying the nature and locations of the disease. This way, we can transform the TCM syndrome differentiation auxiliary decision-making task from the original multiclassification task into a multilabel classification task, where the disease's nature and location are regarded as labels. However, doctors give clinical prescriptions based on the diagnosed syndromes. As a result, extensive TCM clinical expertise and excellent TCM diagnostic understanding are required in the syndrome differentiation approach. In order to imitate the TCM dialectics process, we need to introduce the information implied by the symptoms and chief complaints in each medical record into the medical record text representation. So, we constructed a knowledge graph (KG) based on TCM features (TCMKG)

to excavate hidden information in the medical records to solve this problem.

The language model of a deep neural network has developed into an efficient text representation method in recent years. BERT [1] has shown promise in some natural language processing (NLP) tasks. We adopted BERT for EMRs multilabel classification and extended the model with knowledge representation with hidden information provided in the shared task, such as four diagnostic information, chief complaint, and physical examination. Experiments on a test set of EMRs demonstrated the efficacy of our method.

The following are the contributions of this paper: (1) To get around the problem of not being able to use EMRs for information-based dialectics, we turned TCM's standard syndrome differentiation work into a multilabel categorization assignment. (2) In this paper, we proposed syndrome differentiation assistant decision-making based on knowledge (KDSD) to integrate implicit knowledge graph embedding (KGE) from TCMKG into the syndrome differentiation assistant task. (3) We designed a method to adopt the fusion of medical text and knowledge graph entity representation (F-MT-KER) to alleviate the apparent impact of the personalization of the TCM medical records on decision-making.

2. Related Works

Machine learning often accomplishes multilabel classification tasks by changing the current algorithm or employing a binary classification technique. Unlike traditional machine learning, the deep learning method uses various neural network structures to extract the semantic embedding of input text. XML-CNN [2] used a one-dimensional convolutional neural network along the sequence length and word embedding dimension to represent the text input. Slice considers supervised pretrained dense embeddings from XML-CNN model as the input of its hierarchical linear model. Recently, based on XML-CNN, AttentionXML [3] used BiLSTM, added attention to labels to design scoring functions, and used a hierarchical tag tree to warm up the model.

Many studies have been committed to pretraining language representation models to obtain language information from a text and then use this information for specific NLP tasks. The Elmo [4], GPT [5], and BERT [1] models have achieved better results in multiple NLP tasks. After finetuning, these pretraining models can be applied to various tasks. However, due to the lack of information linkages between vertical applications and open domains, these models cannot be easily translated to specialized domains. The model can be pretrained in specific fields to solve this problem, but it is not desirable in terms of computational time and cost. For example, RoBERTa [6], XLNET [7], SpanBERT [8], and ERNIE [9] are trained in the abovementioned way. Recent research has demonstrated that in addition to the input, adding other relevant knowledge to the model can improve the model's performance to variable degrees, such as reading comprehension [10], text classification [11], natural language inference [12], knowledge acquisition [13],

and question answer [14]. Therefore, we argued that additional knowledge information could effectively benefit the existing pretraining model. Some studies have tried effectively controlling the external KGs through the joint representation learning of words and entities to achieve promising results [15–18]. Chen et al. [19] used BiLSTM to process text and introduced additional knowledge through specific attention mechanisms. Zhang et al. [20] tried to use external KG to enrich embedded information and improve language understanding.

In TCM data mining, Zhou [21] built a data model from multiple information entities and their interactions in outpatient data to fulfill large-scale clinical data integration and preprocessing activities, as well as to uncover useful clinical knowledge from the data. Liu et al. [22] approached a clinical data warehouse based on structured EMR data. Clinical terminology is collected to establish clinical hypotheses and aid in the discovery of clinical knowledge from large-scalereal-world TCM clinical data. The majority of TCM data mining involves constructing data into a database and then mining information from it to realize the application [23]. In recent years, the auxiliary decision-making systems based on expert systems [24] and the model-assisteddecision-making systems [25] based on deep learning have dominated TCM diagnosis. Joyce et al. [26] described the cognitive work of TCM tongue diagnosis and used it to find TCM diagnostic thinking patterns. An expert panel must assess the correctness of the diagnosis. Xie et al. [27] mined links between hidden items and inferred pathways from symptoms to syndromes using TCM knowledge graphs and reinforcement learning techniques. Yang et al. [28] utilized KGE to find triples connection pairings of syndromes and symptoms, and then created a score algorithm to find suggested syndromes. To apply clinically aided diagnostic and treatment models, Ruan et al. [29] stacks semantic-aware graph convolutions to learn efficient low-dimensional representations of nodes via metagraphs and self-attention, and predicts correct patterns via clustering and linking. Xie [30] employed multiclassifier integrated TCM syndrome element classification to help with diagnosis, mainly using the Nave Bayes, Weighted bipartite graph, SVM, and ProSVM algorithms. Since the number of syndromes in actual diagnosis is unclear, we split syndromes into labels of illness type and disease location, used multilabel classification to predict labels, and utilize a certain number of labels to replace ambiguity.

3. Methods

3.1. Overview. The KDSD model is divided into two components, as illustrated in Figure 1, the EMRs module and the KG module. We first develop the TCMKG using the EMRs and then represent all entities and relationships in KG through the KG embedding model. Input the EMRs into the model, and the EMRs module will extract the chief complaint and symptom information (four diagnostic information and physical examination). Meanwhile, the embedding representation of the knowledge entities in the
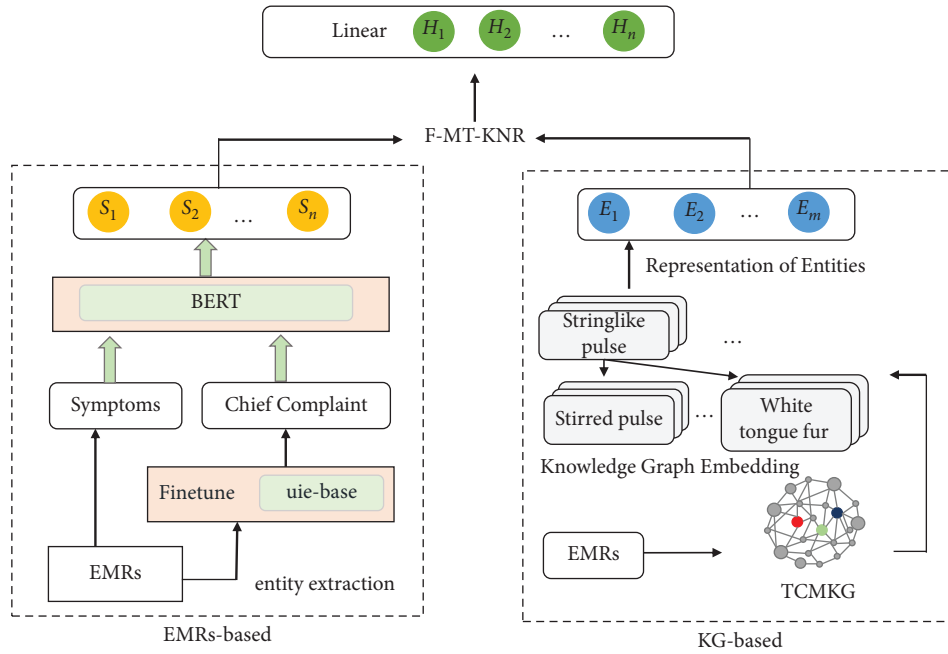
FIGURE 1: The architecture of the KDSD model.

KG module is obtained by extracting and linking the entities of the EMRs module. Finally, we fused text representations and KG entity embedding representations through the F-MT-KNR method designed in this paper to predict the nature and location of diseases through multilabel classification. The implementation details of the model are described below.

3.2. Construction of Knowledge Graph.

Different doctors use different syndrome differentiation procedures, which results in them prescribing different syndromes for the same illness and even achieving different syndrome outcomes for the same patient. In TCM, syndromes include the relationship between evil and positive, internal and external pathogenic factors, the location of the disease, and the nature of the disease. Therefore, it is objective to use the nature and location of the disease to assist doctors in making clinical decisions.

In the data mining phase, we extracted the TCM symptom entities and chief complaint information from the desensitized EMRs. We then decomposed the information on the nature of the illness and its location in the syndrome. We designed the schema layer top-down [31], which is actually to classify these entities. As a result, the schema layer's categories and the data layer's medical entities establish "class-example" connections. The data layer is filled with entities simultaneously, and the TCMKG is completed. Figure 2 shows the partial TCMKG.

After desensitization, entity extraction, and relationship extraction from 12533 EMRs, we created TCMKG. The TCMKG database has 115012 triples, which include 18078 symptom entities, 2098 chief complaint entities, and 157 nature and location of the disease entities. In addition, Neo4j is employed as a data storage tool for TCMKG.

3.3. Representation of Knowledge Graph.

In the EMRs of TCM, there is a very complex relationship between the four diagnostic information and syndrome or disease nature and location, which experts of TCM can only understand. By contrast, KG is a multirelational graph comprised of many different entities and their relationships [32]. In this paper, KG can accurately describe the relationship between symptoms and syndromes in TCM through information extraction, data mining, and knowledge reasoning. It can describe the evolution process and development law of knowledge. To provide accurate, traceable, interpretable, and inferential knowledge data for syndrome auxiliary decision-making. Besides, the representation learning of KG is the method of transforming these knowledge data with implicit information into a vector representation.

Given a knowledge graph K, which includes a collection of entities ε and relations $\mathscr{R}$, that is, K$\subseteq \varepsilon \times \mathscr{R} \times \varepsilon$. The triples are expressed as (h, r, t), where $h, r \in \varepsilon$ denote the head and tail entities, respectively, and $r \in \mathscr{R}$ denotes the relationship between them.

For each entity e and relationship r in K, the knowledge graph representation methods generate $e_v \in R^{d_e}$ and $r_v \in R^{d_r}$, where e_v and r_v are d_e and d_r dimensional vectors, respectively. Each embedding technique incorporates a scoring mechanism, $\mathscr{F}: \varepsilon \times \mathscr{R} \times \varepsilon \longrightarrow R$ will give some scores $\mathscr{F}(h, r, t)$ is assigned to a possible triple $(h, r, t), h, t \in \varepsilon,$, and $r \in \mathscr{R}$. The model is trained in such a way that for each correct triple $(h, r, t) \in$ K and wrong triple $(h', r', t') \notin$ K, the model assigns a score such that $\mathscr{F}(h, r, t) > 0$ and $\mathscr{F}(h', r', t') < 0$. Typically, a scoring function is a function of (e_h, e_r, e_t).

We selected several representative KG embedding methods to conduct experiments on TCMKG (see the experimental part) and evaluated the indicators, among which TransE performed best.
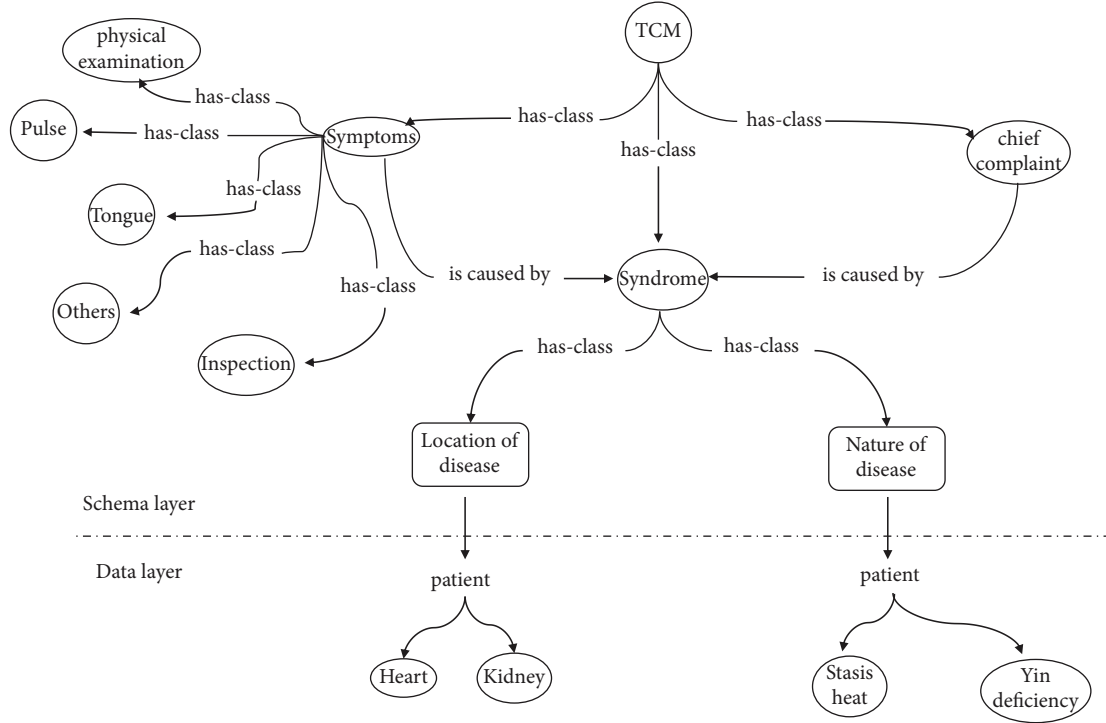
Figure 2: Ontology representation of TCMKG.

Bordes et al. [33] proposed TransE based on the assumption that the additional $h + t$ embedding should be near to that of t and that the scoring function should be defined as follows under L_1 or L_2 constraints:

$$f_r(h,t) = \|h + r - t\|_{L1/L2}. \tag{1}$$

As mentioned above, $h, t \in \varepsilon$, and $r \in \mathscr{R}$. After the KG embedding is completed, formula (1) makes $f(h, r, t) > 0$ for all true triples, and $f(h, r, t) < 0$ for all wrong triples.

3.4. EMR-Based Module. The BERT layer is used by the EMRs module to extract features from incoming text input. The embeddings of the KDSD model differ from those of the BERT model. We use outside information to complement the text's representation. We separate the information about the principal complaint from the information about the symptoms when we analyze the input data, and the format of the input sequence is specified by

$$[CLS]\text{symptoms}[UNK]\text{chief complaint}[UNK][SEP], \tag{2}$$

where [CLS] and [SEP] are the BERT model-specific special symbols, and [UNK] is the separator inside the text. Symptoms include pulse-taking, tongue examination, listening and smelling, inspection, and physical examination. Each symptom and chief complaint field is separated by [UNK] to facilitate the integration of subsequent KGE.

After entering the EMRs into the model, it is required to improve the text representation via the hidden layer. However, the symptom information and the chief complaint information are stated differently in EMRs, as is the length of the text. Since the maximum input length of BERT is

generally 512 characters, considering that this paper needs to separate symptoms and chief complaints, and the description of private complaints is too complicated, the length of some medical records exceeds the limit of the model. In this paper, most of the information on the four diagnostic information is relatively standardized. To decrease the text length, we apply the fine-tuned Uie-base [34] model to extract the main complaint information from the text. Then, as stated in the following equations, we employed the multi-headself-attention [35] process to integrate the principal complaint information into the text representation of symptom information.

$$Q = K = V = W^S \text{Concat}([C]; S_{1\ldots M}, C_{1\ldots N}), \tag{3}$$

$$[C'] = \text{Concat}(\text{head}_1, \ldots, \text{head}_i)W^O, \tag{4}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \tag{5}$$

where $[C]$ is the representation of [CLS] in the hidden layer state, $[C']$ is the representation of input text. $S_{1\ldots M}$ and $C_{1\ldots N}$ are symptom information embedding containing M values and chief complaint information embedding containing N values. W^s, W^O, W^Q, W^K, and W^V are trainable parameters.

Integration of additional medical knowledge.

The fusion information part aims to integrate the output of KGE into the output of BERT. BERT is to train the representation of each character in the unit of characters, while in the KGE module, we can only train the entities in KG into triples. Therefore, we proposed F-MT-KER to realize the fusion of characters and entity embeddings in KG, as shown in Figure 3.
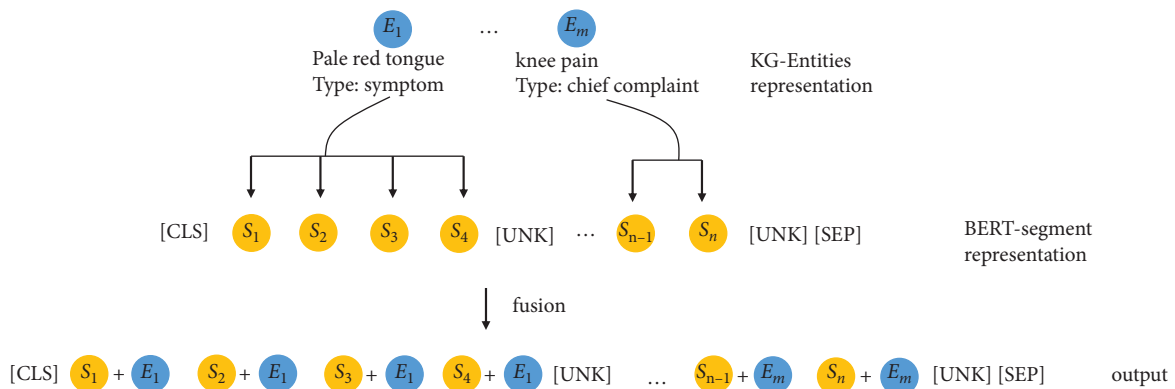
FIGURE 3: The process of fusion of knowledge graph entity information and character information of the BERT model.

$$S = \sum_{i=1}^{i=k}\left[\left(S_j, \ldots, S_{j+L(e_k)}\right) + e_k\right]. \qquad (6)$$

In equation (6), e_k is the embedding of an entity in KG, $S_j, \ldots, S_{j+L(e_k)}$ corresponds to symptom character set or chief complaint character set of e_k. L is the function of calculating the length of the entity and f is to select the corresponding e_k. The embedded representation of the same entity is added to the characters of each entity. Through the linear layer transformation, the fused vector dimension is changed from the dimension of the hidden layer to the dimension of the label quantity.

4. Results

4.1. Overview. The process of auxiliary decision-making for syndrome differentiation can be divided into the KG construction and embedding module, the text representation of the BERT model, and multilabel classification based on model fusion. For the medical records of the input model, we first take the medical record entity set through TCMKG and then link the entities in this entity set to the vector space embedded in TCMKG through entity matching. As a result, we can obtain vector representations of symptom entities and chief complaint entities. Ultimately, the nature and location of the disease recommended by the medical record can be obtained by inputting the embeddings containing hidden information into the model.

4.2. Dataset and Experimental Setup. We conducted experiments on the EMRs dataset and TCMKG. As seen in Figure 4, we removed sensitive information from the original EMRs and extract entities for pulse diagnosis, tongue diagnosis, and the other four diagnostic information based on the punctuation marks. After fine-tuning the medical record annotation data, the symptom information in the chief complaint is retrieved using the Uie-base model. For example, from the text information, the symptoms of "oral ulcer" and "unfavorable urination" may be extracted from "oral ulcer and unfavorable urination" that recur repeatedly for 2-3 years. It is worth mentioning that when the

probability of the anticipated symptom item is less than 0.5, the entity is deleted.

The dataset includes 12533 EMRs from QIHUANG TCM. 10026 EMRs were used for training and 2507 were used for testing. Furthermore, our criterion for dividing the test set is that only the test data labels that appear in the training set can be divided into the test set. This part of the data accounts for a tiny proportion, so we used the model to generate their input randomly. There are 157 classifications for disease nature and disease location. Every text in the input model is a medical case, and each text has an average of 102 characters. The relationship, number of edges, and triples in the knowledge graph constructed in this paper are shown in Table 1.

We constructed the knowledge ontology based on Yu [31] and defined entity and relationship description as the core. It includes some knowledge in EMRs, such as pulse diagnosis, tongue diagnosis, smell diagnosis, observation diagnosis, chief complaint, and other knowledge.

This paper preprocessed EMRs through data desensitization, cleaning, structuring, filtering, and diagnostic label standardization. In data filtering, duplicate information, and information that has little impact on diagnosis (such as the date in chief complaint) will be deleted. On the one hand, it can fulfill the BERT model input's format criteria. In TCMKG, on the other hand, it may also maintain concealed information. Our BERT model was bert-base-chinese, with the following primary settings: hidden size 768, maximum position embedding 512, number of epochs 50, number of attention heads 12, number of hidden layers 12, maximum input length 256, learning rate $2e - 5$, and batch size 8. Our NVIDIA T4 GPU is the backbone of all our experiments (17.18G).

4.3. Evaluation Metrics. Table 2 displays the experimental outcomes on the EMRs dataset. P@k (Precision at k) was used as evaluation metrics.

$$P@k = \frac{1}{k}\sum_{l=1}^{k} y_{\text{rank}(l)}. \qquad (7)$$
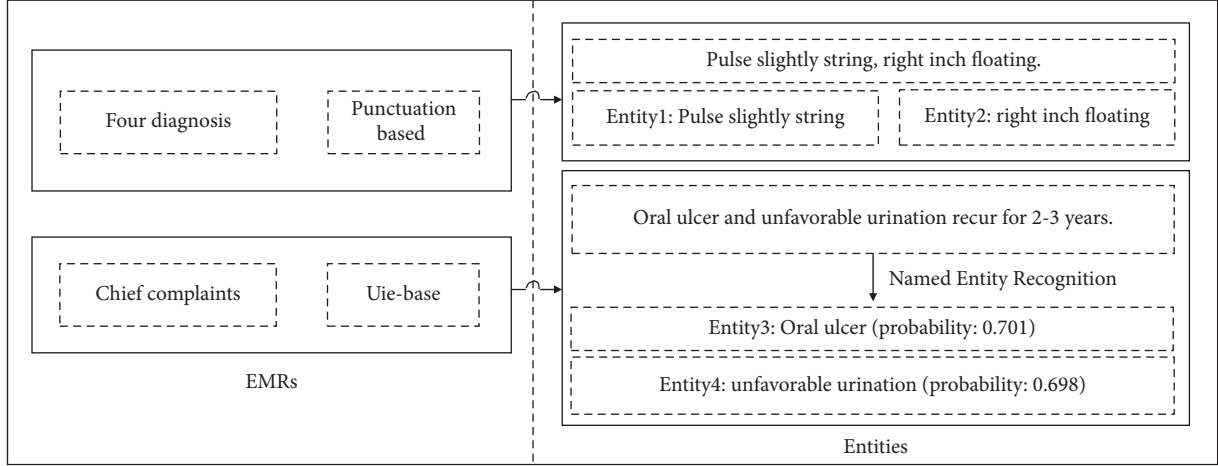
FIGURE 4: Entity extraction method for four diagnosis and chief complaint information.

The index of the lth highest prediction label is $\text{rank}(l)$ and y is the true binary vector.

4.3.1. Hamming Loss.

The Hamming loss is the proportion of wrongly predicted labels. The higher the performance, the lower the Hamming loss value. $\widehat{y}_i$ is the predicted value for the jth label of a given sample, y_i is the corresponding true value, and n_{labels} is the number of classes or labels.

$$L_{\text{Hamming}}(y, \widehat{y}) = \frac{1}{n_{\text{labels}}} \sum_{j=0}^{n_{\text{labels}}-1} 1\left(\widehat{y}_i \neq y_i\right). \tag{8}$$

4.3.2. Average Precision.

The weighted mean of precisions acquired at each threshold is used to summarize a precision-recall curve, with the increase in recall from the previous threshold used as the weight. Where P_n and R_n are the precision and recall at the nth threshold.

$$AP = \sum_n (R_n - R_{n-1})P_n. \tag{9}$$

4.3.3. Label Ranking Loss.

The label ranking loss function computes the ranking loss, which is weighted by the inverse of the number of ordered pairs of false and true labels.

$$\text{ranking}_{\text{loss}(y,\widehat{f})} = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{sample}-1}} \frac{1}{\|y_i\|_0 \left(n_{\text{labels}} - \|y_i\|_0\right)} |\left\{(k,l): \widehat{f}_{ik} < \widehat{f}_{il}, y_{ik} = 1, y_{il} = 0\right\}. \tag{10}$$

4.4. Performance on EMRs Dataset.

Liu et al. [36] used a data improvement and model fusion technique to accomplish syndrome differentiation classification of lung cancer diagnosis using the neural network text classification models TextCNN [37], TextRNN [38], RCNN [39], and fastText [40]. Hu et al. [41] employed the CNN and fastText models to complete the task of yin and yang deficient syndrome text categorization.

In addition, this study employs AttentiveConvNet [42], DPCNN [43], Transformer [35], and AttentionXML as comparative models for the multilabel text classification problem. A variety of context information is added to local convolution processes by AttentiveConvNet. Long-distance text dependencies may be extracted using the DPCNN. AttentionXML is a structural model for tag trees based on the attention mechanism. Using a multilabel attention approach for text multilabel categorization, the most essential feature information is collected for each label, allowing the relationship information between the input text and each label to be properly comprehended. Moreover, the KG-based algorithm used the KG's information to forecast

relationships that will deliver the desired results. This paper proposes the KDSD model. Bert-base-chinese is the pre-training model that is employed. By linking the linear layer after feature extraction using the model, a multilabel classification result is achieved.

4.5. The Results of Knowledge Graph Representation.

In order to demonstrate the KGE model's efficacy, we used the mainstream KGE methods [33, 44–46] to evaluate the TCMKG. For each KGE approach, we produce the associated knowledge representation and add it to the KDSD model. Table 3 illustrates the KDSD model's performance when various KGE models are employed as knowledge supplements.

We employ the MRR (Mean Radical Ranking), MR (Mean Rank), and Hits@N to analyze the performance metrics of KGE [47]. The MRR is the average of the triples' reciprocal rankings. The value ranges from 0 to 1; the greater the number, the better the model. MR is the average of all the triples' rankings. The value ranges from 1 (ideal condition,

TABLE 1: Number of relation categories and examples of triples in TCMKG.

Relationship categories	Number of edges	Triples	
		Chinese	English
Chief complaint-syndrome	16277	<失眠, 主诉_证候, 胆热痰扰>	< Insomnia, chief complaint-syndrome, gallbladder heat, and phlegm disturbance>
Symptom-syndrome	62308	<两寸略弱, 主诉_证候, 胆热痰扰>	< Two inches slightly weaker, symptom-syndrome, gallbladder heat, and phlegm disturbance>
Syndrome-nature of disease	19327	<胆热痰扰,证候_病性, 热>	<Gallbladder heat and phlegm disturbance, syndrome-nature of disease, and heat>
Syndrome-location of disease	17100	<胆热痰扰,证候_病位, 胆>	< Gallbladder heat and phlegm disturbance, syndrome-location of disease, and gallbladder>

TABLE 2: The results on the EMRs dataset.

Models	P@1	P@3	P@5	Average precision	Hamming loss	Label ranking loss
AttentiveConvNet	0.416	0.404	0.384	0.263	0.081	0.136
DPCNN	0.690	0.63	0.574	0.528	0.045	0.079
Transformer	0.721	0.665	0.61	0.578	0.04	0.086
AttentionXML	0.749	0.702	0.645	0.537	0.043	0.078
TextCNN	0.753	0.686	0.636	0.603	0.039	0.069
TextRNN	0.769	0.719	0.668	0.646	0.034	0.068
FastText	0.809	0.748	0.686	0.676	0.034	0.046
BERT	0.89	0.876	0.845	0.857	0.026	0.016
KG-based	0.059	0.073	0.066	—	—	—
KDSD	0.9	0.879	0.852	0.866	0.029	0.014

TABLE 3: Performance of different embedding models on KDSD.

Methods	P@1	P@3	P@5
BERT	0.89	0.876	0.845
+TransE	0.9	0.879	0.852
+DistMult	0.878	0.856	0.829
+ComplEx	0.868	0.852	0.825
+ConvKB	0.886	0.862	0.833

all rankings are equal) to the number of corruptions. Hits@n is the proportion of calculated ranks that are higher than (or equal to) a rank of n. The value ranges from 0 to 1; the greater the number, the better the model.

5. Discussion

5.1. Principal Results. As seen in Table 2, based on pre-training models, BERT and KDSD models outperform fastText, TextCNN, Text RNN, Transformer, and other models. KDSD increases by 1%, 0.3%, and 0.52% in the metrics P@1, P@3, and P@5, respectively, as compared to the BERT model.

From the experimental results, the results of the pre-training model are better than the previous text-based multilabel classification method. Furthermore, including KGE information in the pretraining model can improve the model's performance. However, all KG-Based indicators were terrible, far lower than the KDSD results. This might be due to one of the two factors. One is that KG-based method only used the information embedded in the graph and obtains the labels through link prediction. Another possibility is that TCMKG is a KG constructed for TCM diagnosis. Although the diagnosis of each medical record is different, the reuse rate of the information of the four diagnoses and the chief complaint is high. Besides, various clinicians define diseases differently, the number of triples for some conditions may be inadequate for good prediction.

From Table 3, it can be seen that the average precision of the KDSD model is 0.9% higher than that of the BERT model. This is due to the fact that in actual text task processing, some training corpora are difficult to obtain; their overall number and the total number of words contained are very small, making them unsuitable for training models with embedding layers. However, these data are suitable for training models without embedding layers. In addition, it provides useful rules that may be extracted by the model. Using a pretrained model to encode the original text is an excellent choice in this instance because the pre-trained model is derived from a large corpus and can make the current text meaningful, despite the fact that these implications may not be domain-specific. However, these deficiencies can be remedied by utilizing fine-tuned models.

5.2. Results of KGE. The KG-based experiment is another experiment undertaken for this investigation. If only the knowledge graph and its embedded data are used, the link prediction is performed on multiple triples associated with the same medical record, followed by the application of the KG's embedded scoring function. The expected labels received were evaluated, and the label with the highest score was finally selected. The results are bad due to the high rate of recurrence of the four diagnoses in TCM, the excessive number of triple diagnoses that may be derived from a medical file, and the inability to establish a connection between them. When this experiment is compared to the KDSD model and the BERT model, it is clear that using the information of the KG alone will not provide excellent results; instead, a deep learning model must be used in conjunction with it to get superior outcomes.

Although the KG-based method has no advantages in various indicators, the results of KDSD are indeed better than BERT in some indicators. It shows that the fusion with the embedded information from TCMKG can improve the performance of syndrome decision-making model. Moreover, as shown in Table 3, KDSD has the best performance after using the TransE model, surpassing BERT in all indicators. However, the results of some models have even declined after being fused with BERT. Here we evaluated the performance of KGE. Table 4 shows the performance evaluation findings for several embedding strategies. As seen in the table, TransE outperforms other models in MRR, MR, Hits@N, and has the greatest performance in the KDSD model for link prediction.

We analyze the reasons for the following reasons. First, complicated embedding approaches, such as adding complex vector space and convolution methods, will not function when the repeat rate of the four diagnosis information and the principal complaint information is high. Second,

TABLE 4: Performance evaluation of knowledge graph embedding.

Methods	MRR	MR	Hit@1	Hit@10	Hit@100
TransE	0.827	171	0.764	0.912	0.946
DistMult	0.203	492	0.13	0.357	0.633
ComplEx	0.201	530	0.128	0.35	0.625
HolE	0.208	1103	0.133	0.363	0.539
ConvKB	0.157	1011	0.072	0.361	0.551

since the triples of TCMKG construction adopt "symptom-nature\location" and "chief complaint-nature\location," we think the TransE model's distance-based scoring function calculates the Euclidean distance between them, which has more extraordinary expression ability, utilizing this intuitive distance-based technique.

5.3. Limitations. In this section, we discuss some limitations of the KDSD model. First, the new medical record cannot ensure that the input symptoms and chief complaints can match the entities in the TCMKG to ensure the existing label. Therefore, the embedded knowledge integrated into the model has little information, affecting the model's performance. Second, due to the particularity of TCM dialectics, although there is enough information on the syndrome, some syndromes' descriptions are still too simple. For example, if the syndrome is damp-heat obstructing the meridian (湿热阻滞经脉), then, we can get that the location of the disease is the meridians (经脉), and the nature of the disease is obstruction (阻滞) and damp-heat (湿热). This syndrome contains little information about the nature and location of the disease, so such medical records will have an adverse impact on the model's prediction of multiple labels when tested.

6. Conclusion

The auxiliary syndrome differentiation work has been viewed as a multiclassification issue. This paper presents the KDSD model for converting multiclassification into a multilabel classification. The KDSD model improves the model's performance by integrating the textual information of the EMR and the entity information embedded in the KG. We constructed TCMKG by using EMRs. Then, we represented TCMKG by TransE, and connected the results to the input layer of BERT. Finally, the results are obtained through the full connection output of the model. The results show that although BERT has been an excellent multilabel classification model in recent years, adding domain-specific knowledge graph information can improve the model's performance.

In the future, we will further study the construction of KG in TCM and explore the critical technologies of model fusion to improve the accuracy of the syndrome differentiation decision-making system. We find that some symptom entities in EMRs are not included in TCMKG (we only consider the existence of labels when building the dataset, some symptom entities will generate vectors randomly when predicting) to introduce other TCM knowledge that contains more symptom, and syndrome entities is an effective feature for syndrome differentiation.

Abbreviations

EMRs:	Electronic medical records
TCM:	Traditional Chinese medicine
KG:	Knowledge graph
KDSD:	Knowledge-based decision support model for syndrome differentiation
TCMKG:	Knowledge graph based on TCM features
F-MT-KER:	Fusing medical text with KG entity representation
KGE:	Knowledge graph embedding.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Qing Ye and Rui Yang contributed equally to this work.

Acknowledgments

References

[1] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the naacL-HLT*, pp. 4171–4186, Association for Computational Linguistics, Minneapolis, Minnesota, June 2019.

[2] J. Z. Liu, W. C. Chang, Y. X. Wu, and Y. M. Yang, "Deep learning for extreme multi-label text classification," in *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pp. 115–124, ACM, Shinjuku, Tokyo, Japan, August 2017.

[3] R. H. You, Z. H. Zhang, Z. Y. Wang, and S. Y Dai, "Attentionxml: label tree-basedattention-aware deep model for high-performance extreme multi-label text classification," in *Proceedings of the 33rd Conference on Neural Information*

Processing Systems (NeurIPS 2019), ACM, Vancouver, Canada, 2019.

[4] J. Sarzynska-Wawer, A. Wawer, A. Pawlak et al., "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Research*, vol. 304, no. 10, pp. 114135–135, 2021.

[5] T. Brown, B. Mann, N. Ryder, and M. Subbuah, "Language models are few-Shot Learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[6] Y. H Liu, M. Ott, N. Goyal, and J. F. Du, "Roberta: a robustly optimized bert pretraining approach," 2019, https://arxiv.org/abs/1907.11692.

[7] Z. Yang, Z. Dai, Y. Yang, and J. Carbonell, "Generalized autoregressive pretraining for language understanding," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, pp. 5753–5763, ACM, Vancouver, Canada, December 2019.

[8] M. Joshi, D. Q. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy, "Spanbert: improving pre-training by representing and predicting spans," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 64–77, 2020.

[9] Y. Sun, S. H. Wang, Y. K. Li, and S. K. Feng, "Ernie: enhanced representation through knowledge integration," 2019, https://arxiv.org/pdf/1904.09223.pdf.

[10] T. Mihaylov and A. Frank, "Knowledgeable reader: Enhancing cloze-style reading comprehension with external common-sense knowledge," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 821–832, Association for Computational Linguistics, Melbourne, Australia, July 2018.

[11] K. L. Zhang, X. Zhao, L. Zhuang, and Q. Xie, "Knowledge-enabled diagnosis assistant based on Obstetric EMRs and knowledge graph," in *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pp. 1155–1165, Chinese National Conference on Computational Linguistics, Haikou, China, October 2020.

[12] Q. Chen, X. D. Zhu, Z. H. Ling, and D. Inkpen, "Neural natural language inference models enhanced with external knowledge," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers)*, pp. 2406–2417, Association for Computational Linguistics, Melbourne, Australia, July 2017.

[13] X. Han, Z. Y. Liu, and M. S. Sun, "Neural knowledge acquisition via mutual attention between knowledge graph and text," in *Proceedings of the Thirty-second AAAI conference on artificial intelligence*, pp. 4832–4839, AAAI, April 2018.

[14] A. Saxena, A. Tripathi, and P. Talukdar, "Improving multi-hop question answering over knowledge graphs using knowledge base embeddings," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 4498–4507, Association for Computational Linguistics, July 2020.

[15] Z. Wang, J. W. Zhang, J. L. Feng, and Z. Chen, "Knowledge graph and text jointly embedding," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1591–1601, Association for Computational Linguistics, Doha, Qatar, October 2014.

[16] K. Toutanova, D. Q. Chen, P. Pantel, and H. Poon, "Representing text for joint embedding of text and knowledge bases," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1499–1509, Association for Computational Linguistics, Lisbon, Portugal, September 2015.

[17] X. Han, Z. Y. Liu, and M. S. Sun, "Joint representation learning of text and knowledge for knowledge graph completion," 2016, https://arxiv.org/pdf/1611.04125.pdf.

[18] Y. X. Cao, L. Hou, J. Z. Li, and Z. Y. Liu, "Joint representation learning of cross-lingual words and entities via attentive distant supervision," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 227–237, Brussels, Belgium, November 2018.

[19] J. D. Chen, Y. Z. Hu, J. P. Liu, and Y. H. Xiao, "Deep short text classification with knowledge powered attention," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, pp. 6252–6259, AAAI, January 2019.

[20] Z. Y. Zhang, X. Han, Z. Y. Liu, and X. Jiang, "ERNIE: enhanced language representation with informative entities," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1441–1451, Association for Computational Linguistics, Florence, Italy, July 2019.

[21] X. Z. Zhou, S. B. Chen, B. Y. Liu et al., "Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support," *Artificial Intelligence in Medicine*, vol. 48, no. 2-3, pp. 139–152, 2010.

[22] B. Y. Liu, X. Z. Zhou, Y. H. Wang et al., "Data processing and analysis in real-world traditional Chinese medicine clinical data: challenges and approaches," *Statistics in Medicine*, vol. 31, no. 7, pp. 653–660, 2012.

[23] S. S. Fang, L. Dong, L. Liu et al., "HERB: a high-throughputexperiment-andreference-guided database of traditional Chinese medicine," *Nucleic Acids Research*, vol. 49, no. D1, pp. 1197–1206, 2021.

[24] M. Casal-Guisande, A. Comesaña-Campos, A. Pereira, J. B. Bouza-Rodríguez, and J. Cerqueiro-Pequeño, "A decision-making methodology based on expert systems applied to machining tools condition Monitoring," *Mathematics*, vol. 10, no. 3, p. 520, 2022.

[25] Y. Xie, "A tcm question and answer system based on medical records knowledge graph," in *Proceedings of the 2020 International Conference on Computing and Data Science (CDS)*, pp. 373–376, IEEE, Stanford, CA, USA, August 2020.

[26] J. K. Anastasi, L. M. Currie, and G. H. Kim, "Understanding diagnostic reasoning in TCM practice: tongue diagnosis," *Alternative Therapies in Health & Medicine*, vol. 15, no. 3, pp. 18–28, 2009.

[27] Y. H. Xie, L. Y. Hu, X. X. Chen, J. Feng, and D. Zhang, "Auxiliary diagnosis based on the knowledge graph of tcm syndrome," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 481–494, 2020.

[28] R. Yang, Q. Ye, C. L. Cheng, S. H. Zhang, Y. Lan, and J. Zou, "Decision-making system for the diagnosis of syndrome based on traditional Chinese medicine knowledge graph," *Evidence-based Complementary and Alternative Medicine*, vol. 20229 pages, Article ID 8693937, 2022.

[29] C. Y. Ruan, Y. Wu, Y. Yang, and G. S. Luo, "Semantic-aware graph convolutional networks for clinical auxiliary diagnosis and treatment of traditional Chinese medicine," *IEEE Access*, vol. 9, no. 99, pp. 8797–8807, 2021.

[30] Y. H. Xie, Y. Y. Yan, J. Y. Li, and D. Zhang, "Research on assistant diagnostic method of TCM based on multi classifier integration," *Advances in Intelligent Systems Research*, vol. 153, pp. 371–376, 2017.

[31] T. Yu, J. H. 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.

[32] S. X. Ji, S. R. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: representation, acquisition, and

applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494–514, 2022.

[33] A. Bordes, N. Usunier, A. G. Duran, and J. Weston, "Translating embeddings for modeling multi-relational data," *Advances in Neural Information Processing Systems*, vol. 2, pp. 2787–2795, 2013.

[34] Y. J. Lu, Q. Liu, D. Dai, and X. Y. Xiao, "Unified structure generation for Universal information extraction," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 5755–5772, Association for Computational Linguistics, Dublin, Ireland, May 2022.

[35] A. Vaswani, N. Shazeer, N. Parmar, and J. Uszkoreit, "Attention is all you need," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, NIPS, Long Beach, CA, USA, 2017.

[36] Z. Q. Liu, H. Y. He, S. X. Yan, Y. Wang, T. Yang, and G. Z. Li, "End-to-End models to imitate traditional Chinese medicine syndrome differentiation in lung cancer diagnosis: model development and Validation," *JMIR Medical Informatics*, vol. 8, no. 6, Article ID e17821, 2020.

[37] Y. Kim, "Convolutional neural network for sentence classification," 2014, https://arxiv.org/pdf/1408.5882.pdf.

[38] P. F. Liu, X. P. Qiu, and X. J. Huang, "Recurrent neural network for text classification with multi-task learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2873–2879, ACM, New York, USA, July 2016.

[39] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, IEEE, Columbus, OH, USA, September 2014.

[40] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of Tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 427–431, Association for Computational Linguistics, Valencia, Spain, April 2017.

[41] Q. A. Hu, T. Yu, J. H. Li, Q. Yu, L. Zhu, and Y. Gu, "End-to-end syndrome differentiation of Yin deficiency and Yang deficiency in traditional Chinese medicine," *Computer Methods and Programs in Biomedicine*, vol. 174, pp. 9–15, 2019.

[42] W. P. Yin and H. Schütze, "Attentive convolution: Equipping cnns with rnn-style attention mechanisms," *Transactions of the Association for Computational Linguistics*, vol. 6, pp. 687–702, 2018.

[43] R. Johnson and T. Zhang, "Deep Pyramid convolutional neural networks for text categorization," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 562–570, Association for Computational Linguistics, Vancouver, Canada, July 2017.

[44] T. Trouillon, J. Welbl, S. Riedel, and E. Gaussier, "Complex embeddings for simple link prediction," in *Proceedings of the 33 rd International Conference on Machine Learning*, pp. 2071–2080, ACM, New York, NY, USA, June 2016.

[45] B. Yang, W. Yih, X. He, J. F. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," 2014, https://arxiv.org/abs/1412.6575.

[46] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung, "A novel embedding model for knowledge base completion based on convolutional neural network," in *Proceedings of the NAACL-HLT*, pp. 327–333, Association for Computational Linguistics, New Orleans, Louisiana, June 2017.

[47] Y. Q. Zhang, Q. M. Yao, T. X. Shao, and L. Chen, "NSCaching: simple and efficient negative sampling for knowledge graph embedding," in *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 614–625, IEEE, Macao, China, April 2019.

Research Article

Exploring the Molecular Mechanism of Tong Xie Yao Fang in Treating Ulcerative Colitis Using Network Pharmacology and Molecular Docking

Menglong Zou 🄳 and Ying Zhu 🄳

The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan 410007, China

Correspondence should be addressed to Ying Zhu; bjdf1992@126.com

Objective. The purpose of this study was to investigate the mechanisms of action of Tong Xie Yao Fang (TXYF) against ulcerative colitis (UC) by employing a network pharmacology approach. *Methods.* The network pharmacology approach, including screening of the active ingredients and targets, construction of the active ingredient-drug target network, the active ingredient-diseasetarget network, the protein–protein interaction (PPI) network, enrichment analyses, molecular docking, and targets validation, was used to explore the mechanisms of TXYF against UC. *Results.* 34 active ingredients and 129 and 772 targets of TXYF and UC, respectively, were identified. The intersection of the active ingredient-drug target network, the active ingredient-disease target network, and the PPI network suggested that kaempferol, beta-sitosterol, wogonin, and naringenin were the core ingredients and prostaglandin-endoperoxide synthase 2 (PTGS2) was the core target. Enrichment analyses showed that regulation of exogenous protein binding and other functions were of great significance. Nuclear factor-kappa B (NF-κB) signaling pathway, interleukin-17 (IL-17) signaling pathway, and tumor necrosis factor (TNF) signaling pathway were important pathways. Results of molecular docking indicated that the core ingredients and the target molecule had strong binding affinities. We have validated the high levels of expression of PTGS2 in UC by analyzing three additional datasets from the Gene Expression Omnibus (GEO) database. *Conclusions.* There are multiple ingredients, targets, and pathways involved in TXYF's effectiveness against UC, and these findings will promote further research and clinical applications.

1. Introduction

Ulcerative colitis (UC) is a chronic nonspecific inflammatory bowel disease with diffuse inflammation of both the colon and rectum mucosa [1]. The disease is also associated with abdominal pain, diarrhea, hematochezia, and systemic symptoms of varying severity [1]. Globally, there are estimated to be 8 million cases of UC, and rates are on the rise [2]. A hospital stay is required in approximately 20% of UC patients, and up to 10% are eventually required to undergo colectomy surgery [3, 4]. Additionally, people with UC have a higher incidence of colorectal cancer than people without [5]. Because of this, UC has a major impact on both mind and body. While medicinal treatments, such as 5-ASA and corticosteroids, are effective treatments for UC, their

side effects limit their use [6, 7]. It is therefore necessary to find other alternative treatment options for UC. The use of traditional Chinese medicine (TCM) has been around for thousands of years to treat refractory and common diseases [8]. Its validity, however, is often questioned since there is a lack of high-level evidence [9]. TCM has multiple ingredients and targets, and there is some evidence that some Chinese medicines improve the intestinal functional barrier in UC [10].

Tong Xie Yao Fang (TXYF) is a classic and famous Chinese medicine prescription [11]. Four herbs (Table 1) are included in TXYF, which were first described in a Yuan dynasty book named "Dan Brook Heart Law." It has been used for hundreds of years to treat diarrhea and abdominal pain associated with UC [12]. Previously, we reviewed the

TABLE 1: The formulation of Tong Xie Yao Fang (one dose).

Herb		Medicinal part	Amount in application (g)
Latin name	Chinese name		
Atractylodes macrocephala Koidz	Bai zhu	Rhizomes	12
Paeoniae Radix Alba	Bai shao	Root	12
Citrus Reticulata	Chen pi	Peel	9
Saposhnikoviae Radix	Fang feng	Root	6

mechanisms of action of these treatments, including anti-inflammatory and immunomodulatory effects, regulation of intestinal flora and the brain-gut axis, and promotion of mucosal healing [11]. The results of clinical studies also show that TXYF is effective and has few adverse effects in the treatment of UC [13]. There is, however, still much to learn about its molecular mechanisms.

In treating diseases, herbs contain many ingredients and target multiple targets, making it difficult to identify their pharmacodynamic ingredients and molecular mechanisms. With the advancement of network technology, the combination of network pharmacology, molecular docking, and bioinformation provides a powerful method for uncovering the complex molecular mechanisms of TCM [14]. They increase the likelihood of discovering the best drug candidates and significantly cut down on the initial time and expense of conducting experiments to determine drug-target interactions [15–17]. Thus, the goal of this study was to elucidate the potential mechanism by which TXYF against UC, using a network pharmacological strategy and molecular docking.

2. Materials and Methods

2.1. Screening the Active Ingredients of TXYF. "Baizhu," "Baishao," "Chenpi," and "Fangfeng" were used as search terms in the Traditional Chinese Medicine System Pharmacology Database (TCMSP, https://tcmspw.com/tcmsp.php) [18] and the Encyclopaedia of Traditional Chinese Medicine (ETCM, https://www.tcmip.cn/ETCM/) [19] to acquire related ingredients, respectively. The ingredients meeting both bioavailability (OB) ≥ 30% and drug-likeness (DL) ≥ 0.18 were screened out as the active ingredients [20].

2.2. Predicting the Targets of the Active Ingredients. We extracted the corresponding targets of the active ingredients of TXYF from the TCMSP database. To make the targets of TXYF more comprehensive, we retrieved the SymMap database (https://www.symmap.org/) [21] to add descriptions of target information. And then we converted the targets to gene symbols by using UniProt KB (https://www.uniprot.org/) [22]. Nonhuman genes were removed. Moreover, theingredients corresponding to nonhuman gene also were deleted.

2.3. Active Ingredient-Drug Target Network Construction. A network of active ingredient-drug target was created to visualize the complex pharmacodynamics relationships between herbs of TXYF by the employment of Cytoscape

v3.8.0. In Cytoscape, we used the plug-in cytoNCA to calculate the value of degree centrality (DC) and closeness centrality (CC) of all nodes in the network. As a standard for filtering ingredients and targets, we considered the nodes with values greater than the median value of DC and CC as the key nodes in our network [23]. Based on the new network, we can intuitively identify relationships between TXYF's active ingredients and its target targets.

2.4. Seeking Out Disease-Related Targets. From the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) [24], study number GSE65114 gene expression profile chip data have been gathered. The chip platform was GPL16686. This data set consists of mucosal biopsy specimens collected from 16 patients with UC and 12 healthy people. Identifying differential genes based on these criteria: $|\log FC| > 1.5$, $p < 0.05$.

2.5. Active Ingredient-Disease Target Network Construction. Targets from TXYF and UC were intersected to determine intersections. Cytoscape software was used to visualize the active ingredient-disease target network based on the active ingredients of TXYF and the intersection targets.

2.6. Protein–Protein Interaction (PPI) Network Construction. PPI network was built by importing the intersection targets into the STRING database (https://www.string-db.org/) [25], and the result was exported in tab-separated value (TSV) format. Subsequently, PPI data were visualized with Cytoscape software.

2.7. Enrichment Analysis. Using R software, we performed enrichment analysis on the target genes of TXYF in the treatment of UC based on Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG).

2.8. Molecular Docking. To determine what is the core ingredients of TXYF in the treatment of UC, we intersected steps 2.3 and 2.5. In the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) [26], we entered the core ingredients to obtain small molecule ligand structures. Likewise, we intersected steps 2.3 and 2.6 to identify the core targets of TXYF in the treatment of UC. The core target proteins were entered into the RCSB PDB database (https://www.rcsb.org/) [27] and the 3D structure was downloaded. Molecular docking was performed with AutoDockTools software after preparing the ligand files and the receptor file.

Table 2: Active ingredients of TXYF.

No.	Drug	Molecule ID	Ingredient name	OB	DL	Reference
1	Fangfeng	MOL011737	Divaricatacid	87.00	0.32	[32]
2	Fangfeng	MOL004793	Marmesine	84.77	0.18	[32]
3	Fangfeng	MOL000392	Formononetin	69.67	0.21	[32]
4	Fangfeng	MOL000011	(2R,3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-methylol-2,3-dihydropyrano[5,6-h][1,4]benzodioxin-9-one	68.83	0.66	[32]
5	Fangfeng	MOL005429	Hancinol	64.01	0.37	[33]
6	Fangfeng	MOL011732	Anomalin	59.65	0.66	[33]
7	Fangfeng	MOL004792	Nodakenin	57.12	0.69	[33]
8	Fangfeng	MOL001944	Marmesin	50.28	0.18	[33]
9	Fangfeng	MOL011730	11-hydroxy-sec-o-beta-d-glucosylhamaudol_qt	50.24	0.27	[32]
10	Fangfeng	MOL000417	Calycosin	47.75	0.24	[33]
11	Fangfeng	MOL002392	Deltoin	46.69	0.37	[32]
12	Fangfeng	MOL001942	Isoimperatorin	45.46	0.23	[34]
13	Fangfeng	MOL011749	Phelloptorin	43.39	0.28	[32]
14	Fangfeng	MOL001494	Mandenol	42.00	0.19	[34]
15	Fangfeng	MOL002644	Phellopterin	40.19	0.28	[34]
16	Fangfeng	MOL007514	Methyl icosa-11,14-dienoate	39.67	0.23	[34]
17	Fangfeng	MOL013077	Decursin	39.27	0.38	[34]
18	Fangfeng	MOL011753	5-O-methylvisamminol	37.99	0.25	[34]
19	Fangfeng	MOL000358	Beta-sitosterol	36.91	0.75	[32]
20	Fangfeng	MOL000359	Sitosterol	36.91	0.75	[32]
21	Fangfeng	MOL000358	Beta-sitosterol	36.91	0.75	[32]
22	Fangfeng	MOL003588	Prangenidin	36.31	0.22	[32]
23	Fangfeng	MOL001941	Ammidin	34.55	0.22	[32]
24	Fangfeng	MOL011747	Ledebouriellol	32.05	0.51	[32]
25	Fangfeng	MOL011740	Divaricatol	31.65	0.38	[34]
26	Fangfeng	MOL000173	Wogonin	30.68	0.23	[32]
27	Chenpi	MOL011737	Divaricatacid	87.00	0.32	[35]
28	Chenpi	MOL005815	Citromitin	86.90	0.51	[35]
29	Chenpi	MOL004793	Marmesine	84.77	0.18	[35]
30	Chenpi	MOL000392	Formononetin	69.67	0.21	[35]
31	Chenpi	MOL000011	(2R,3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-methylol-2,3-dihydropyrano[5,6-H][1,4]benzodioxin-9-one	68.83	0.66	[36]
32	Chenpi	MOL005429	Hancinol	64.01	0.37	[35]
33	Chenpi	MOL005828	Nobiletin	61.67	0.52	[35]
34	Chenpi	MOL011732	Anomalin	59.65	0.66	[37]
35	Chenpi	MOL004328	Naringenin	59.29	0.21	[35]
36	Chenpi	MOL004792	Nodakenin	57.12	0.69	[38]
37	Chenpi	MOL001924	Paeoniflorin	53.87	0.79	[38]
38	Chenpi	MOL001944	Marmesin	50.28	0.18	[37]
39	Chenpi	MOL011730	11-hydroxy-sec-O-beta-D-glucosylhamaudol_Qt	50.24	0.27	[36]
40	Chenpi	MOL000417	Calycosin	47.75	0.24	[38]
41	Chenpi	MOL005100	5,7-dihydroxy-2-(3-hydroxy-4-methoxyphenyl)chroman-4-one	47.74	0.27	[36]
42	Chenpi	MOL002392	Deltoin	46.69	0.37	[37]

TABLE 2: Continued.

No.	Drug	Molecule ID	Ingredient name	OB	DL	Reference
43	Chenpi	MOL001942	Isoimperatorin	45.46	0.23	[35]
44	Chenpi	MOL011749	Phelloptorin	43.39	0.28	[37]
45	Chenpi	MOL001494	Mandenol	42.00	0.19	[35]
46	Chenpi	MOL002644	Phellopterin	40.19	0.28	[37]
47	Chenpi	MOL007514	Methyl icosa-11,14-dienoate	39.67	0.23	[38]
48	Chenpi	MOL013077	Decursin	39.27	0.38	[38]
49	Chenpi	MOL011753	5-O-methylvisamminol	37.99	0.25	[37]
50	Chenpi	MOL000359	Sitosterol	36.91	0.75	[35]
51	Chenpi	MOL000358	Beta-sitosterol	36.91	0.75	[35]
52	Chenpi	MOL000359	Sitosterol	36.91	0.75	[35]
53	Chenpi	MOL003588	Prangenidin	36.31	0.22	[37]
54	Chenpi	MOL001941	Ammidin	34.55	0.22	[37]
55	Chenpi	MOL011747	Ledebouriellol	32.05	0.51	[37]
56	Chenpi	MOL011740	Divaricatol	31.65	0.38	[37]
57	Chenpi	MOL000173	Wogonin	30.68	0.23	[37]
58	Baizhu	MOL009431	Stemonine	81.75	0.72	[39]
59	Baizhu	MOL000392	Formononetin	69.67	0.21	[39]
60	Baizhu	MOL000022	14-acetyl-12-senecioyl-2E,8Z,10E-atractylentriol	63.37	0.3	[40]
61	Baizhu	MOL000020	12-senecioyl-2E,8E,10E-atractylentriol	62.40	0.22	[40]
62	Baizhu	MOL000021	14-acetyl-12-senecioyl-2E,8E,10E-atractylentriol	60.31	0.31	[40]
63	Baizhu	MOL005360	Malkangunin	57.71	0.63	[39]
64	Baizhu	MOL005384	Suchilactone	57.52	0.56	[40]
65	Baizhu	MOL000049	3β-acetoxyatractylone	54.07	0.22	[40]
66	Baizhu	MOL009154	Tuberostemoenone	53.90	0.73	[40]
67	Baizhu	MOL009387	Didehydrotuberostemonine	51.91	0.74	[40]
68	Baizhu	MOL000028	α-Amyrin	39.51	0.76	[40]
69	Baizhu	MOL009436	Stemotinine	38.69	0.46	[40]
70	Baizhu	MOL000358	Beta-sitosterol	36.91	0.75	[40]
71	Baizhu	MOL000033	(3S,8S,9S,10R,13R,14S,17R)-10,13-dimethyl-17-[(2R,5S)-5-propan-2-yloctan-2-yl]-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1h-cyclopenta[a]phenanthren-3-ol	36.23	0.78	[41]
72	Baizhu	MOL000072	8β-ethoxy atractylenolide III	35.95	0.21	[40]
73	Baizhu	MOL009361	13,15-dideoxyaconitine	34.67	0.25	[41]
74	Baishao	MOL001918	Paeoniflorgenone	87.59	0.37	[42]
75	Baishao	MOL003958	Evodiamine	86.02	0.64	[42]
76	Baishao	MOL004793	Marmesine	84.77	0.18	[42]
77	Baishao	MOL001925	paeoniflorin_qt	68.18	0.4	[42]
78	Baishao	MOL001928	albiflorin_qt	66.64	0.33	[42]
79	Baishao	MOL007016	Paeoniflorigenone	65.33	0.37	[42]
80	Baishao	MOL001910	11alpha,12alpha-epoxy-3beta-23-dihydroxy-30-norolean-20-en-28,12beta-olide	64.77	0.38	[42]
81	Baishao	MOL000785	Palmatine	64.60	0.65	[42]
82	Baishao	MOL005807	Sen-byakangelicol	58.00	0.61	[43]
83	Baishao	MOL004792	Nodakenin	57.12	0.69	[43]
84	Baishao	MOL000211	Mairin	55.38	0.78	[43]

TABLE 2: Continued.

No.	Drug	Molecule ID	Ingredient name	OB	DL	Reference
85	Baishao	MOL000492	(+)-catechin	54.83	0.24	[42]
86	Baishao	MOL001924	Paeoniflorin	53.87	0.79	[42]
87	Baishao	MOL001924	Paeoniflorin	53.87	0.79	[43]
88	Baishao	MOL001944	Marmesin	50.28	0.18	[43]
89	Baishao	MOL000096	(−)-Catechin	49.68	0.24	[42]
90	Baishao	MOL001921	Lactiflorin	49.12	0.8	[42]
91	Baishao	MOL002710	Pyrethrin Ii	48.36	0.35	
92	Baishao	MOL001942	Isoimperatorin	45.46	0.23	[42]
93	Baishao	MOL001919	(3S,5R,8R,9R,10S,14S)-3,17-dihydroxy-4,4,8,10,14-pentamethyl-2,3,5,6,7,9-hexahydro-1h-cyclopenta[a]phenanthrene-15,16-dione	43.56	0.53	[44]
94	Baishao	MOL000422	Kaempferol	41.88	0.24	[42]
95	Baishao	MOL002662	Rutaecarpine	40.30	0.6	[43]
96	Baishao	MOL002644	Phellopterin	40.19	0.28	[42]
97	Baishao	MOL002464	1-Monolinolein	37.18	0.3	[44]
98	Baishao	MOL007061	Methylenetanshinquinone	37.07	0.36	[44]
99	Baishao	MOL000359	Sitosterol	36.91	0.75	[42]
100	Baishao	MOL000358	Beta-sitosterol	36.91	0.75	[42]
101	Baishao	MOL000359	Sitosterol	36.91	0.75	[42]
102	Baishao	MOL000358	Beta-sitosterol	36.91	0.75	[42]
103	Baishao	MOL001454	Berberine	36.86	0.78	[42]
104	Baishao	MOL006346	Itraconazole	36.67	0.33	[43]
105	Baishao	MOL005789	Neobyakangelico L	36.18	0.31	[43]
106	Baishao	MOL001939	Alloisoimperatorin	34.80	0.22	[42]
107	Baishao	MOL001930	Benzoyl paeoniflorin	31.27	0.75	[42]
108	Baishao	MOL007025	Isobenzoylpaeoniflorin	31.14	0.54	[42]

Docking modules are considered more stable when their binding energy is smaller [28]. In terms of minimum binding energy, a value lower than -5.0 kcal/mol indicates a good affinity between receptor and ligand, and a value lower than -7.0 kcal/mol indicates a very strong affinity between receptor and ligand [29, 30]. To better evaluate the docking result, we also downloaded mesalazine-D3 (CID: 71750020) from PubChem database as ligand to perform molecular docking. Visualization of all docking results was performed with PyMOL v2.1.4 software [31].

2.9. Target Validation. Three additional data sets were retrieved from GEO for validation. This includes GSE36807 (7 healthy controls and 15 UC), GSE38713 (13 healthy controls and 22 UC), and GSE10616 (11 healthy controls and 10 UC). These three validation data sets were then used to verify the expression levels of the core targets.

3. Results and Analysis

3.1. Active Ingredients and Targets of TXYF. We obtained 108 active ingredients (Table 2) of TXYF in the TCMSP database and the ETCM database that satisfy both screening factors of $OB \geq 30\%$ and $DL \geq 0.18$ simultaneously. Among them, there are 35 ingredients for Baishao, 16 ingredients for Baizhu, 31 ingredients for Chenpi, and 26 ingredients for Fangfeng. A total of 34 active ingredients of TXYF were obtained after deleting 40 repeated ingredients and 34 ingredients corresponding to nonhuman gene. We identified 129 targets corresponding to 34 active ingredients in TXYF by searching TCMSP database and SymMap database. The Cytoscape software was used to draw a network diagram of the 34 active ingredients and the 129 drug targets. Next, we used the CytoNCA for further analysis of the topology of the network. The median value of DC and CC of the network was 2 and 0.320158103, respectively. According to the median value of BC and DC, 57 nodes were further screened (Figure 1).

3.2. Differential Gene in UC. GSE65114 data was analyzed using R software for screening differential genes for UC. 772 differential genes were screened according to screening criteria, 607 of which were upregulated and 165 were downregulated (Figure 2(a)). The first 50 differentially expressed genes of UC mucosal biopsy specimens and healthy mucosal biopsy specimens are selected, and a differentially expressed gene heatmap is drawn using the R software. Gene expression levels are represented by the color of the heatmap: blue represents decrease, red represents increase, and increasing and decreasing degrees of gene expression are represented by the brightness of the color, as shown in Figure 2(b).

3.3. Active Ingredient-Disease Target Network Construction. The 129 drug gene symbols and the 772 disease gene symbols were intersected to find 15 common targets. Using these 15 common targets, we constructed an active ingredient-

disease target network. The network shown in Figure 3 illustrates the complex relationship between TXYF and UC.

3.4. PPI Network Construction. STRING platform provided us with a map of the action relationships between 15 overlapping genes (Figure 4(a)). We exported the PPI network in "TSV format," and then analyzed its topological properties using Cytoscape software. The median value of DC and CC of the network is 16 and 0.30952381, respectively. We then selected all genes with values greater than the median value of DC and CC and generated 5 targets of more significant networks (Figure 4(b)).

3.5. GO Enrichment Analysis. 15 overlapping genes were categorized and enriched according to three modules by GO enrichment analysis ($p < 0.05$ and $q < 0.05$): biological process (BP), molecular function (MF), and cellular component (CC), with 751 GO terms enriched, of which the proportion of BP terms was relatively high, with 650 GO terms, mainly showing how some proteins interact with biological pathways and how they transport, such as BP involved in symbiotic interaction (GO: 0044403), entry into host (GO: 0044409), acute-phase response (GO: 0006953), and movement in host environment (GO: 0052126). 35 GO terms are present in CC, and these overlapping genes are closely related to those cell biofilms, like external side of plasma membrane (GO: 0009897), membrane raft (GO: 0045121), and membrane microdomain (GO: 0098857). The overlapping genes enriched 66 GO terms for MF, and MF analysis contributed to understanding which receptor activities the overlapping genes influence as well as how partial protein binding occurs, mainly virus receptor activity (GO: 0001618), exogenous protein binding (GO: 0140272), and protease binding (GO: 0002020). Figure 5 shows q value intercepts for the top ten terms from small to large for an abbreviated presentation of GO enrichment results.

3.6. KEGG Enrichment Analysis. R software was used to perform the KEGG enrichment analysis, with screening conditions set at $p < 0.05$ and $q < 0.05$. As illustrated in Figure 6, TXYF is mainly involved in treating UC through the lipid and atherosclerosis (hsa05417), the IL-17 signaling pathway (hsa04657), the NF-kappa B signaling pathway (hsa04064), the TNF signaling pathway (hsa04668), the AMPK signaling pathway (hsa04152), and the PPAR signaling pathway (hsa03320).

3.7. Molecular Docking. Based on the active ingredient-drug target network diagram (Figure 1), MOL000358, MOL000422, MOL000173, MOL004328, MOL005828, MOL011753, MOL000011, MOL000049, MOL013077, and MOL011740 may contribute to the pharmacodynamic material basis of TXYF. As well, PTGS2, DPP4, NCOA2, HSP90AB1, SCN5A, GABRA1, AR, ESR1, and NOS2 may be the ten most important targets of TXYF used to treat diseases. Figure 3 shows that MOL000173, MOL000422, MOL005828, MOL004328, MOL011740, MOL000358,
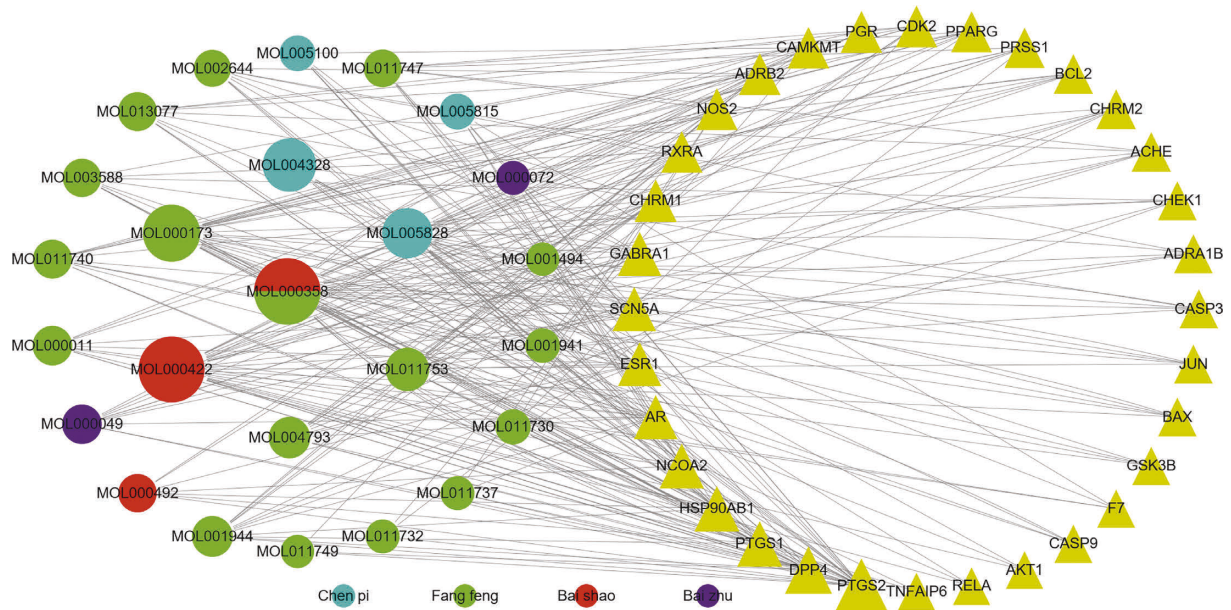
FIGURE 1: The network diagram of active ingredient-drug target. The circle nodes represent ingredients, and the targets are indicated by triangle nodes. Node size is proportional to its DC value.
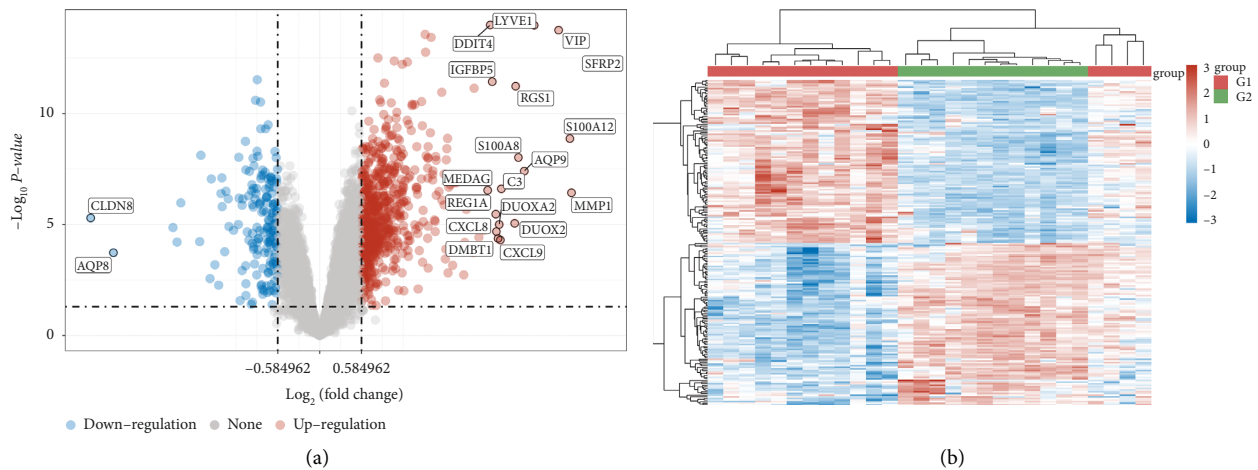

(a)

(b)

FIGURE 2: (a) Volcano map. (b) Heatmap. *Note.* G1, mucosal biopsy specimens of healthy people; G2, mucosal biopsy specimens from patients with UC; red, high gene expression; blue, low gene expression.

MOL011730, MOL000011, MOL000358, and MOL011747 may be ten of the important active ingredients of TXYF against UC. The PPI network diagram (Figure 4(b)) indicates that TXYF may target 5 highly important targets in the treatment of UC: ICAM1, PPARG, FN1, PTGS2, and CXCL8. Besides the interactions between drugs and diseases, we must also consider the effects of the drugs themselves. Hence, we obtained 1 core target (PTGS2) by taking the intersection of 10 important targets in the active ingredient-drug target network and 5 important targets in the PPI network. Likewise, we obtained 7 core ingredients (MOL011740, MOL000358, MOL000422, MOL000173, MOL004328, MOL005828, and MOL000011) by taking the intersection of 10 important ingredients in the active ingredient-drug target network and 10 important

ingredients in the active ingredient-diseases target network. Notably, the core target and core ingredients obtained here were screened according to their topological parameters. Based on the published literature and preliminary information in the TCMSP database, we investigated whether the core ingredients directly interact with the core target. All the core ingredients can directly affect the core target, which is interesting. Molecular docking is a methodology for predicting drug and protein binding modes and binding affinity based on receptor characteristics [45]. A ligand's conformational stability to the receptor is determined by its binding energy: lower binding energy means higher stabilization [29, 30]. After performing the analysis described above and predicting the results, we selected 7 core ingredients in TXYF to dock with PTGS2 to validate the results
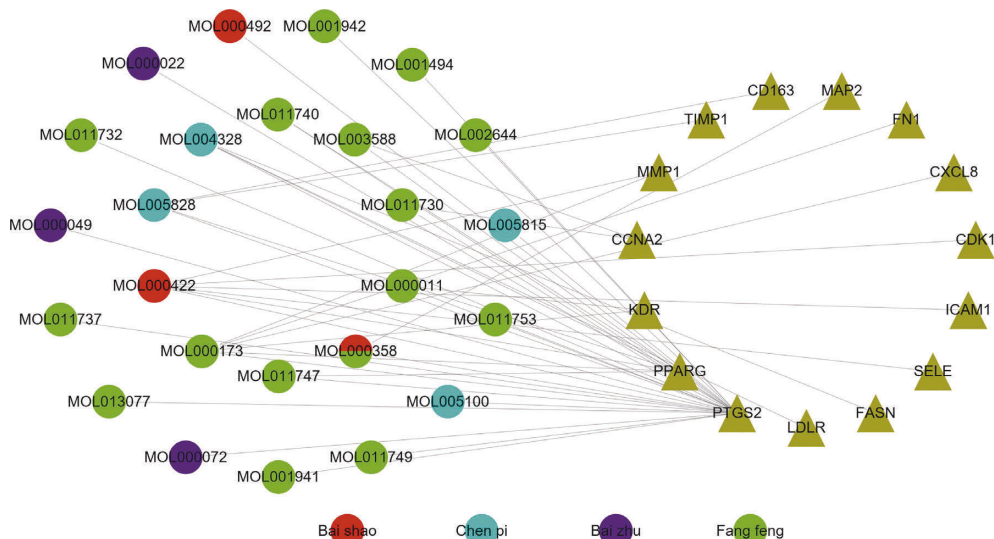
FIGURE 3: Active ingredient-disease target network. Circle nodes represent active ingredients in TXYF, triangle nodes represent overlapping gene symbols between disease and drug, with edges indicating that nodes can interact.
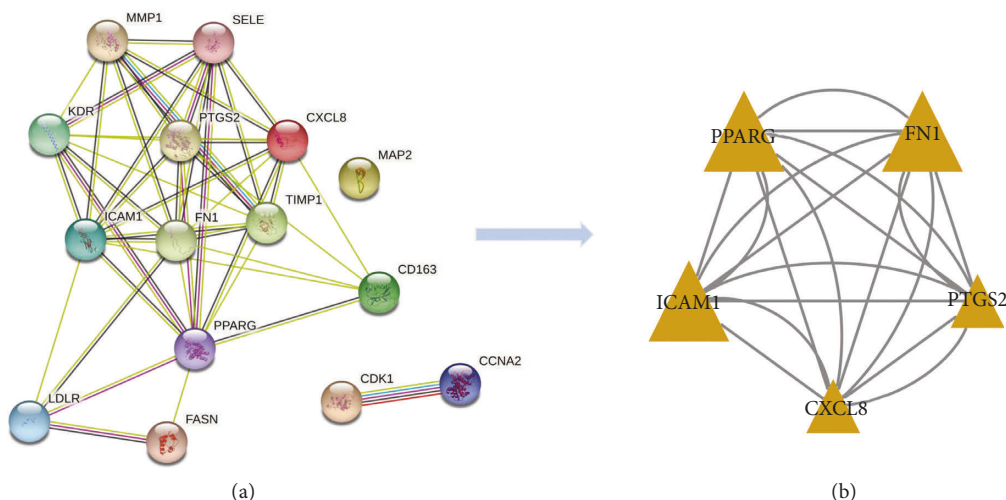

(a)

(b)

FIGURE 4: (a) The PPI network obtained from the STRING database platform. (b) The PPI network that extracted genes with value greater than the median value of DC and CC.

(Figure 7). As well, we downloaded mesalazine-D3's two-dimensional structure from the PubChem database for molecular docking to compare with TXYF's seven core ingredients (Figure 7). Based on the results, all the core ingredients exhibited strong binding affinity to the core target (Table 3). (2R, 3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-methylol-2,3-dihydropyrano[5,6-h][1,4]ben-zodioxin-9-one (MOL000011) and kaempferol (MOL000422) showed the smallest binding energy and considered as the most effective ingredients against PTGS2 protein. The largest binding energy was observed for mesalazine-D3 with −7.1 kcal/mol.

3.8. Validation of the Core Target. Three independent datasets (GSE36807, GSE38713, and GSE10616) were analyzed for PTGS2 expression data to validate the core target's expression. In the validation datasets, PTGS2 was similarly upregulated with statistical significance in UC (Figure 8). Therefore, PTGS2 may be a promising therapeutic target for TXYF treatment of UC.

4. Discussion

TCM has been used for thousands of years to treat various diseases. There are many classical prescriptions recorded in TCM literature that are still used in China today with clinical effectiveness. TXYF, one of classical TCM prescriptions, is composed of *Atractylodes macrocephala* Koidz (Baizhu), Paeoniae Radix Alba (Baishao), *Citrus Reticulata* (Chenpi), and *Saposhnikoviae* Radix (Fangfeng) [46]. Our study explored possible mechanisms of TXYF related to UC using network pharmacology and molecular docking, which
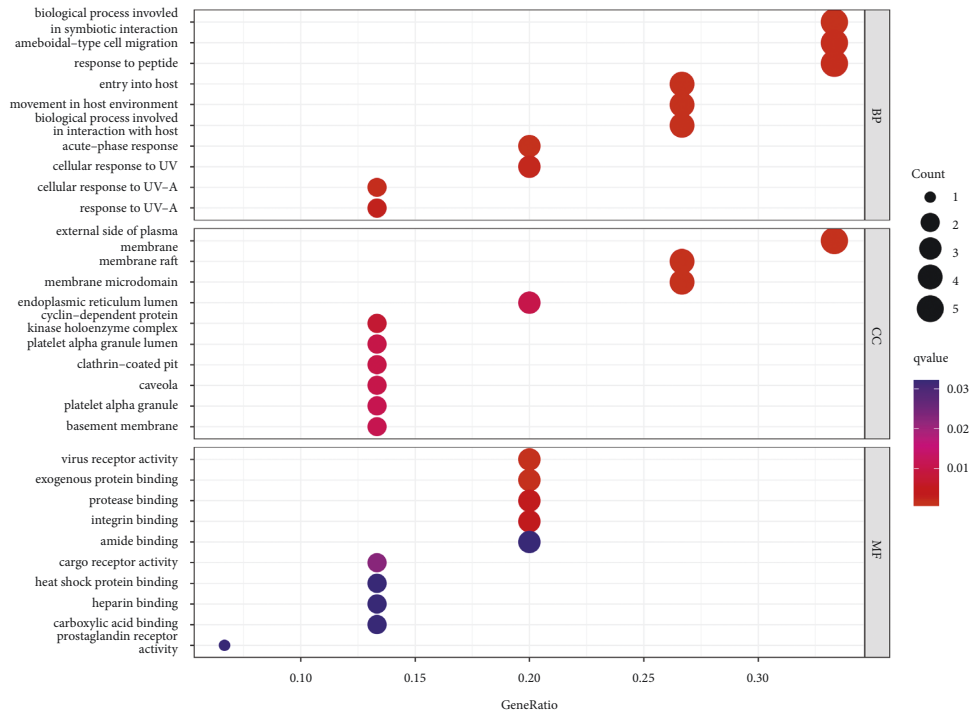
FIGURE 5: GO enrichment analysis of the overlapping targets.
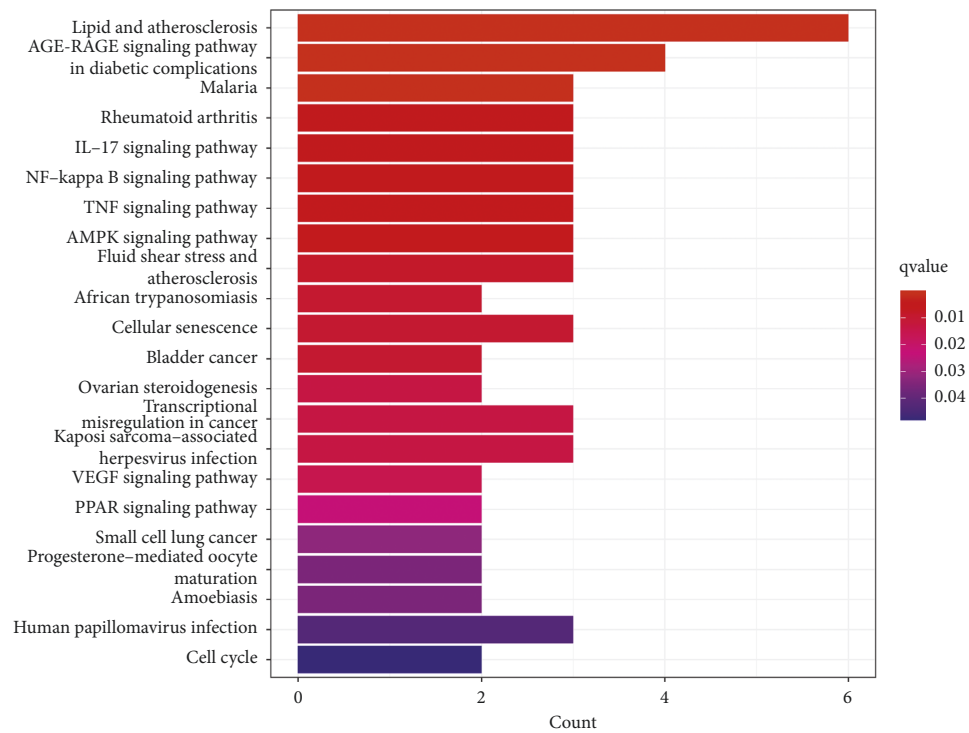
FIGURE 6: KEGG enrichment analysis of overlapping targets.

indicated that 7 ingredients and 15 target genes, especially PTGS2 (prostaglandin-endoperoxide synthase 2), were related to UC. According to the results of GO and KEGG pathway enrichment analyses, we found that the effects of

the TXYF against UC may be due to the core ingredients of the TXYF, especially for kaempferol (MOL000422) and (2R,3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-m ethylol-2,3-dihydropyrano[5,6-h][1,4]benzodioxin-9-one

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
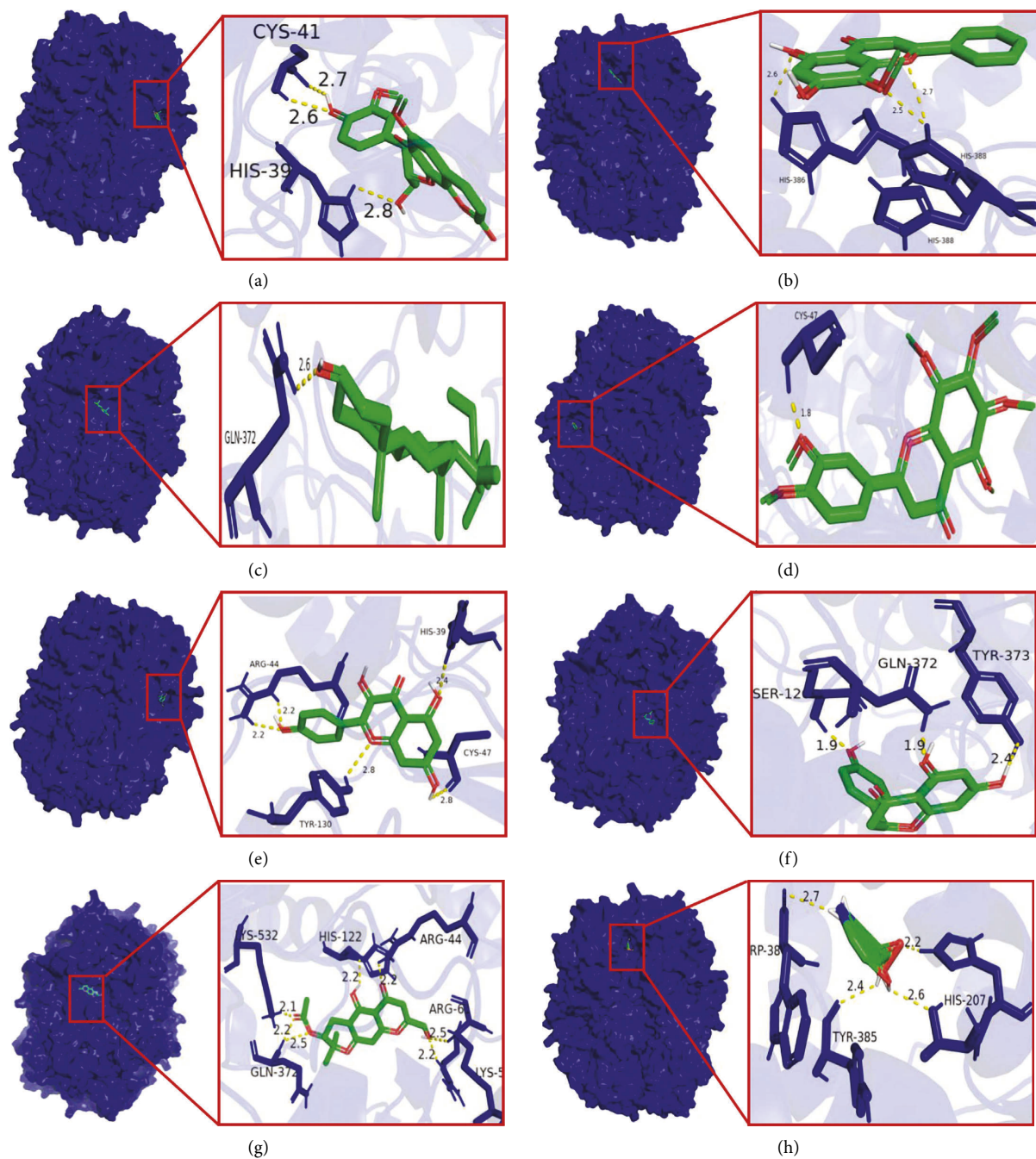
Figure 7: The core ingredients' interactions with PTGS2. (a) MOL000011 with PTGS2, (b) MOL000173 with PTGS2, (c) MOL000358 with PTGS2, (d) MOL005828 with PTGS2, (e) MOL000422 with PTGS2, (f) MOL004328 with PTGS2, (g) MOL011740 with PTGS2, and (h) mesalazine-D3 with PTGS2.

Table 3: Results of molecular docking.

Target	Binding energy (kcal/mol)							
	MOL000011	MOL000173	MOL000358	MOL005828	MOL000422	MOL004328	MOL011740	Mesalazine-D3
PTGS2	−9.0	−8.7	−8.5	−8.4	−9.0	−8.0	−0.84	−7.1
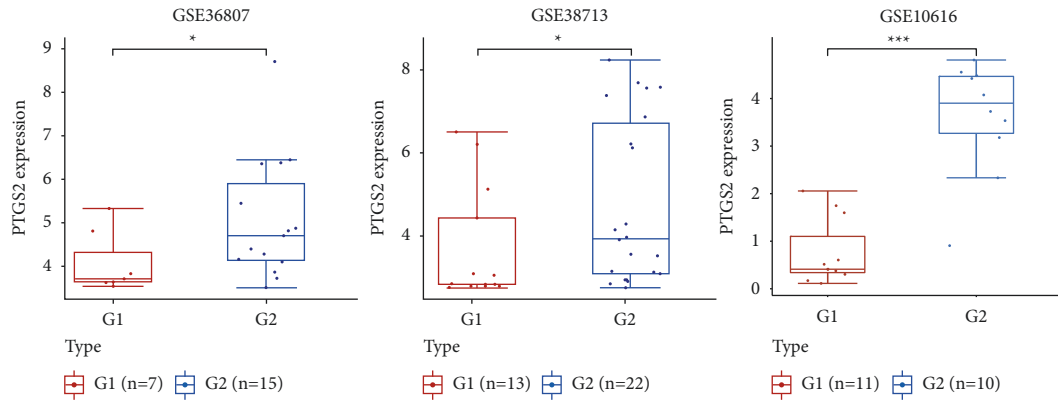
FIGURE 8: Validation of PTGS2 expression in UC patients. *Note.* Boxplot shows PTGS2 expression in mucosal biopsy specimens of patients with UC and healthy people in three GEO datasets (GSE36807, GSE38713, and GSE10616). G1, mucosal biopsy specimens of healthy people; G2, mucosal biopsy specimens from patients with UC. *Note.* $*<0.05$, $**<0.01$, $***<0.001$.

(MOL000011) which could influence the regulation of the NF-κB (nuclear factor-kappa B), IL-17 (interleukin-17), and TNF (tumor necrosis factor) signaling pathway. They may also be related to virus receptor activity, movement in host environment, exogenous protein binding, and protease binding. PTGS2 was selected as a core target based on the results of the PPI network analysis and the active ingredient-drug target network analysis, and it shows a greater affinity with the core ingredients than mesalazine-D3 based on molecular docking results.

The etiology and mechanisms of UC are complex [47]. Animal and clinical studies have confirmed the potential of herbal immunomodulators in UC [48–50]. There is evidence that mucosal immune cells produce inflammation factors that play an essential role in the pathogenesis of UC [51]. Active macrophages produce proinflammatory cytokines such as IL-17 and TNF-α, which can aggravate UC [52]. A balance between pro- and anti-inflammatory cytokines is essential to the control of UC [53]. In UC, anti-TNF induces mucosal healing, which reflects TNF's important role in pathogenesis [54]. A multitude of evidence points to the NF-κB pathway as having a vital role in the pathogenesis of UC [55, 56]. NF-κB overexpression in mucosal macrophages causes the production of proinflammatory cytokines, which harms mucosal tissues [57]. It has been shown that TXYF treatment reduces expression of NF-κB p65 gene and protein in rats, indicating that TXYF prevents overactivation of the NF-κB pathway [58]. Proinflammatory and anti-inflammatory factors are regulated by TXYF, which inhibits the expression of proinflammatory promoters and increases the expression of anti-inflammatory inhibitors [58]. In a study of 62 patients with UC, the control group was given sulfasalazine and the treatment group was given a combination of TXYF and sulfasalazine. One month after treatment, IL-17 and interferon-γ levels decreased statistically in the treatment group compared with the control group [59]. By soothing the liver and strengthening the spleen, TXYF may have an overall regulating effect, thereby repairing colon tissue, inhibiting the expression of TNF-α, and eliminating inflammation [60]. In our study, the NF-κB, IL-17 and the TNF signaling pathways are key pathways involved treatment of TXYF in UC.

PTGS2, also known as cyclooxygenase-2 (COX-2), encodes a prostaglandin synthase that catalyzes the synthesis of arachidonic acid derivatives (prostaglandins) [61]. In colonic epithelia of UC mice, mTOR complex 1 (mTORC1) was hyperactive [62]. It was found that colonic epithelial TSC1 (mTORC1 negative regulator) disruption increased the level of mTORC1 activity in the colon epithelia and aggravated UC [63]. Importantly, COX-2 inhibitor reversed elevated proinflammatory mediator levels caused by TSC1 deficiency, and subsequently mice models of UC were reduced in symptom severity and pathological characteristics [64]. When the prostaglandin E receptor 2 (PGE2) signaling pathway synergizes with TNF-α, it increases TNF-α-induced inflammatory responses, thereby escalating PG-mediated inflammation [65]. Reducing inflammation may be a key mechanism for treating UC, and anti-inflammatory agents may prove useful in preventing UC. A study conducted by Guo and Yan [66] revealed that TXYF can promote ulcer healing in rats with UC by downregulating the expression of COX-2 in the colonic mucosa, which is consistent with our results. It is widely known that LPS can trigger the release of inflammatory mediators like COX-2, IL-6, and IL-8. A study conducted by Wang et al. [67] showed wogonin inhibited the expression of COX-2 and restricted the translocation of NF-κB to the nucleus, thereby maintaining intestinal barrier integrity in LPS-inducedCaco-2 cells. β-sitosterol is a common sterol found in herbal medicines. Lee et al. [68] found that β-sitosterol improved the colon mucosa barrier by suppressing the expression of inflammatory factors TNF-α and COX-2, as well as the activation of NF-κB. The flavonoid kaempferol has been shown to have anti-inflammatory properties as well as to be immune-modulating [69]. Several structural changes were observed in colonic samples from DSS-induced mice, including mucosal ulcerations, crypt destruction, and loss of goblet cells [70]. On the other hand, histological analysis of the colons of mice treated with kaempferol revealed reduced levels of tissue damage [71]. In UC model, cytokines are controlled by many cells involved in inflammation, such as macrophages, neutrophils, and others. Myeloperoxidase (MPO) reflects neutrophil recruitment. When compared to other flavonols, Regasini et al.

[72] found that kaempferol derivatives significantly reduced MPO activity. Similarly, Kanashiro et al. [73] showed kaempferol inhibited human neutrophil degranulation significantly. The increased levels of COX-2-derived PGE2 are found at sites of intestinal inflammation and correlate with disease activity, exactly as described previously. It was reported by Park et al. [74] that consumption of kaempferol-supplemented diets are able to significantly lower colonic COX-2 mRNA expression. Likewise, nobiletin inhibited inflammation by downregulating the expression of COX-2 and inducible nitric oxide synthase (iNOS) in colitic rats [75]. In our study, we preliminarily explored the mechanism of the TXYF in treating UC by using network pharmacology and molecular docking. NF-κB, IL-17, and TNF signaling pathways are closely regulated by kaempferol, (2R,3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-methylol-2,3-dihydropyrano[5,6-h][1,4] benzodioxin-9-one, nobiletin, and divaricatol, and so on, according to our findings. PTGS2 was found to be the core target of the TXYF in treating UC and it reflects a strong affinity with the core ingredients, such as aempferol, and so on.

Nonetheless, some limitations were present in our study. In the evaluations, the ingredients and targets were derived mostly from databases. However, some ingredients and targets may have been omitted. Moreover, dose of herbs was not taken into consideration. Further experimentation is therefore needed to verify TXYF's multiple mechanisms of action for treating UC, both in vivo and in vitro.

5. Conclusion

We obtained the core active ingredients of TXYF in the treatment of UC, namely, kaempferol, (2R,3R)-3-(4-hydroxy-3-methoxy-phenyl)-5-methoxy-2-methylol-2,3-dihydropyrano[5,6-h][1,4]benzodioxin-9-one, nobiletin, divaricatol, beta-sitosterol, wogonin, and naringenin. By constructing the active ingredient-drug target network and PPI network, 1 core gene was screened, namely, PTGS2. GO functional enrichment analysis showed that the cross gene mainly performed exogenous protein binding and other functions, and the results of KEGG pathway enrichment analysis showed that the cross gene was mainly involved in NF-κB, IL-17, TNF, and other signaling pathways. Finally, molecular docking is vital to testing and exploring the binding between active ingredients and targets.

Data Availability

Specific study data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] I. Ordás, L. Eckmann, M. Talamini, D. C. Baumgart, and W. J. Sandborn, "Ulcerative colitis," *The Lancet*, vol. 380, no. 9853, pp. 1606–1619, 2012.

[2] F. Carbonnel, J. F. Colombel, J. Filippi et al., "Methotrexate is not superior to placebo for inducing steroid-free remission, but induces steroid-free clinical remission in a larger proportion of patients with ulcerative colitis," *Gastroenterology*, vol. 150, no. 2, pp. 380–388.e4, 2016.

[3] P. Hindryckx, V. Jairath, and G. D'Haens, "Acute severe ulcerative colitis: from pathophysiology to clinical management," *Nature Reviews Gastroenterology & Hepatology*, vol. 13, no. 11, pp. 654–664, 2016.

[4] F. Magro and M. M. Estevinho, "Is tofacitinib a game-changing drug for ulcerative colitis?" *United European Gastroenterology Journal*, vol. 8, no. 7, pp. 755–763, 2020.

[5] O. Olén, R. Erichsen, M. C. Sachs et al., "Colorectal cancer in ulcerative colitis: a scandinavian population-based cohort study," *The Lancet*, vol. 395, no. 10218, pp. 123–131, 2020.

[6] P. Sehgal, J. F. Colombel, A. Aboubakr, and N. Narula, "Systematic review: safety of mesalazine in ulcerative colitis," *Alimentary Pharmacology & Therapeutics*, vol. 47, no. 12, pp. 1597–1609, 2018.

[7] J. Langhorst, H. Wulfert, R. Lauche et al., "Systematic review of complementary and alternative medicine treatments in inflammatory bowel diseases," *Journal of Crohn's and Colitis*, vol. 9, no. 1, pp. 86–106, 2015.

[8] C. Q. Zhao, Y. Zhou, J. Ping, and L. M. Xu, "Traditional Chinese medicine for treatment of liver diseases: progress, challenges and opportunities," *Journal of Integrative Medicine*, vol. 12, no. 5, pp. 401–408, 2014.

[9] R. Teschke, A. Wolff, C. Frenzel, A. Eickhoff, and J. Schulze, "Herbal traditional Chinese medicine and its evidence base in gastrointestinal disorders," *World Journal of Gastroenterology*, vol. 21, no. 15, pp. 4466–4490, 2015.

[10] C. Wang, X. Tang, and L. Zhang, "Huangqin-tang and ingredients in modulating the pathogenesis of ulcerative colitis," *Evidence Based Complement Alternative Medicine*, vol. 2017, Article ID 7016468, 7 pages, 2017.

[11] K. Chen, Y. Lou, and Y. Zhu, "Tong Xie Yao Fang: a classic Chinese medicine prescription with potential for the treatment of ulcerative colitis," *Evidence-Based Complementary and Alternative Medicine*, vol. 2021, Article ID 5548764, 12 pages, 2021.

[12] T. Zhang, M. Shi, C. M. Liao, and F. L. Liu, "Research progress on mechanism of tongxieyao formula in the treatment of ulcerative colitis," *World Chinese Medicine*, vol. 10, no. 16, pp. 1643–1648, 2021.

[13] Y. J. Li, N. Deng, X. Y. Lin, Z. Fang, J. M. Liu, and C. F. Liu, "Meta-analysis of tongxie yaofang in the treatment of ulcerative colitis," *Journal of Emergency in Traditional Chinese Medicine*, vol. 31, no. 1, pp. 7–10.

[14] M. Lee, H. Shin, M. Park, A. Kim, S. Cha, and H. Lee, "Systems pharmacology approaches in herbal medicine research: a brief review," *BMB Report*, vol. 5666, 2022.

[15] Q. Huang, J. Lin, S. Huang, and J. Shen, "Impact of qi-invigorating traditional Chinese medicines on diffuse large B cell lymphoma based on network pharmacology and experimental validation," *Frontiers in Pharmacology*, vol. 12, Article ID 787816, 2021.

[16] M. Liu, G. Luo, L. Dong et al., "Network pharmacology and in vitro experimental verification reveal the mechanism of the

hirudin in suppressing myocardial hypertrophy," *Frontiers in Pharmacology*, vol. 13, Article ID 914518, 2022.

[17] R. Sun, G. Xu, D. Gao, Q. Ding, and Y. Shi, "To predict anti-inflammatory and immunomodulatory targets of guizhi decoction in treating asthma based on network pharmacology, molecular docking, and experimental validation," *Evidence-Based Complementary and Alternative Medicine*, vol. 2021, Article ID 9033842, 17 pages, 2021.

[18] J. Ru, P. Li, J. Wang et al., "TCMSP: a database of systems pharmacology for drug discovery from herbal medicines," *Journal of Cheminformatics*, vol. 6, no. 1, p. 13, 2014.

[19] H. Y. Xu, Y. Q. Zhang, Z. M. Liu et al., "ETCM: an encyclopaedia of traditional Chinese medicine," *Nucleic Acids Research*, vol. 47, no. D1, pp. D976–d982, 2019.

[20] N. Li, J. Sun, J. L. Chen, X. Bai, and T. H. Wang, "Gene network mechanism of zhilong huoxue tongyu capsule in treating cerebral ischemia-reperfusion," *Frontiers in Pharmacology*, vol. 13, Article ID 912392, 2022.

[21] 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, no. D1, Article ID D1110, 2019.

[22] Y. Wang, Q. Wang, H. Huang et al., "A crowdsourcing open platform for literature curation in UniProt," *PLoS Biology*, vol. 19, no. 12, Article ID e3001464, 2021.

[23] S. Zhou, N. Jiang, M. Zhang et al., "Analyzing active constituents and optimal steaming conditions related to the hematopoietic effect of steamed panax notoginseng by network pharmacology coupled with response surface methodology," *BioMed Research International*, vol. 2020, Article ID 9371426, 2020.

[24] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets-update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, 2013.

[25] D. Szklarczyk, A. L. Gable, K. C. Nastou et al., "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Research*, vol. 49, no. D1, pp. D605–d612, 2021.

[26] S. Kim, J. Chen, T. Cheng et al., "PubChem in 2021: new data content and improved web interfaces," *Nucleic Acids Research*, vol. 49, no. D1, Article ID D1388, 2021.

[27] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[28] K. M. Elokely and R. J. Doerksen, "Docking challenge: protein sampling and molecular docking performance," *Journal of Chemical Information and Modeling*, vol. 53, no. 8, pp. 1934–1945, 2013.

[29] C. Feng, M. Zhao, L. Jiang, Z. Hu, and X. Fan, "Mechanism of modified danggui sini decoction for knee osteoarthritis based on network pharmacology and molecular docking," *Evidence-based Complementary and Alternative Medicine*, vol. 2021, Article ID 6680637, 11 pages, 2021.

[30] Y. Lin, C. Shen, F. Wang, Z. Fang, and G. Shen, "Network pharmacology and molecular docking study on the potential mechanism of yi-qi-huo-xue-tong-Luo formula in treating diabetic peripheral neuropathy," *Journal of Diabetes Research*, vol. 2021, Article ID 9941791, 16 pages, 2021.

[31] Y. Xie, K. Zhou, Y. Wang et al., "Revealing the mechanism of astragali radix against cancer-related fatigue by network pharmacology and molecular docking," *Evidence-Based Complementary and Alternative Medicine*, vol. 2021, Article ID 7075920, 10 pages, 2021.

[32] J. Kang, J. H. Sun, L. Zhou et al., "Characterization of compounds from the roots of *Saposhnikovia divaricata* by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 22, no. 12, pp. 1899–1911, 2008.

[33] L. Chen, X. Chen, L. Su, Y. Jiang, and B. Liu, "Rapid characterisation and identification of compounds in Saposhnikoviae radix by high-performance liquid chromatography coupled with electrospray ionisation quadrupole time-of-flight mass spectrometry," *Natural Product Research*, vol. 32, no. 8, pp. 898–901, 2018.

[34] H. X. Zhong, A. L. Xu, and G. L. Xiao, "Identification of characteristic constituents in *Saposhnikovia* divaricata and its adulterant libanotis laticalycina by UPLC-Q-TOF-MS/MS and TLC," *Traditional Chinese Drug Research & Clinical Pharmacology*, vol. 33, no. 2, pp. 242–248.

[35] Y. Y. Zheng, X. Zeng, W. Peng, Z. Wu, and W. W. Su, "Characterisation and classification of citri reticulatae pericarpium varieties based on UHPLC-Q-TOF-MS/MS combined with multivariate statistical analyses," *Phytochemical Analysis*, vol. 30, no. 3, pp. 278–291, 2019.

[36] X. Lin, S. Cao, J. Sun, D. Lu, B. Zhong, and J. Chun, "The chemical compositions, and antibacterial and antioxidant activities of four types of Citrus essential oils," *Molecules*, vol. 26, pp. 3412–11, 2021.

[37] Y. Wang, J. Qian, J. Cao et al., "Antioxidant capacity, anticancer ability and flavonoids composition of 35 Citrus (*Citrus reticulata* blanco) varieties," *Molecules*, vol. 22, pp. 1114–7, 2017.

[38] L. Duan, L. Guo, L. L. Dou et al., "Discrimination of *Citrus reticulata* Blanco and *Citrus reticulata* "Chachi" by gas chromatograph-mass spectrometry based metabolomics approach," *Food Chemistry*, vol. 212, pp. 123–127, 2016.

[39] C. Zhai, J. Zhao, A. G. Chittiboyina, Y. Meng, M. Wang, and I. A. Khan, "Newly generated atractylon derivatives in processed rhizomes of *Atractylodes macrocephala* Koidz," *Molecules*, vol. 25, pp. 5904–24, 2020.

[40] Y. X. Wu, W. W. Lu, Y. C. Geng et al., "Antioxidant, antimicrobial and anti-inflammatory activities of essential oil derived from the wild rhizome of *Atractylodes macrocephala*," *Chemistry and Biodiversity*, vol. 17, no. 8, Article ID e2000268, 2020.

[41] H. Jiang, J. Shi, and Y. Li, "Screening for compounds with aromatase inhibiting activities from *Atractylodes macrocephala* Koidz," *Molecules*, vol. 16, no. 4, pp. 3146–3151, 2011.

[42] Y. Wu, Y. Jiang, L. Zhang et al., "Chemical profiling and antioxidant evaluation of *Paeonia lactiflora* pall. "Zhongjiang" by HPLC-ESI-MS combined with DPPH assay," *Journal of Chromatographic Science*, vol. 59, no. 9, pp. 795–805, 2021.

[43] X. Sun, L. Chen, H. Yan et al., "An efficient high-speedcounter-current chromatography method for the preparative separation of potential antioxidants from *Paeonia lactiflora* Pall. combination of in vitro evaluation and molecular docking," *Journal of Separation Science*, vol. 45, no. 11, pp. 1856–1865, 2022.

[44] N. N. Tong, X. Y. Zhou, L. P. Peng, Z. A. Liu, and Q. Y. Shu, "A comprehensive study of three species of Paeonia stem and leaf phytochemicals, and their antioxidant activities," *Journal of Ethnopharmacology*, vol. 273, Article ID 113985, 2021.

[45] Y. Wufuer, X. Yang, L. Guo et al., "The antitumor effect and mechanism of total flavonoids from *Coreopsis tinctoria* nutt (snow Chrysanthemum) on lung cancer using network

pharmacology and molecular docking," *Frontiers in Pharmacology*, vol. 13, Article ID 761785, 2022.

[46] Y. Wang, Y. Q. Huang, S. L. Zhu et al., "Efficacy of Tong-Xie-Yao-Fang granule and its impact on whole transcriptome profiling in diarrhea-predominant irritable bowel syndrome patients: study protocol for a randomized controlled trial," *Trials*, vol. 21, no. 1, p. 908, 2020.

[47] L. R. Glick, A. S. Cifu, and L. Feld, "Ulcerative colitis in adults," *JAMA*, vol. 324, no. 12, pp. 1205-1206, 2020.

[48] L. Huang, Z. Cai, Y. Zhu, and H. Wan, "Treatment of ulcerative colitis with spleen and kidney yang deficiency by kuijiening plaster: a randomized controlled study," *Zhongguo Zhen Jiu*, vol. 33, no. 7, pp. 577–581, 2013.

[49] J. Wang, X. Wang, X. Ma et al., "Therapeutic effect of Patrinia villosa on TNBS-induced ulcerative colitis via metabolism, vitamin D receptor and NF-κB signaling pathways," *Journal of Ethnopharmacology*, vol. 288, Article ID 114989, 2022.

[50] L. Ju, F. Ke, and P. Yadav, "Herbal medicine in the treatment of ulcerative colitis," *Saudi Journal of Gastroenterology*, vol. 18, no. 1, pp. 3–10, 2012.

[51] Y. Zhang, A. Desai, S. Y. Yang et al., "Tissue regeneration. Inhibition of the prostaglandin-degrading enzyme 15-PGDH potentiates tissue regeneration," *Science*, vol. 348, no. 6240, Article ID aaa2340, 2015.

[52] Z. Wang, W. Shi, D. Tian et al., "Autotaxin stimulates LPA2 receptor in macrophages and exacerbates dextran sulfate sodium-induced acute colitis," *Journal of Molecular Medicine (Berlin)*, vol. 98, no. 12, pp. 1781–1794, 2020.

[53] M. F. Neurath, "Cytokines in inflammatory bowel disease," *Nature Reviews Immunology*, vol. 14, no. 5, pp. 329–342, 2014.

[54] A. Cholapranee, G. S. Hazlewood, G. G. Kaplan, L. Peyrin-Biroulet, and A. N. Ananthakrishnan, "Systematic review with meta-analysis: comparative efficacy of biologics for induction and maintenance of mucosal healing in Crohn's disease and ulcerative colitis controlled trials," *Alimentary Pharmacology & Therapeutics*, vol. 45, no. 10, pp. 1291–1302, 2017.

[55] P. D. Lu and Y. H. Zhao, "Targeting NF-κB pathway for treating ulcerative colitis: comprehensive regulatory characteristics of Chinese medicines," *Chinese Medicine*, vol. 15, no. 1, p. 15, 2020.

[56] B. Liu, X. Piao, W. Niu et al., "Kuijieyuan decoction improved intestinal barrier injury of ulcerative colitis by affecting TLR4-dependent PI3K/AKT/NF-κB oxidative and inflammatory signaling and gut microbiota," *Frontiers in Pharmacology*, vol. 11, p. 1036, 2020.

[57] L. Wang, B. Walia, J. Evans, A. T. Gewirtz, D. Merlin, and S. V. Sitaraman, "IL-6 induces NF-kappa B activation in the intestinal epithelia," *The Journal of Immunology*, vol. 171, no. 6, pp. 3194–3201, 2003.

[58] X. D. Zhu, Y. F. Cao, Y. Wang, and Y. Q. Duan, "Effect of tongxieyaofang on the expression of NF-κB p65 gene and protein in colonic tissue of ulcerative colitis rats," *Chinese Journal of Gerontology*, vol. 34, no. 5, pp. 1288–1281, 2021.

[59] Q. F. Li, S. J. Liu, and D. H. Jing, "An Inquiry into treatment mechanism of ulcerative colitis by Tongxie Yaofang based on clinical efficacy and Inflammatory balance change in Th1/Th2 and Th17/Treg," *LiaoNing Journal Of Traditional Chinese Medicine*, vol. 8, no. 45, pp. 1569–1572.

[60] X. D. Zhu, L. Z. Li, Y. Q. Duan, Y. F. Cao, and B. Wang, "Influence of tongxieyao formula on inflammatory factors in rats with ulcerative colitis," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 20, no. 4, pp. 41–43, 2021.

[61] B. J. Colleypriest, S. G. Ward, and D. Tosh, "How does inflammation cause Barrett's metaplasia?" *Current Opinion in Pharmacology*, vol. 9, no. 6, pp. 721–726, 2009.

[62] S. Sedda, V. Dinallo, I. Marafini et al., "mTOR sustains inflammatory response in celiac disease," *Scientific Reports*, vol. 10, no. 1, Article ID 10798, 2020.

[63] G. Iyer, M. I. Milowsky, and D. F. Bajorin, "Novel strategies for treating relapsed/refractory urothelial carcinoma," *Expert Review of Anticancer Therapy*, vol. 10, no. 12, pp. 1917–1932, 2010.

[64] X. Lin, Q. Sun, L. Zhou et al., "Colonic epithelial mTORC1 promotes ulcerative colitis through COX-2-mediated Th17 responses," *Mucosal Immunology*, vol. 11, no. 6, pp. 1663–1673, 2018.

[65] T. Aoki and S. Narumiya, "Prostaglandin E(2)-EP2 signaling as a node of chronic inflammation in the colon tumor microenvironment," *Inflammation and Regeneration*, vol. 37, no. 1, p. 4, 2017.

[66] J. X. Guo and L. G. Yan, "Effects of Tong Xie Yao Fang and Fangfeng on COX-2 and IL-6 in peripheral blood of colonic mucosa in rats with ulcerative colitis caused by liver depression and spleen deficiency," *Traditional Chinese Medicinal Research*, vol. 30, no. 11, pp. 73–77, 2021.

[67] W. Wang, T. Xia, and X. Yu, "Wogonin suppresses inflammatory response and maintains intestinal barrier function via TLR4-MyD88-TAK1-mediated NF-κB pathway in vitro," *Inflammation Research*, vol. 64, no. 6, pp. 423–431, 2015.

[68] I. A. Lee, E. J. Kim, and D. H. Kim, "Inhibitory effect of β-sitosterol on TNBS-induced colitis in mice," *Planta Medica*, vol. 78, no. 09, pp. 896–898, 2012.

[69] A. Y. Chen and Y. C. Chen, "A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention," *Food Chemistry*, vol. 138, no. 4, pp. 2099–2107, 2013.

[70] M. E. V. Johansson, J. K. Gustafsson, J. Holmén-Larsson et al., "Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis," *Gut*, vol. 63, no. 2, pp. 281–291, 2014.

[71] Y. Qu, X. Li, F. Xu et al., "Kaempferol alleviates murine experimental colitis by restoring gut microbiota and inhibiting the LPS-TLR4-NF-κb Axis," *Frontiers in Immunology*, vol. 12, Article ID 679897, 2021.

[72] L. O. Regasini, J. C. R. Vellosa, D. H. S. Silva et al., "Flavonols from *Pterogyne* nitens and their evaluation as myeloperoxidase inhibitors," *Phytochemistry*, vol. 69, no. 8, pp. 1739–1744, 2008.

[73] A. Kanashiro, J. G. Souza, L. M. Kabeya, A. E. C S Azzolini, and Y. M. Lucisano-Valim, "Elastase release by stimulated neutrophils inhibited by flavonoids: importance of the catechol group," *Zeitschrift Für Naturforschung C*, vol. 62, no. 5-6, pp. 357–361, 2007.

[74] M. Y. Park, G. E. Ji, and M. K. Sung, "Dietary kaempferol suppresses inflammation of dextran sulfate sodium-induced colitis in mice," *Digestive Diseases and Sciences*, vol. 57, no. 2, pp. 355–363, 2012.

[75] Y. Xiong, D. Chen, C. Yu et al., "Citrus nobiletin ameliorates experimental colitis by reducing inflammation and restoring impaired intestinal barrier function," *Molecular Nutrition & Food Research*, vol. 59, no. 5, pp. 829–842, 2015.