

Research Article

Predicting Coupled Herbs for the Treatment of Hypertension Complicated with Coronary Heart Disease in Real-World Data Based on a Complex Network and Machine Learning

Jia-Ming Huan ,¹ Yun-Lun Li,² Xin Zhang,³ Jian-Liang Wei,³ Wei Peng,³ Yi-Min Wang,² Xiao-Yi Su,² Yi-Fei Wang ,³ and Wen-Ge Su ³

¹School of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250014, China

²First School of Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250014, China

³The Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250014, China

Correspondence should be addressed to Yi-Fei Wang; 15853150460@163.com and Wen-Ge Su; 71000768@sutcm.edu.cn

Received 16 October 2021; Revised 20 November 2021; Accepted 5 January 2022; Published 22 January 2022

Academic Editor: Xuezhong Zhou

Copyright © 2022 Jia-Ming Huan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hypertension and coronary heart disease are the most common cardiovascular diseases, and traditional Chinese medicine is applied as an auxiliary treatment for common cardiovascular diseases. This study is based on 3 years of electronic medical record data from the Affiliated Hospital of Shandong University of Traditional Chinese Medicine. A complex network and machine learning algorithm were used to establish a screening model of coupled herbs for the treatment of hypertension complicated with coronary heart disease. A total of 5688 electronic medical records were collected to establish the prescription network and symptom database. The hierarchical network extraction algorithm was used to obtain core herbs. Biological features of herbs were collected from public databases. At the same time, five supervised machine learning models were established based on the biological features of the coupled herbs. Finally, the K-nearest neighbor model was established as a screening model with an AUROC of 91.0%. Seventy coupled herbs for adjuvant treatment of hypertension complicated with coronary heart disease were obtained. It was found that the coupled herbs achieved the purpose of adjuvant therapy mainly by interfering with cytokines and regulating inflammatory and metabolic pathways. These results show that this model can integrate the molecular biological characteristics of herbs, preliminarily screen combinations of herbs, and provide ideas for explaining the value in clinical applications.

1. Introduction

Every year, 10.4 million people die of complications of hypertension worldwide [1]. Organ damage caused by hypertension and cardiovascular disease (CVD) are currently the main causes of death [2]. Due to the abnormal increase in arterial pressure, the coronary artery is more likely to experience increased local tensile stress, causing endothelial injury, the accumulation of lipid particles, induction of inflammatory reactions, and acceleration of the growth of plaques [3]. More than 50% of hypertension sufferers have multiple cardiovascular risk factors, and 25%–30% of coronary heart disease (CHD) patients have hypertension [4–6].

Traditional Chinese medicine (TCM) is recommended as a complementary and alternative therapy in the treatment of hypertension and CHD in China. An existing systematic review shows that TCM herbs can improve the vascular endothelial function of patients with hypertension, inhibit inflammatory reactions, regulate blood lipids, and improve mood. Adjuvant TCM therapy more easily achieves the targeted blood pressure and improves the comfort of patients, protecting target organs and reducing cardiovascular events [7–9].

In the theory of TCM, herbs are combined to form a prescription according to the principle of “Jun-Chen-Zuo-Shi.” Among them, the herbs of “Jun” that play a core role

are often applied as drug combinations, usually no more than three herbs for the core disease. On the other hand, there are obvious individual differences in concurrent diseases. TCM theory emphasizes that medication for individual patients should have suitable herb pairing rules. But it is difficult to reflect its curative effect by relying on traditional randomized controlled trials (RCTs). To fully reflect the disease characteristics of the individual and groups of patients, it is necessary to analyze them on the basis of real-world data (RWD) [10]. The traditional summary of drug compatibility is often based on the physician's long-term diagnosis and treatment experience, and the screening period takes too long to apply to rapidly changing disease spectrum. Through the algorithm analysis of RWD, we can quickly find the drug compatibility information hidden in the effective prescriptions, carry out bioinformatics analysis to achieve preliminary screening, and provide ideas for further experimental verification and clinical applications to accelerate and summarize the application rules of herbs.

Therefore, we developed a mining method based on RWD to explore effective coupled herbs for the treatment of hypertension complicated with CHD by combining their symptom information and target information. Existing network pharmacology and bioinformatics techniques have been widely used to discover the core targets of herbs, and the understanding of multiple targets with TCM therapy has become increasingly profound [11, 12]. However, most of the existing studies are based on the relationship between herbs, compounds, targets, and diseases. Previous studies [13] used the Dijkstra algorithm to integrate the symptom information emphasized by TCM into the herb information network. However, it is still unable to fully evaluate the closeness of herb combinations and complicated diseases, including the target similarity of different herbs and their contribution to the curative effect.

Supervised machine learning can aggregate a variety of herb feature information to generate a model, matching the input features of the herbs with the expected output to form a learning function and to complete the classifier after adjusting the parameters by cross-validation. Commonly used models include K-nearest neighbor (KNN), support vector machine (SVM), gradient boosting decision tree (GBDT), Bayesian network (BN), and logistic regression (LR). KNN is more sensitive to the local information in the feature space of the input herbs, while SVM and others reflect its global characteristics.

In this study, we established a prescription database and symptom database of patients with hypertension complicated with CHD. First, we used the hierarchical network extraction algorithm to extract the main herbs and symptoms from the database, collected biological information, established a biological network, including herbal compounds, targets, and related disease symptom information, and then used supervised machine learning models compared with the classical Apriori algorithm. The best model was used to evaluate the pertinence of each coupled herb in the treatment of hypertension complicated with CHD.

2. Materials and Methods

2.1. Data Preparation. In this study, 5688 electronic medical records (EMRs) of hypertension complicated with CHD collected from the Affiliated Hospital of Shandong University of Traditional Chinese Medicine (between July 1, 2014, and May 31, 2017) were used to extract and standardize the symptoms in the prescription and medical history of TCM [14], and the prescription database and symptom database were established. After that, we took the herbs in the prescription database as the node and the frequency of the two herbs as the weight. The hierarchical network extraction algorithm uses Liquorice software [15] to obtain the core herbs of the prescription network based on the degree coefficient prescription = 1.9.

2.1.1. Identifying Compounds and Targets of Core Herbs. We identified compounds and targets of core herbs from online public databases: Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP) [16], SymMap [17], and The Encyclopedia of Traditional Chinese Medicine (ETCM) [18], and published biomedical literature in the PubMed and CNKI databases. The names of the compounds were merged after being unified by PubChem and UniProt. The target of the core herbs was imported into the STRING database [19], and a protein–protein interaction network (PPIN) with a confidence ≥ 0.9 of *Homo sapiens* was established.

2.1.2. Identifying the TCM Symptoms of Core Herbs. The SymMap database contains TCM symptoms corresponding to herbs and their compounds, and the TCM symptoms of the core herbs can be obtained from the SymMap database and compared with the symptom database in EMRs to screen and determine the symptoms of hypertension complicated with CHD that can be effectively treated by each herb.

2.1.3. Collecting the Related Genes of Hypertension and CHD. The expression data of hypertension and CHD were retrieved from MalaCards [20] and the NCBI GEO database. The GSE76845 dataset contains 5 hypertension patient samples and 5 healthy control samples. The GSE71226 dataset contains 3 coronary heart disease patient samples and 3 healthy control samples. Then, the differentially expressed genes with $Q < 0.05$ and adj. $P < 0.05$ were analyzed by the GEO2R tool.

2.2. Evaluating the Features of the Coupled Herbs. The coupling of herbs can increase the pertinence of disease treatment. In this study, the core herbs coupled with each other. The main clinical and biological features of the herbs were quantified to evaluate the action characteristics of the coupled herbs.

2.2.1. Frequency Assessment of the Coupled Herbs. To evaluate the correlation of the two matched herbs in the prescription database, we established the frequency matrix of the herbs in the prescriptions. The Manhattan distance [21] between two herbs was calculated to evaluate their coupling characteristics in the prescription.

2.2.2. Symptom Similarity Assessment. The Jaccard similarity coefficient is used to compare the similarity between the sample sets. We obtained the symptom set of hypertension complicated with CHD and the herbal regulation symptom sets by screening the information of the RWD symptom database and the SymMap database. The Jaccard similarity coefficient was used to compare the similarity between the sample sets. We obtained the TCM symptom information of compound-regulated hypertension complicated with CHD by screening the information of the EMRs symptom database and the SymMap database, which was used to calculate the Jaccard similarity coefficient to evaluate the closeness of the herb-related symptoms to the disease-related symptoms.

2.2.3. Bioavailability Assessment. Oral bioavailability (OB) represents the percentage of oral doses reaching the systemic circulation, and high OB is usually a key indicator for identifying bioactive molecules as having therapeutic properties. In this study, the OB values of the herbal compounds obtained from the SymMap database were added to evaluate the bioavailability of the coupled herb.

2.2.4. Herb Target Identification and Functional Enrichment Analysis. A full understanding of human biological function cannot be realized by individual genes, only by the ubiquitous interaction between different genes. Therefore, the random walk with restart (RWR) algorithm was used in PPIN to evaluate the connection degree of herbal targets to disease-related genes. The disease-related genes in PPIN were used as seed node sets, and the restart probability was 0.75 [22]. The stable probability of diffusion to each herbal target was obtained by the RWR operation, which was realized in the PyRWR package (version 1.0.0) in Python 3.7.5. The summation was used to evaluate the regulation of disease-related proteins of herbs.

The semantic comparison of gene ontology (GO) annotations provides a method to calculate the similarity between genes and genomes. To measure the similarity between herbal targets and disease-related genes, GO biological process semantic similarity (GoSim) was used to evaluate its effectiveness. We relied on the annotated data provided by *Bioconductor* and used the algorithm designed by Wang et al. [23], implemented in the GOSemSim package [24] (version 2.12.1) in R 3.6.3.

2.3. Machine Learning Model Training. We took the lift value of the coupled herbs calculated by the Apriori algorithm as the classification criterion and the evaluation value of the effectiveness of the coupled herbs as the input information for the

machine learning tasks (Figure 1) and used the *kknn* (version 1.3.1), *e1071* (version 1.7-8), *gbm* (version 2.1.8), and *klaR* (version 0.6-15) packages in R 3.6.3 to complete the model training of KNN, SVM, GBDT, BN, and LR. Ten cross-validations were used to evaluate the performance of the models. We, in particular, randomly divide the performance data of the coupled herbs into 10 subsets, selected one as the test set in turn, and repeated the other 9 training sets 10 times. According to the verification of a large amount of data [25], 10% can obtain the best error estimate, which can be used to prevent overfitting of the model, and all drug pair data can be used as training sets and test sets to effectively avoid data waste.

2.4. Analysis of the Mechanism of Coupled Herbs. We used the *clusterProfiler* package [26] (version 3.14.3) in R 3.6.3 to annotate genes with *org.Hs.eg.db* (version 3.10.0) and analyzed the related genes of the effective coupled herbs by KEGG enrichment analysis. Based on the hypergeometric distribution, $Q < 0.05$ was considered a significant enrichment pathway, and the same method was used for GO enrichment analysis, $Q < 0.05$ and adj. $P < 0.05$ as significant enrichment. Hierarchical clustering (HCT) was used to classify the herbs and pathways to distinguish the biological processes of intervention. At the same time, we introduced genes into *Metascape* [27] for multigene list meta-analysis, including functional proteomics and gene screening.

3. Results

3.1. The Core Herbs. A total of 5,689 electronic medical records data of Chinese medicines for the treatment of hypertension and CHD were collected to establish a database. Among them, the TCM prescription database contains 3697 prescriptions, which included 442 herbs used 85662 times, and each prescription contained, on average, 23.17 ± 10.38 herbs. A total of 234 disease-related symptoms were found in the symptoms database. Through the hierarchical extraction algorithm, we obtained 18 core herbs, including *Atractylodis Macrocephalae Rhizoma*, *Citri Reticulatae Pericarpium*, *Glycyrrhizae Radix Et Rhizoma*, *Rhizoma Pinelliae*, *Poria*, *Radix Salviae Liguliobae*, *Codonopsis Radix*, *Rhizoma Coptidis*, *Radix Angelicae Sinensis*, *Astragali Radix*, *Radix Paeoniae Alba*, *Puerariae Lobatae Radix*, *Chuanxiong Rhizoma*, *Radix Notoginseng*, *Ophiopogonis Radix*, *Rehmanniae Radix*, *Schisandrae Chinensis Fructus*, *Ziziphi Spinosa Semen*.

After 18 herbs were coupled with each other, 77 coupled herbs (Figure 2) were selected as the main herb pairs through the Apriori algorithm, of which *Schisandrae Chinensis Fructus*, *Codonopsis Radix*, *Astragali Radix*, *Radix Notoginseng*, *Chuanxiong Rhizoma*, *Ziziphi Spinosa Semen*, *Radix Angelicae Sinensis*, *Atractylodis Macrocephalae Rhizoma*, *Radix Paeoniae Alba*, and *Rhizoma Coptidis* accounted for a high proportion (Table 1).

3.2. Genes Related to Hypertension Complicated with CHD. As shown in Figure 3, 211 hypertension-related genes were obtained from GSE76845. The main upregulated genes

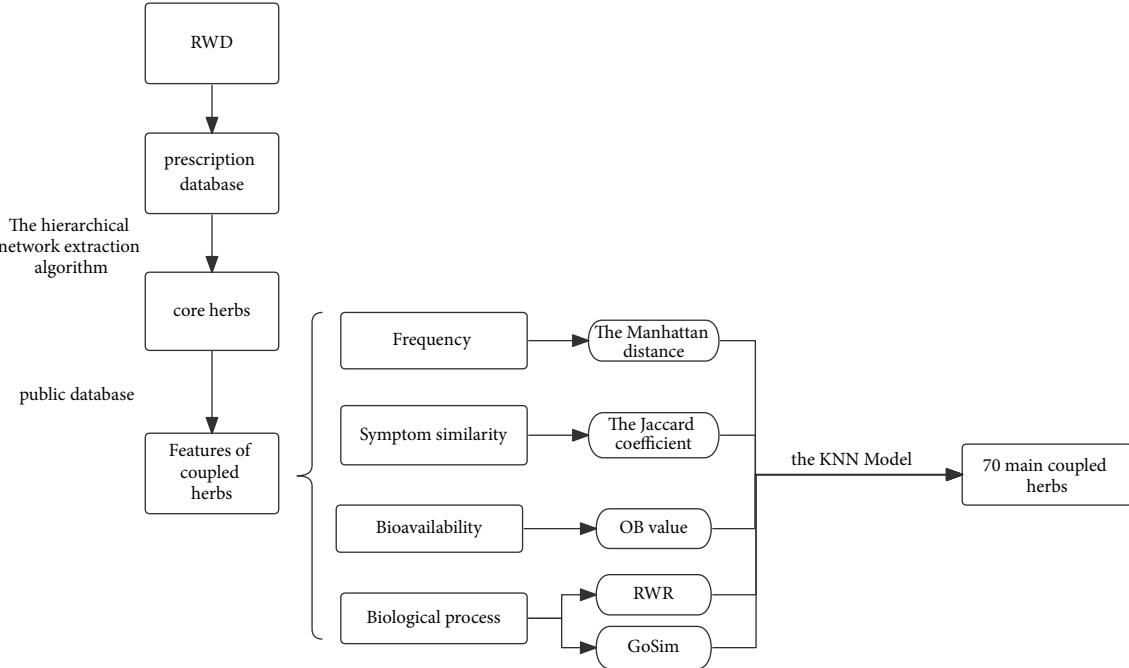


FIGURE 1: Overview of herb screening: first, we used the hierarchical extraction algorithm to analyze the core herbs from the prescription database. Second, we obtained the features of the coupled herbs from a public database. Finally, we input the features into the machine learning model to obtain 70 main coupled herbs.

included RNF6, UBE2L5, MTX1, FTL, SLITRK2, OGG1, HMGA1, RHAG, and TRIM8. The main downregulated genes included SPAG17, ALPG, ENO1, ESYT2, DNAJB1, HFE, RAPSN, DLGAP5, and OST4. Then, 246 CHD-related genes were obtained from GSE71226. The main upregulated genes included TM2D1, SRSF11, SLC33A1, NFATC2IP, GPRIN3, FRG1JP, DROSHA, and CHD2. The main downregulated genes included YBX3, TMCC3, OAZ2, MME, KRTTN1, IL1R2, GYPA, and HYPA. MalaCards obtained 33 hypertension-related genes and 20 CHD-related genes.

3.3. Performance of the Models. In this study, 5 machine learning models, KNN, SVM, GBDT, BN, and LR, were used to analyze the effectiveness feature data of the coupled herbs and establish the coupled herbs classification. As shown in Figure 4, the area under the receiver operating characteristic (ROC) curve (AUROC) was used to evaluate the performance of the models. We found that the 5 models had good classification performance. Among them, the KNN effect was the best, with an AUROC of 91.0%. BN, GBDT, and SVM were all over 80%, 87.3%, 82.9%, and 80.4%, respectively. The GBDT was 76.6%.

3.4. Prediction of the Coupled Herbs. The performance of the KNN model was the best when the parameter k was 9. According to its best cutoff point value, we obtained a herb combination relationship network between the paired drugs containing 18 herbs. There were 70 main coupled herbs. The top 15 effective coupled herbs screened by the KNN model

are shown in Table 2. Combining the information of Tables 1 and 2, we found that Schisandrae Chinensis Fructus, Codonopsis Radix, Radix Notoginseng, Atractylodis Macrocephalae Rhizoma, Ziziphi Spinosae Semen, and Radix Paeoniae Alba were the 6 herbs most related to the other drugs. The percentages of Radix Notoginseng, Atractylodis Macrocephalae Rhizoma, Ziziphi Spinosae Semen, and Radix Paeoniae Alba increased compared to the result of Apriori, while the percentages of Astragali Radix, Chuanxiong Rhizoma, and Radix Angelicae Sinensis decreased. At the same time, there were Citri Reticulatae Pericarpium-Rhizoma Pinelliae, Citri Reticulatae Pericarpium-Poria, Schisandrae Chinensis Fructus-Ophiopogonis Radix, and Radix Notoginseng-Rhizoma Coptidis which are individually strongly associated drugs.

3.5. Validation of the Models' Effect. To verify the effect of the KNN model, we used the data diagnosed as hypertensive nephropathy (HN) in the standardized electronic medical record data from the Affiliated Hospital of Shandong University of Traditional Chinese Medicine. GEO2R was used to analyze GSE99325 to obtain HN-related genes. The final AUROC obtained using the KNN model was 94.2% with parameter k being 4. The effective therapeutic coupled herbs obtained include Radix Salviae Ligulobae-Astragali Radix, Radix Achyranthis Bidentatae-Radix Angelicae Sinensis, Astragali Radix-Radix Angelicae Sinensis, and Ophiopogonis Radix-Poria. According to recent HN-related studies [28–32], all of the coupled herbs screened in the KNN model have therapeutic effects on HN.

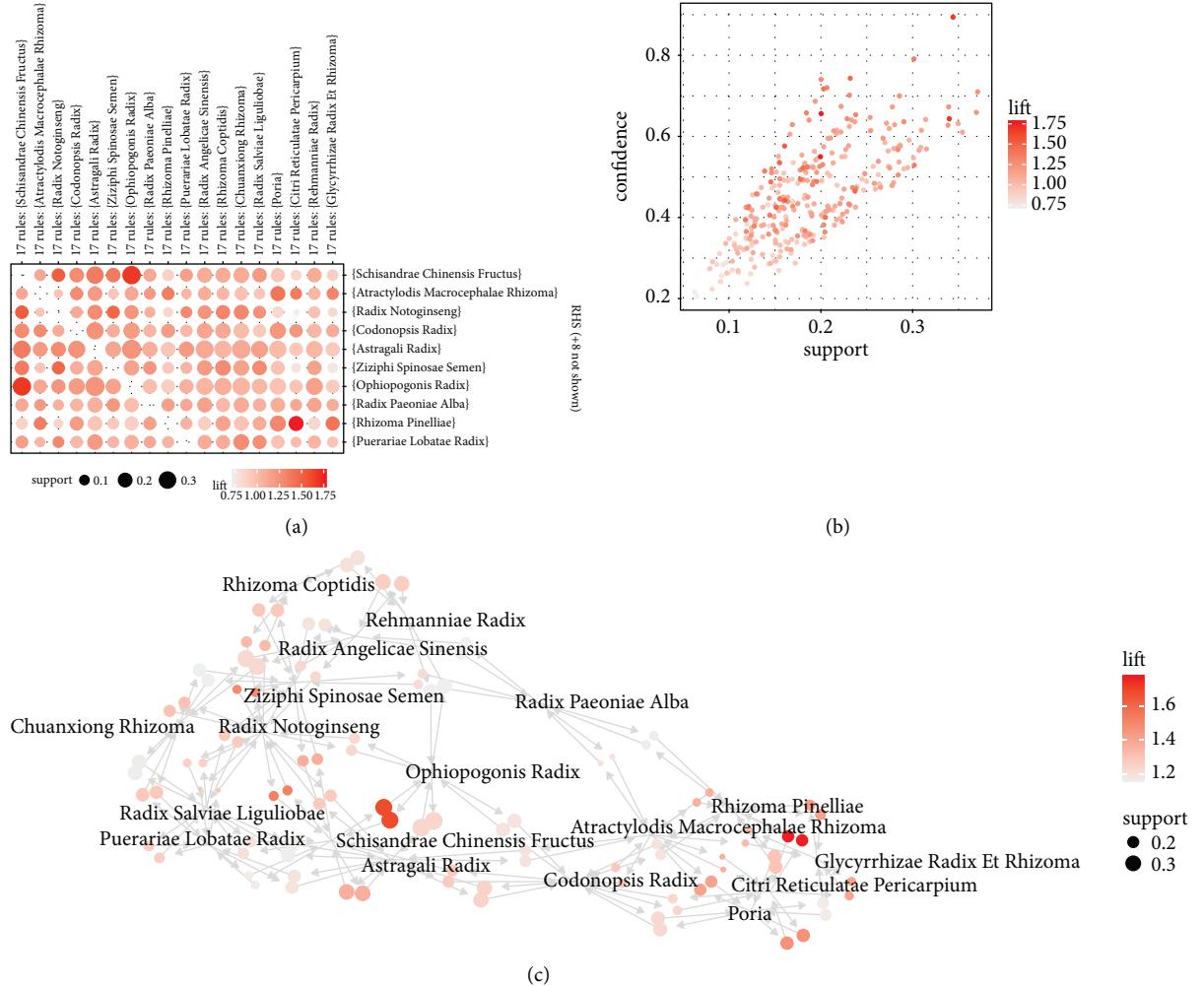


FIGURE 2: Results of apriori: (a) and (c) are the paired relationships between the core herbs obtained by the apriori algorithm. The size of the dot indicates the support between two herbs, and the darker the color of the dot, the larger the lift. (b) shows the coupling feature of all herbs obtained by the apriori algorithm.

TABLE 1: Regional division of the core herbs.

Core herbs	Proportion in the main coupled herbs		Regional division
	According to apriori (%)	According to the KNN model (%)	
Radix paeoniae alba	5.84	6.52	K1
Atractylodis macrocephala rhizoma	5.84	7.97	K1
Chuanxiong rhizoma	6.49	5.80	K1
Rehmanniae radix	4.55	1.45	K1
Ophiopogonis radix	4.55	5.07	K1
Astragali radix	6.49	5.80	K2
Puerariae lobatae radix	5.19	5.07	K3
Ziziphi spinosae semen	6.49	7.25	K3
Codonopsis radix	7.14	7.97	K4
Schisandrace chinensis fructus	8.44	7.97	K4
Rhizoma pinelliae	5.19	5.07	K5
Citri reticulatae pericarpium	3.90	4.35	K5
Radix angelicae sinensis	6.49	5.80	K5
Radix salviae liguliobae	4.55	4.35	K5
Poria	3.25	2.90	K5
Glycyrrhizae radix et rhizoma	3.25	2.90	K5
Rhizoma coptidis	5.84	5.80	K5
Radix notoginseng	6.49	7.97	K5

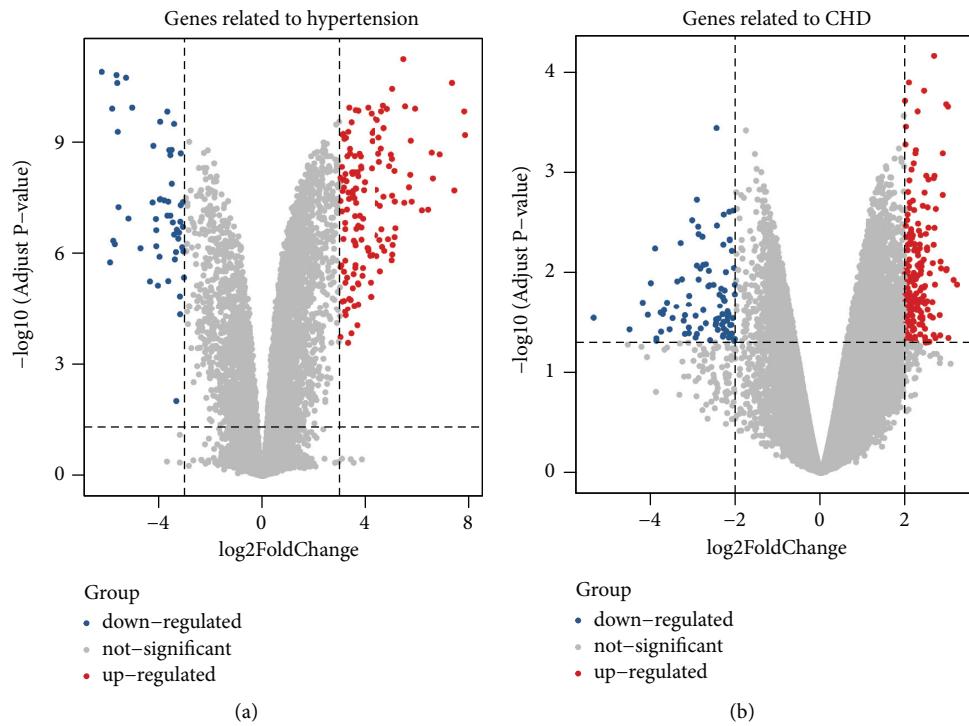


FIGURE 3: Genes related to hypertension (a) and CHD (b): a blue dot indicates a downregulated gene, and a red dot indicates an upregulated gene.

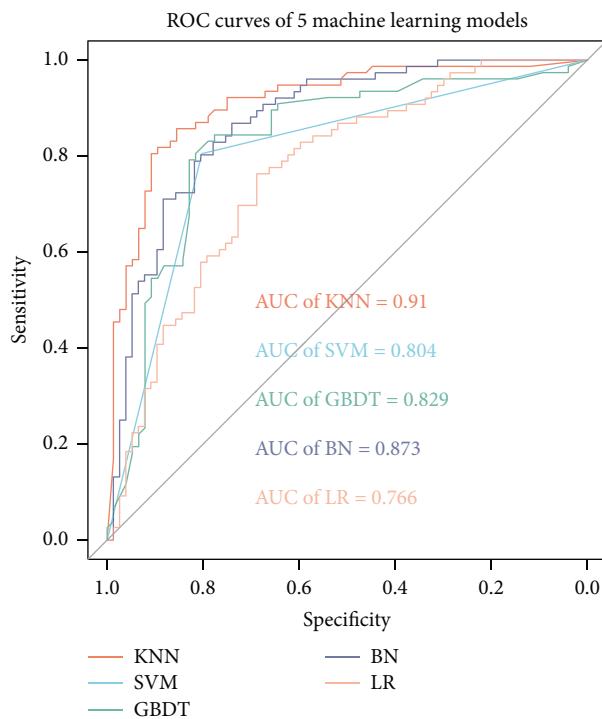


FIGURE 4: ROC curves of 5 machine learning models.

TABLE 2: The top 15 effective coupled herbs screened by the KNN model.

Coupled herbs	
Astragali radix	Ophiopogonis radix
Ophiopogonis radix	Schisandrae chinensis fructus
Atractylodis macrocephalae rhizoma	Citri reticulatae pericarpium
Citri reticulatae pericarpium	Poria
Rhizoma pinelliae	Citri reticulatae pericarpium
Radix notoginseng	Ziziphi spinosae semen
Chuanxiong rhizoma	Radix angelicae sinensis
Codonopsis radix	Schisandrae chinensis fructus
Ziziphi spinosae semen	Schisandrae chinensis fructus
Radix paeoniae alba	Atractylodis macrocephalae rhizoma
Rhizoma pinelliae	Poria
Puerariae lobatae radix	Radix notoginseng
Atractylodis macrocephalae rhizoma	Poria
Radix paeoniae alba	Codonopsis radix
Radix notoginseng	Schisandrae chinensis fructus

3.6. Therapeutic Mechanism of the Coupled Herbs. The targets of the core herbs were analyzed by GO and KEGG enrichment analysis, and the pathways in which the number of enriched genes for each herb was greater than the quartile of related genes of the herb were retained.

A total of 215 KEGG signaling pathways are obtained. We established an 18×215 -dimensional feather profile of the core pathways. HCT divided the KEGG signaling pathways into 7 parts. As shown in Figure 5, based on the KEGG signaling pathways, 18 herbs were divided into 5 parts, each with similar enrichment results. The reserved coupled herbs screened by the KNN model were 59 in different clusters and 10 in the same cluster. Among the excluded coupled herbs, there were 59 in different clusters and 25 in the same cluster. In the reserved coupled herbs, herbs in different clusters were more abundant ($P = 0.04$). In addition, the number of pathways in which the number of enriched genes was greater than the average number of herbs coupled with herbs in different clusters was larger ($P < 0.01$), indicating that the biological processes of the pairing of herbs in different clusters were more extensive.

For example, Atractylodis Macrocephalae Rhizoma, Ophiopogonis Radix, Radix Paeoniae Alba, Chuanxiong Rhizoma, and Ziziphi Spinosa Semen in K1 regions have fewer enriched genes in the inflammatory pathways, such as the HIF-1 signaling pathway, TNF signaling pathway, NOD-like receptor signaling pathway, NF-kappa B signaling pathway, and 31 coupled herbs related to them; 16 were combined with herbs in the K5 region to supplement the regulation of the inflammatory pathway. At the same time, compared with the K5 region, the K1 region had fewer enriched genes in glucose and lipid metabolism, such as insulin resistance, insulin secretion, cholesterol metabolism, and the regulation of lipolysis in adipocytes, so it was necessary to combine the K5 region herbs. A total of 37.5% of the coupled herbs contained herbs in K2 and K4, and the herbs in K4 were mainly enriched in leukocyte transendothelial migration, the calcium signaling pathway, platelet activation, and other pathways related to the formation of coronary artery plaques.

GO enrichment found that the number of enriched genes in the biological process (BP) category of the KNN-coupled herbs was different from that of the deleted coupled herbs ($P = 0.04$), and the molecular function (MF) and cellular component (CC) categories were not different. From the results of GO enrichment in Figure 6(a), the regulatory effects of the core herbs were mainly cytokines, chemokines, growth factors, and the regulation of cell metabolism. As shown in Figure 6(b), the regulatory proteins of the core herbs and the genes related to hypertension complicated with CHD can be coenriched on the main nodes of the GO term network, suggesting that the paired core herbs can cooperate and complement each other, resulting in the regulation of hypertension complicated with CHD.

4. Discussion

The prescription rule of “Jun-Chen-Zuo-Shi” means that different combinations of herbs in the prescription need to be effective against multiple symptoms of clinical diseases. At the same time, the combined use of herbs is in keeping with the TCM theory which emphasizes that the core herbs enhance the efficacy of specific disease symptoms. Herbs are rich in components and match the network regulation mechanism of the disease. Experiments verification for the total mechanism is tedious and expensive. In recent years, with advances in bioinformatics research and the advent of the concept of multitarget drugs, information on compounds and targets has been extracted and screened in large herb databases and biological databases. Evaluating and predicting the efficacy of herb compound prescriptions and deducing their action mechanism has become a widely used research model [13, 15, 33].

The existing herbal pairing research is based on a single commonly used prescription, establishes a drug-gene-disease network and explores the correlation characteristics between nodes, analyzes the interaction law of herbal combinations after prescription decomposition, and discusses the rationality of its curative effect [34–36]. However, in clinical practice, such prescriptions are often used as “Jun” or “Chen” in patients’ prescriptions to regulate major disease

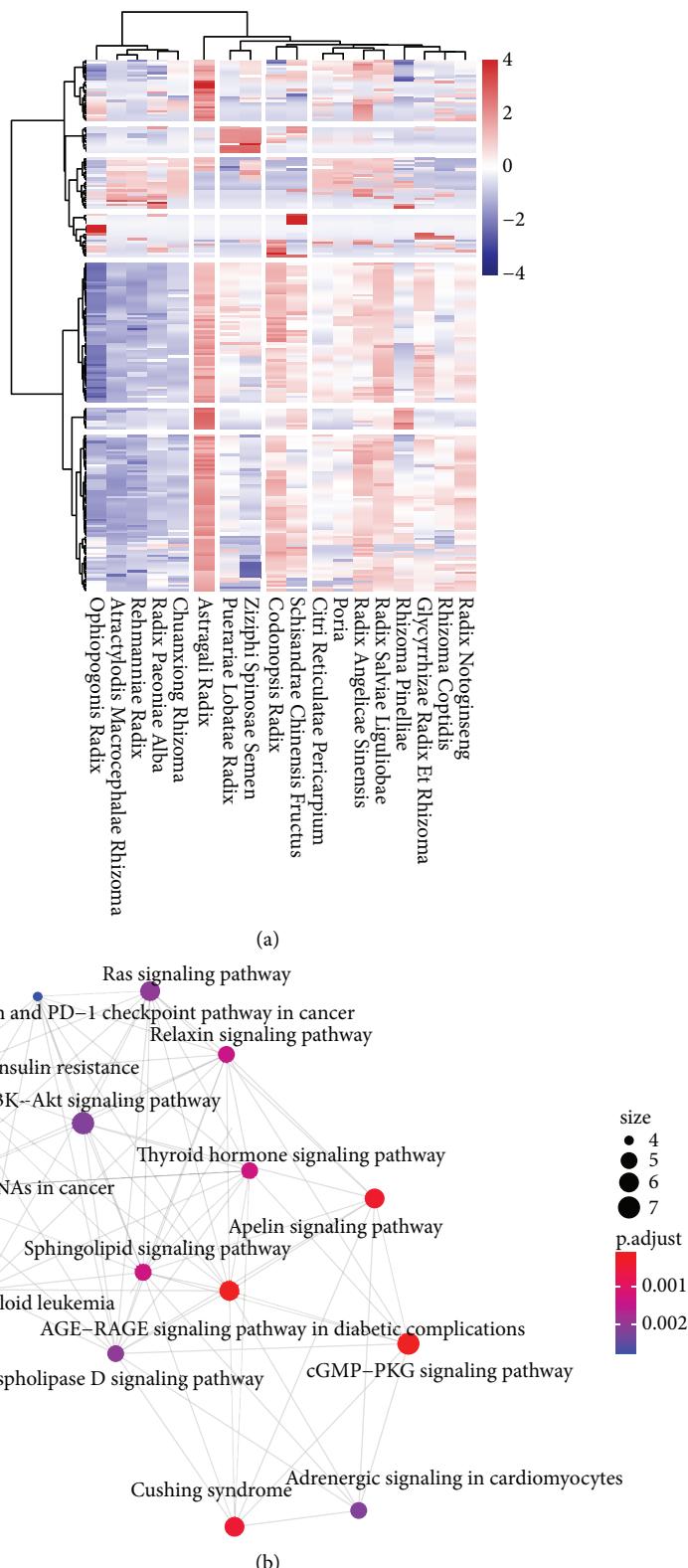


FIGURE 5: The results of KEGG pathway enrichment: heatmap and hierarchical clustering result of the KEGG signaling pathways (a) and the KEGG pathway network (b).

symptoms, which seriously limits the summary of the findings on the treatment of concurrent diseases and symptoms. Therefore, this study used 5688 EMRs as the data

source, extracted the patients' TCM symptoms, and used the hierarchical network extraction algorithm to extract common drugs for the treatment of hypertension complicated

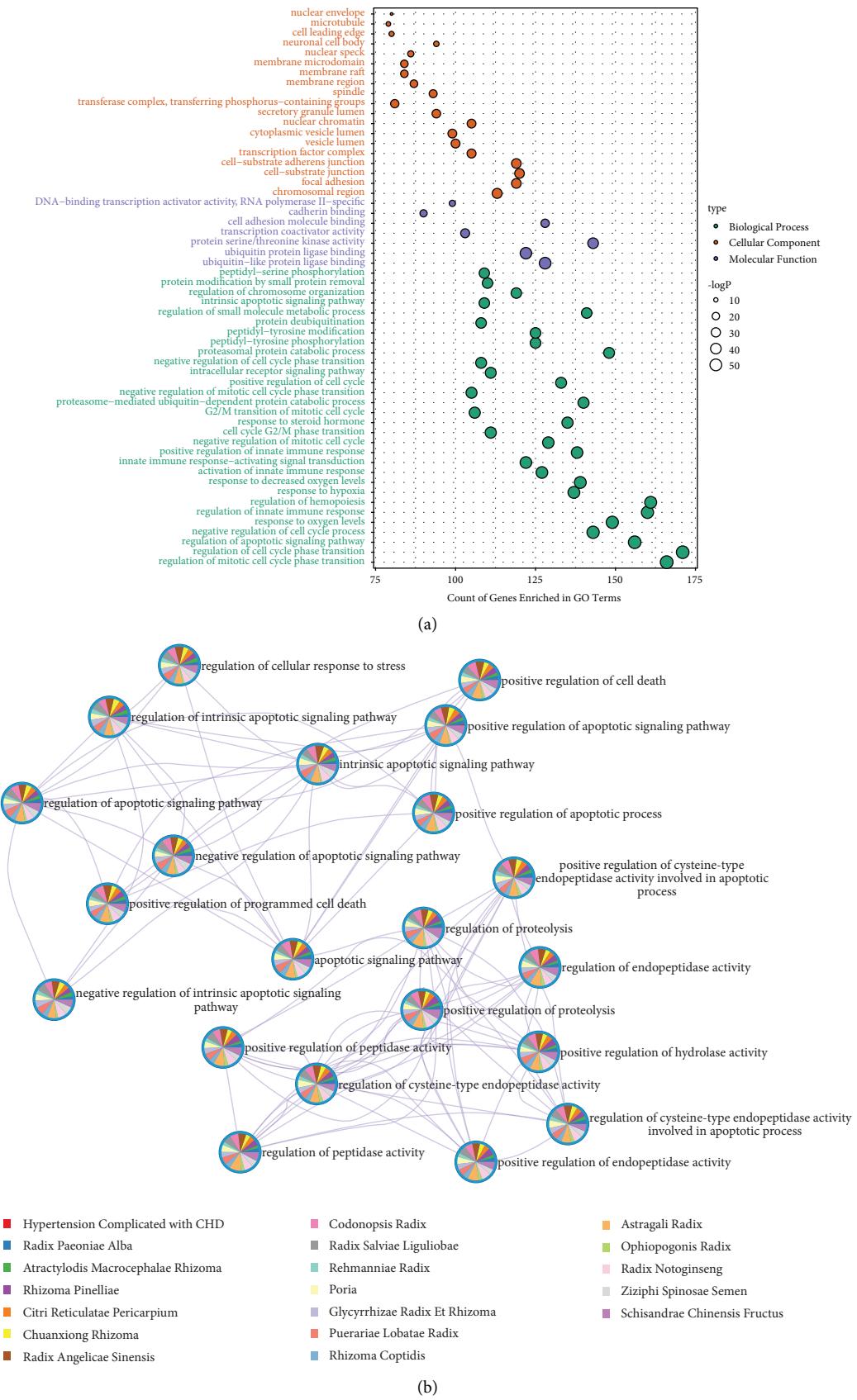


FIGURE 6: The results of the GO term enrichment: (a) is the enrichment result of biological process, cellular component and molecular function. (b) is a network of enriched GO terms represented as pie charts, where pie pieces are color-coded based on the identities of the genes in the herbs and hypertension complicated with CHD. The thicker the line, the more common the targets of the nodes and the closer their interaction.

with CHD in the prescription network, which is not limited to a single doctor or genre. Treatment experience can be widely collected and summarized, but it also increases the heterogeneity of the dataset. To screen out truly effective paired drugs, we established a prediction model for screening the effectiveness of herbs, which integrates different types of information, including the bioavailability of compounds, the degree of association between herbs and the clinical symptoms and the biological functions.

According to previous research [33], the effective coupled herbs in the complete prescriptions of TCM include drugs with different pharmacological directions and application frequencies. Therefore, this study quantifies the effectiveness of coupled herbs on the basis of PPIN. First, we incorporate TCM symptom information into herbal PPIN but, at the same time, it also brings more noise and biological dimensions. By using RWR to amplify the connection information between nodes, the connection probability between herbal targets and disease symptom-related proteins was evaluated, and their closeness was quantified. At the same time, we used Manhattan distance to evaluate the correlation between the two herbs in the clinical application, the Jaccard coefficient to evaluate the similarity between paired herbal symptoms and the disease symptom database, and the OB to determine the bioavailability of the herbs. After calculating the effectiveness of five classical classification machine learning models, the best KNN is selected as the classification model with an AUROC of 91.0%, and when the HN data are used to evaluate the effectiveness of the model, the AUROC is 94.2%. Therefore, according to the best cutoff point of the KNN model, we finally identified 70 effective coupled herbs.

Hypertension and CHD are common cardiovascular diseases with overlapping risk factors. After an increase in blood pressure, the renin–angiotensin–aldosterone system (RAAS) is overactivated, which induces an inflammatory reaction of target organs such as blood vessels, myocardium, and kidney, resulting in the upregulation of inflammatory cytokines. The coronary artery induces vascular endothelial injury, promotes lipids to enter the intima, causes platelet aggregation, and accelerates the growth of plaques [3, 37–39]. To further analyze the therapeutic pathways of coupled herbs, we carried out GO and KEGG enrichment analysis on each herbal target. Finally, we found that coupled herbs can cooperate with each other, directly or indirectly acting on the inflammatory pathway, further assist the body in controlling risk factors by regulating blood lipids, proteins, and carbohydrate metabolism, regulate the expression of various cytokines at the cellular level, and regulate their target proteins to inhibit the inflammatory response and interstitial fibrosis of target organs caused by hypertension.

The analysis results of GO and KEGG enrichment showed that the herbal combinations with synergistic effects could be correctly identified by a machine learning model with the quantitative indicators of herbal effectiveness used in this study, but this study also contains many limitations. First, the herbal-related compounds and target information come from public database. Although we used a variety of models to compare and verify to find the best model, the data

deviation of the database may have a potential impact on the research results. Second, although the TCM symptom information of patients has been included in the evaluation system, it fails to make full use of the clinical information of individual differences in EMRS. Third, the sample size for model training is still limited, and information on the interaction between compounds has not been added, so future research needs to improve the performance of the model to design a more effective adjvant therapy for TCM.

5. Conclusion

This study was based on data from 5688 EMRs of hypertension complicated with CHD in the Affiliated Hospital of Shandong University of Traditional Chinese Medicine from 2014 to 2017, and a patient symptom dataset and prescription dataset were established. Eighteen commonly used herbs were obtained, and their biological network was established. Using the interaction information between nodes in the network, we established quantitative data on the effectiveness of coupled herbs. An effective coupled herb screening algorithm based on a machine learning model was proposed, and a total of 70 coupled herbs were obtained. Based on the analysis of various herbs at the pathway level, they can play a multilevel biological regulatory role in controlling the inflammatory response and regulating energy metabolism for hypertension complicated with CHD and its TCM symptoms. Combined with complex networks and machine learning, this study explored the potential law of herbs in the treatment of hypertension complicated with CHD and predicted the curative effect of coupled herbs, providing a direction for summarizing the TCM EMRs and explaining the TCM rules.

Abbreviations

CVD:	Cardiovascular disease
CHD:	Coronary heart disease
TCM:	Traditional Chinese medicine
RCT:	Randomized controlled trial
RWD:	Real-world data
KNN:	K-nearest neighbor
SVM:	Support vector machine
GBDT:	Gradient boosting decision tree
BN:	Bayesian network
LR:	Logistic regression
EMR:	Electronic medical record
TCMSP:	Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform
ETCM:	The Encyclopedia of Traditional Chinese Medicine
PPIN:	The protein–protein interaction network
RWR:	Random walk with restart
GO:	Gene ontology
GoSim:	Gene ontology biological process semantic similarity
OB:	Oral bioavailability
HCT:	Hierarchical clustering

Baizhu:	Atractylodis Macrocephalae Rhizoma
Chenpi:	Citri Reticulatae Pericarpium
Gancao:	Glycyrrhizae Radix Et Rhizoma
Banxia:	Rhizoma Pinelliae
Fuling:	Poria
Danshen:	Radix Salviae Liguliobae
Dangshen:	Codonopsis Radix
Huanglian:	Rhizoma Coptidis
Danggui:	Radix Angelicae Sinensis
Huangqi:	Astragali Radix
Baishao:	Radix Paeoniae Alba
Gegen:	Puerariae Lobatae Radix
Chuanxiong:	Chuanxiong Rhizoma
Sanqi:	Radix Notoginseng
Maidong:	Ophiopogonis Radix
Dihuang:	Rehmanniae Radix
Wuweizi:	Schisandrae Chinensis Fructus
Suanzaoren:	Ziziphi Spinosa Semen
ROC:	Receiver operating characteristic
AUROC:	Area under the receiver operating characteristic curve
HN:	Hypertensive nephropathy
Niuxi:	Radix Achyranthis Bidentatae
Maidong:	Ophiopogonis Radix
RAAS:	Renin–angiotensin–aldosterone system.

Data Availability

The TCM prescription data used to support the findings of this study are available from the corresponding author upon request. The data generated by the analysis process can be found in the article databank.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank the Affiliated Hospital of Shandong University of Traditional Chinese Medicine for its support of this study. This work was supported by The Jinan Science and Technology Bureau Clinical Medical Science and Technology Innovation Program (No. 202019149); based on network pharmacology analysis of the mechanism of real-world core prescriptions for hypertensive renal damage; Shandong Province Medicine and Health Science and Technology Development Project (No. 202103010371); research on the mechanism of ACE2/Ang 1-7/Mas axis regulating hypertensive nephropathy based on proteomics analysis; and the Construction Project of the National Clinical Research Base of Traditional Chinese Medicine for Hypertension (Issued by the State Administration of Traditional Chinese Medicine [2008] No. 23).

Supplementary Materials

1. Apriori: rules of the core herbs obtained for the Apriori algorithm. 2. Features of coupled herbs: The value of Jaccard

similarity coefficient, OB, RWR, GoSim, Manhattan distance, and KNN model output of coupled herbs. 3. The main coupled herbs: Seventy coupled herbs for adjuvant treatment of hypertension complicated with coronary heart disease screened by the KNN model. 4. TCM herbs information: the Chinese name, Pinyin name, Latin name, and English name of herbs in this study. (*Supplementary Materials*)

References

- [1] T. Unger, C. Borghi, F. Charchar et al., “2020 international society of hypertension global hypertension practice guidelines,” *Hypertension*, vol. 75, no. 6, pp. 1334–1357, 2020.
- [2] T. B. Olesen, M. Pareek, J. V. Stidsen et al., “Association between antecedent blood pressure, hypertension-mediated organ damage and cardiovascular outcome,” *Blood Pressure*, vol. 29, no. 4, pp. 232–240, 2020.
- [3] F. Mach, C. Baigent, A. L. Catapano et al., “2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk,” *European Heart Journal*, vol. 41, no. 1, pp. 111–188, 2020.
- [4] D. Etehadi, C. A. Emdin, A. Kiran et al., “Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis,” *The Lancet*, vol. 387, no. 10022, pp. 957–967, 2016.
- [5] G. S. Collins and D. G. Altman, “An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study,” *BMJ (Clinical research ed.)*, vol. 339, Article ID b2584, 2009.
- [6] R. Conroy, K. Pyörälä, A. P. Fitzgerald et al., “Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project,” *European Heart Journal*, vol. 24, no. 11, pp. 987–1003, 2003.
- [7] X. Xiong, P. Wang, Y. Zhang, and X. Li, “Effects of traditional Chinese patent medicine on essential hypertension: a systematic review,” *Medicine*, vol. 94, no. 5, p. e442, 2015.
- [8] X.-J. Xiong, X.-C. Yang, W. Liu et al., “Therapeutic efficacy and safety of traditional Chinese medicine classic herbal formula longdanxiegan decoction for hypertension: a systematic review and meta-analysis,” *Frontiers in Pharmacology*, vol. 9, p. 466, 2018.
- [9] X. Xiong, P. Wang, L. Duan et al., “Efficacy and safety of Chinese herbal medicine xiao yao san in hypertension: a systematic review and meta-analysis,” *Phytomedicine*, vol. 61, Article ID 152849, 2019.
- [10] Z. Shu, Y. Zhou, K. Chang et al., “Clinical features and the traditional Chinese medicine therapeutic characteristics of 293 COVID-19 inpatient cases,” *Frontiers of Medicine*, vol. 14, no. 6, pp. 760–775, 2020.
- [11] S. Li, “Network pharmacology evaluation method guidance-draft,” *World Journal of Traditional Chinese Medicine*, vol. 7, no. 1, pp. 146–154, 2021.
- [12] S. Li and B. Zhang, “Traditional Chinese medicine network pharmacology: theory, methodology and application,” *Chinese Journal of Natural Medicines*, vol. 11, no. 2, pp. 110–120, 2013.
- [13] Y. Yang, K. Yang, T. Hao et al., “Prediction of molecular mechanisms for lianxia ningxin formula: a network pharmacology study,” *Frontiers in Physiology*, vol. 9, p. 489, 2018.
- [14] Y. F. Wang, J. J. Wang, W. Peng et al., “Identification of hypertension subgroups through topological analysis of symptom-based patient similarity,” *Chinese Journal of Integrative Medicine*, vol. 27, pp. 1–10, 2021.

- [15] X. Zhou, S. Chen, B. 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.
- [16] 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.
- [17] 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, pp. D1110–D1117, 2019.
- [18] 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.
- [19] 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.
- [20] N. Rappaport, M. Twik, I. Plaschkes et al., "MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search," *Nucleic Acids Research*, vol. 45, no. D1, pp. D877–D887, 2017.
- [21] D. Sinwar and R. Kaushik, "Study of euclidean and manhattan distance metrics using simple k-means clustering," *International Journal for Research in Applied Science and Engineering Technology*, vol. 2, no. 5, pp. 270–274, 2014.
- [22] J. Yang, S. Tian, J. Zhao, and W. Zhang, "Exploring the mechanism of TCM formulae in the treatment of different types of coronary heart disease by network pharmacology and machine learning," *Pharmacological Research*, vol. 159, Article ID 105034, 2020.
- [23] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen, "A new method to measure the semantic similarity of GO terms," *Bioinformatics*, vol. 23, no. 10, pp. 1274–1281, 2007.
- [24] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang, "GOSemSim: an R package for measuring semantic similarity among GO terms and gene products," *Bioinformatics*, vol. 26, no. 7, pp. 976–978, 2010.
- [25] N. Wang, P. Li, X. Hu et al., "Herb target prediction based on representation learning of symptom related heterogeneous network," *Computational and Structural Biotechnology Journal*, vol. 17, pp. 282–290, 2019.
- [26] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "ClusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [27] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.
- [28] Y. Dong, B. Yue, M. Qian et al., "JYYS granule mitigates renal injury in clinic and in spontaneously hypertensive rats by inhibiting NF- κ B signaling-mediated microinflammation," *Evidence-Based Complementary and Alternative Medicine*, vol. 2018, Article ID 8472963, 13 pages, 2018.
- [29] J. Owoicho Orgah, M. Wang, X. Yang et al., "Danhong injection protects against hypertension-induced renal injury via down-regulation of myoglobin expression in spontaneously hypertensive rats," *Kidney & Blood Pressure Research*, vol. 43, no. 1, pp. 12–24, 2018.
- [30] L. Wu, M. Liu, and Z. Fang, "Combined therapy of hypertension nephropathy with brevicanine injection and anti-hypertensive drugs: a systematic review and a meta-analysis," *Evidence-Based Complementary and Alternative Medicine: eCAM*, vol. 2018, Article ID 2958717, 17 pages, 2018.
- [31] Y. Li, S. Yan, L. Qian, L. Wu, Y. Zheng, and Z. Fang, "Danhong injection for the treatment of hypertensive nephropathy: a systematic review and meta-analysis," *Frontiers in Pharmacology*, vol. 11, p. 909, 2020.
- [32] J. M. Huan, W. G. Su, W. Li et al., "Summarizing the effective herbs for the treatment of hypertensive nephropathy by complex network and machine learning," *Evidence-Based Complementary and Alternative Medicine: eCAM*, vol. 2021, Article ID 5590743, 12 pages, 2021.
- [33] N. Wang, N. Du, Y. Peng et al., "Network patterns of herbal combinations in traditional Chinese clinical prescriptions," *Frontiers in Pharmacology*, vol. 11, Article ID 590824, 2021.
- [34] H. Y. Fang, H. W. Zeng, L. M. Lin et al., "A network-based method for mechanistic investigation of shexiang baoxin pill's treatment of cardiovascular diseases," *Scientific Reports*, vol. 7, Article ID 43632, 2017.
- [35] W. Zhou, X. Cheng, and Y. Zhang, "Effect of liuwei dihuang decoction, a traditional Chinese medicinal prescription, on the neuroendocrine immunomodulation network," *Pharmacology & Therapeutics*, vol. 162, pp. 170–178, 2016.
- [36] G. Yu, Y. Zhang, W. Ren et al., "Network pharmacology-based identification of key pharmacological pathways of yin-huang-qing-fei capsule acting on chronic bronchitis," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 12, pp. 85–94, 2016.
- [37] D. Chen, M. Liang, C. Jin, Y. Sun, D. Xu, and Y. Lin, "Expression of inflammatory factors and oxidative stress markers in serum of patients with coronary heart disease and correlation with coronary artery calcium score," *Experimental and Therapeutic Medicine*, vol. 20, no. 3, pp. 2127–2133, 2020.
- [38] H. Schuett, R. Oestreich, G. H. Waetzig et al., "Transsignaling of interleukin-6 crucially contributes to atherosclerosis in mice," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 32, no. 2, pp. 281–290, 2012.
- [39] Q. Su, L. Li, Y. Sun, H. Yang, Z. Ye, and J. Zhao, "Effects of the TLR4/myd88/NF- κ B signaling pathway on NLRP3 inflammasome in coronary microembolization-induced myocardial injury," *Cellular Physiology and Biochemistry*, vol. 47, no. 4, pp. 1497–1508, 2018.