Kurdish Studies
March 2023
Volume: 11, No: 1, pp. 495-503
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

Received: January 2023 Accepted: February 2023
DOI: 10.53555/ks.v11i1.2941

Herb target prediction using protein complexes detection and machine learning methods in heterogeneous network

Parastoo Fathi[1*], Nasrollah Moghaddam Charkari[2]

[1*]Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran.
[2]Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran, Email:

***Corresponding author:** Parastoo Fathi
*Postal address: Tehran, Chamran Highway,Jalal Ale Ahmad Street ,

Abstract
Traditional Chinese medicine (TCM) is known for its diverse components, multiple therapeutic targets, and intricate mechanisms that offer considerable advantages in disease treatment. Identifying drug targets stands at the core of TCM network pharmacology research. Over the years, various web tools targeting drug discovery with diverse features have been created to aid in target prediction, thereby significantly advancing the field of drug discovery. This study presents the Herb-Symptom-Target Network for predicting herb-target associations via symptom-related links. It aims to create concise feature vectors for herb-protein and herb-symptom interactions in a complex network. Through node resource allocation and gene expression data, protein network weights are assigned to identify protein complexes using core-attachment and second-order neighbors. To understand the molecular mechanisms of herbs based on their clinical effects, machine learning techniques are utilized to predict complex herbal protein interactions. This involves integrating heterogeneous information from herbal medicines, symptoms, corresponding targets, and their relationships. As a result, potential targets of herbs within protein complexes are identified. The prediction of herb targets involves extracting feature vectors for network edges, which are then refined using supervised learning approaches. Experimental results demonstrate the method's effectiveness and validity when compared to state-of-the-art herb-target prediction techniques. Additionally, several predicted herb-target interactions have been manually validated using independent literature, confirming the method's potential to successfully merge diverse information for predicting new herb-target interactions.

Keywords: Heterogeneous network, Protein complex , Herb-target interaction, Herbal medicine

1- Introduction

Traditional Chinese Medicine (TCM) draws on ancient wisdom to create remedies like Fangji [1, 2]. The 'single disease, single drug' approach faces challenges with complex diseases, leading to a focus on drug combinations for increased efficacy, cost reduction, and drug repurposing [3-5]. Understanding interactions between drugs and target proteins is vital for drug discovery [6]. TCM's formula-based treatments encounter difficulties due to complex ingredients and unclear mechanisms [7, 8]. Research targets improving efficacy and reducing side effects through network-based screening and in silico drug selection [9, 10]. A network-based method is recommended for predicting drug groups and examining correlations [11], with the DrugCombDB facilitating exploration of therapeutic effects [12]. Biomarker networks predict drug combinations, while network perturbation analysis enhances efficacy and minimizes side effects [13-14]. Strategies for reducing drug toxicity involve studying target interactions in protein networks and using neural networks for multi-drug side effect modeling [15, 16]. Network methods like CIPHER and drugCIPHER are employed for disease gene and target prediction [17-18].The integration of multi-omics data [19] aids in predicting synergistic cancer drugs, while the CE-BLAST tool calculates antigen distances for antiviral vaccine design[20]. Traditional Chinese Medicine (TCM) research employs systems pharmacology to illustrate herbal treatment effects for complex diseases[21-22]. Integrative pharmacology is set to reshape TCM formula exploration by leveraging the TCMIP database [23]. Furthermore, a standardized template using gene expression data from 102 herbal ingredients unveils the synergistic mechanisms within herbal formula components from a network modular viewpoint [24-25]. Bioinformatics advancements, employing technologies such as multi-omics and biological networks [26–30], have paved the way for innovative drug pathways, particularly in TCM formulas. Progress has been evident in the development of medications derived from TCM herbs [31–34]. Studies have applied data mining, machine learning, and neural networks to forecast outcomes related to herbal formulas [35-39]. Integrating traditional theory with modern pharmacology remains a challenge, treating herbal formula prediction as a multi-label classification task [40-41]. Developing appropriate scores for training herbal formulas is a significant hurdle [42-43]. Topic models, FangNet, and neural networks are utilized in herbal formula prediction, emphasizing the importance of merging classical theory with modern mechanisms [44-47]. The TCMFP method combines Traditional Chinese Medicine expertise, artificial intelligence, and network algorithms to identify the most effective herbal formulas for treating complex diseases [47]. Our paper presents an Herb-Target interaction approach that combines herb, symptom, and protein data to tackle complex biological datasets. The method involves constructing a weighted heterogeneous

network, identifying protein complexes, and employing two learning methods. A new technique, NRAGE-WPN [48], leverages resource allocation and gene expression in Protein-Protein Interaction networks to improve accuracy by detecting dense complex cores. The approach learns low-dimensional vector representations for herb-protein and herb-symptom interactions, integrating contextual and topological details. These representations, along with herb compounds and protein/gene ontology data, are utilized as features for node learning. The model is trained and tested using supervised learning methods, with performance compared against other herb-target prediction techniques to evaluate its efficacy.

2-Materials and Methods

2-1Data Acquiring and Processing

In this study, data on herbs, symptoms, and targets, along with their interactions, are collected from various public databases and literature sources, as outlined in Table 1. Specifically, herb-target relationships are obtained from the HIT2 database [49], serving as the dataset for the classification model's training and evaluation. Associations between herbs and indications are sourced from the Chinese Pharmacopoeia (CHPA, 2015 edition). Herb-herb relationships are formed based on a referenced study [50], while herb-efficacy associations are derived from the CHPA dataset. Using these herb-efficacy connections, vectors representing herbs based on efficacy are created. The cosine similarities between herb pairs are then calculated to establish herb-herb relationships, with these similarity metrics utilized as edge weights in the network.

Table 1. All dataset for finding relation between all nodes.

Name	Composition	Source
Herb-target associations	1254 herbs, 2208 targets and 168797 herb-target links	HIT2
Herb-efficacy associations	829 herbs, 373 efficacies and 3830 herb-efficacy links	CHPA
Herb-symptom associations	465 herbs, 1027 symptoms and 16739 herb-symptom links	CHPA
Herb-herb associations	809 herbs and 1912 links	Herbs linked with similar efficacy
Protein-protein interactions	10622 proteins and 25133 interactions	String10
Symptom-symptom associations	1027 symptoms and 7220 links	SemMedDB

Assume, there are n types of herbs and m types of efficacies. Each herb is represented by a vector of efficacy $V_x = (e_1, e_2, \ldots, e_j, \ldots, e_m)$, where $e_i=1$ indicates that herb x has a relationship with efficacy j, otherwise there is no relationship. The efficacy-based cosine similarity of herbs x and y can be calculated by the following equation.

$$\cos(V_x, V_y) = \frac{V_x . V_y}{|V_x| . |V_y|} \quad (1)$$

The examination of connections between symptoms employed text-mining techniques. initially, these links were sourced from the Semantic MEDLINE Database (SemMedDB) [51-52]. Other studies have also investigated relationships between symptoms. In a study referenced as [52], the significance of each relationship is quantified using Fisher's exact test. Additionally, the Jacquard similarity measure has been employed in other research to assess similarities between symptoms. In our methodology, we identify interactions between symptoms based on shared neighbors within the herb layer. If the overlap between two distinct symptoms involves more than 500 herbs, they are considered to be interacting. Protein-protein interactions have been sourced from a well-known gene-gene interaction network database. Within this study, relationships with weights exceeding 700 [50, 53] are sieved, and a linear normalization process, represented by Equation (2), is applied. This normalization method, also known as min-max normalization, involves linearly transforming the initial data. Moreover, an innovative methodology that integrates resource allocation and gene expression in weighted Protein-Protein Interaction (PPI) networks is applied, comprising two principal phases. Initially, the reliability of protein interaction data is scrutinized by including common neighbor insights and gene expression profiles in constructing the weighted graph. Subsequently, the detection of protein complexes within this newly formed weighted graph is based on core-attachment and second-order neighbor analyses. The procedural steps for implementing this methodology are delineated in Figure 1.

$$Y' = \frac{y - \min(y)}{\max(y) - \min(y)} \quad 2$$

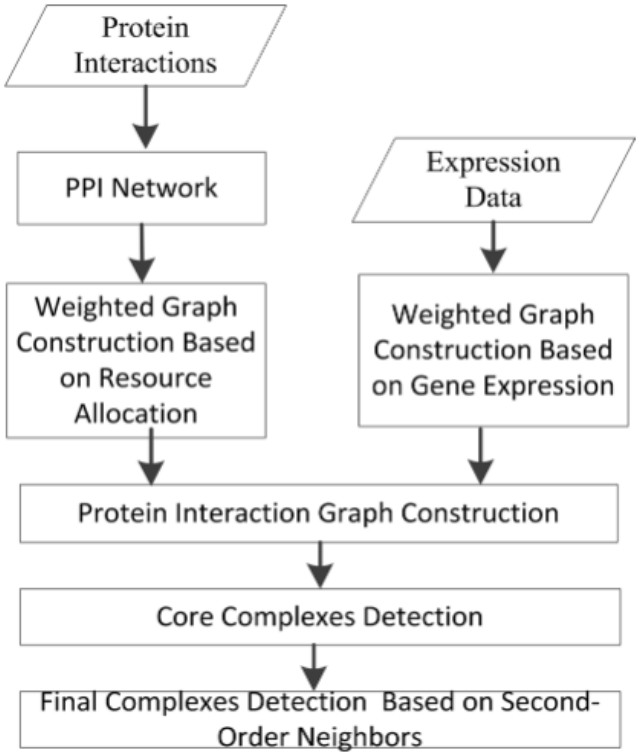

Figure 1 the workflow of construct PPI and finding core complex[48]

2-2 Heterogeneous Network Construction

Referring to Figure 2, a heterogeneous network is constructed utilizing three types of nodes: 1254 herbs, 1027 symptoms, and 2208 protein targets. Additionally, the network encompasses five types of associations, including 1912 herb-herb associations, 7220 associations between symptom, 25133 interactions between protein, 168797 herb-target associations between, and 16739 herb-symptom associations (as detailed in Table 1).

2-3 Detecting protein complexes

Referring to Figure 3, a novel algorithm, influenced by [48], is utilized to form protein complexes, involving two key stages: constructing a weighted graph and identifying core-attachment protein complexes by exploring second-order neighbors. The initial phase combines gene expression data and common neighbor details in weighted graph construction, detailed further in [48]. Equation 3 introduces density and diameter metrics as parameters for defining protein complexes. A node is allocated to the present cluster (subgraph) if it fulfills the conditions specified in equation 4, meeting both criteria. Typically, γ is set around 0.7, and δ is commonly designated as 2.

(1) The density of a node V can be defined as the sum of the weights associated with each connection to this node.
(2) The network diameter refers to the shortest path within a cluster.

$$m = \sum_{e_{ij} \in E} w(e) \tag{3}$$

$$density(G) = \frac{2 * m}{(|N| * (|N| - 1)} \tag{4}$$

$$diameter \leq \delta \; and \; density \geq \gamma$$

2-3 Learning Edge Features

Graph embedding is a technique used to convert and encode the data structure from a high-dimensional space into a lower-dimensional feature space, making it easier for machine learning algorithms to work with [54]. The resulting latent representation, often in the form of a d-dimensional vector, encapsulates the structural properties of nodes by representing their features. In this study, similarity-based approaches are employed to predict herb targets, focusing primarily on the graph's topology. These methods aim to predict missing links by calculating a similarity score, denoted as s(vx, vy), for pairs of nodes (vx and vy) based on the structural graph features. Conversely, extensive research has been conducted on network embedding to automatically recognize and map the structural attributes of the network in a latent space [55]. In previous studies, simple chemical infrastructure and drug-target sequence data were used to represent drugs and targets, requiring abundant molecular information and high-dimensional features [56]. However, when obtaining herb compounds and molecular information to represent herbs and targets, network embedding might not be necessary. In this study, molecular data about drugs and the

structural characteristics of the network are leveraged to predict the undisclosed connections between herbs and targets. To assign weights to the network edges, various metrics such as common neighbors, Jaccard coefficients, Salton Index, Sorenson Index, Adamic Adar, resource allocation, Hub Promoted Index, Preferential Attachment, Leicht Holme Newman, Parameter Dependency, Local Cohesion Structure, Car Based Index, Individual Attraction Index, and mutual information features have been explored.

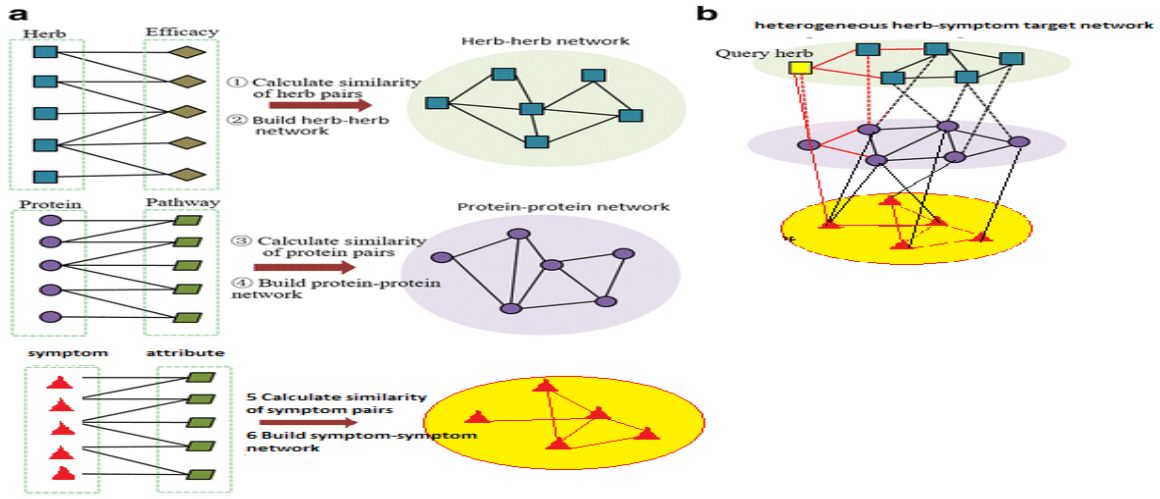

Figure 2 The overview of the proposed method is divided into two parts: (a) Construct a bipartite network between each pair of nodes (refer to Materials and Methods for detailed implementation). (b) Create a heterogeneous network by combining three types of nodes and five types of edges.

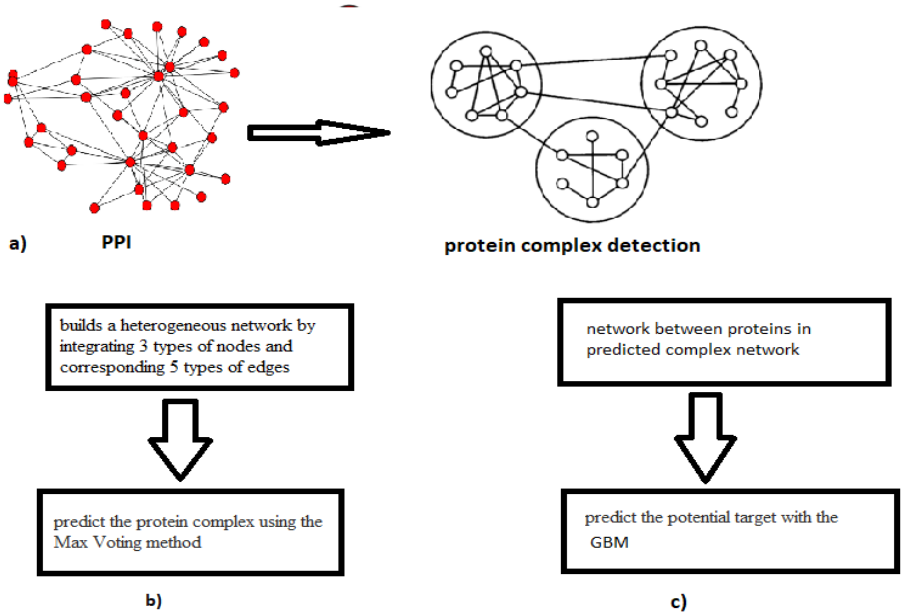

Figure 3 Continuing with the overview of the method in the second part: (a) Construct the Protein-Protein Interaction (PPI) network and identify all protein complexes within the PPI network. (b) In the first phase of learning, focus on predicting the protein complex associated with each herb. (c) In the second phase of learning, aim to predict potential targets using Gradient Boosting Machine (GBM).

2-4 Herb-Target Interaction prediction

Protein complexes refer to clusters of proteins collaborating to perform specific biological functions. Identifying protein complexes through protein-protein interaction networks aids in comprehending cellular mechanisms and holds potential therapeutic applications. Clustering protein interaction networks is a common technique for extracting protein complexes. Genuine protein complexes often demonstrate significant functional similarity, leading to the utilization of function enrichment tests to underscore the biological significance of identified protein complexes. Table 2 showcases the assessment and juxtaposition of p-values for each complex, segmented into four intervals varying from minor to substantial. A p-value exceeding 0.001 indicates that the complex's function is likely randomly assigned and lacks biological significance.

Table 2. The effectiveness of functional enrichment and their associated p-values

No. of clusters	Effective(%)	<E-15(%)	E-15-E-5(%)	E-5-0.001(%)
64	95.27	44.5	45.34	5.43

The feature representations derived from embedding learning, combined with the molecular specifics of herbs, are used as inputs for machine learning applications. These features play a crucial role in the herb-target edge prediction task, which seeks to predict the likelihood of the herb-target connection by leveraging their vector representations. Techniques such as Max or majority voting are part of ensemble supervised learning methods, where predictions from multiple models are combined. In this approach, the final prediction is determined by the prediction that receives the most votes or the majority of votes within the ensemble. Each model in the ensemble initially provides its prediction, and the prediction that appears most frequently across models is selected as the ultimate prediction. In our proposed method, two learning models are employed. Initially, the protein complex is identified using the Max Voting method in the first learner, following which, in the second phase, potential targets are predicted using Gradient Boosting. Gradient Boosting (GBM) utilizes boosting by amalgamating several weak learners to create a strong learner. Base learners in this approach are regression trees, with each subsequent tree in the series built on the mean squared error calculated by the preceding tree.

3-Results

3-1 Overview of method

A novel technique in the paper predicts herb-target interactions from a heterogeneous network with two phases: detecting protein complexes and forecasting herb-target interactions. It begins with building the network with 3 node types and 5 edge types, totaling 4489 nodes and 219801 edges. Latent node representations are extracted using network embedding. Protein complex detection includes building a weighted graph and finding protein complexes via second-order neighbors, merging gene data and common neighbors.

The proposed method encompasses dual learning phases: predicting protein complexes for each herb and forecasting potential protein targets within the proteins of the estimated complex. Ensemble supervised learning models, specifically max voting (used in the first learner) and Gradient Boosting Machine (GBM, utilized in the second learner), are employed to build classification models for forecasting herb-target interactions.

3-2 Experimental Results

In our proposed method, we have chosen the HIT2 database as the gold standard for herb-target interactions, containing a total of 168798 meticulously curated herb-target interactions. To assess the performance of the approach, a 10-fold cross-validation technique has been employed. In this study, an ensemble supervised learning approach involving Max Voting for the first step and Gradient Boosting Machine (GBM) for the second step has been implemented to evaluate the method's performance. The effectiveness of the method is evaluated using the Area Under the Receiver Operating Characteristic (AUROC) curve and the Area Under the Precision-Recall Curve (AUPR). Results from ensemble learning are depicted in Table 3. This study employs various supervised learning methods such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, and Neural Network classification algorithms to generate a range of models from the dataset instead of relying on Max Voting. The output of each model is compared with others, and the targets with the highest occurrence are selected as the desired herb targets. In the context of ensemble learning based on Max Voting, the target with the greatest frequency is considered the final prediction for determining herb targets. It is noteworthy that five targets with the highest probability are selected, along with their respective probabilities, as the output of each model. As illustrated in Figure 4, the proposed approach's prediction performance is compared with the PRINCE method, an iterative algorithm that disseminates information from known nodes to other nodes in the network. The evaluation was conducted on a dataset [57] comprising 310 herbs, 1065 targets, and 8933 herb-target connections. The findings reveal that the proposed method surpasses the PRINCE method with a 17% higher AUROC (67%) and a 2% higher AUPR (86%) [58, 59] caution that ROC can be overly optimistic when assessing algorithm performance, especially on imbalanced data sets. In contrast, the Max Voting model demonstrates superior results in this context. The AUPR performance of the Max Voting model highlights its effectiveness over PRINCE in predicting herb-target interactions within heterogeneous networks. This contrast can be attributed to their underlying methodologies. PRINCE initiates by focusing on target nodes to attract other nodes sharing similar information, ideal for extracting local features in evenly distributed networks. In contrast, the proposed method integrates a hybrid functional and structural feature learning approach, enabling the incorporation of more domain-specific information to comprehensively learn node information within the heterogeneous network.

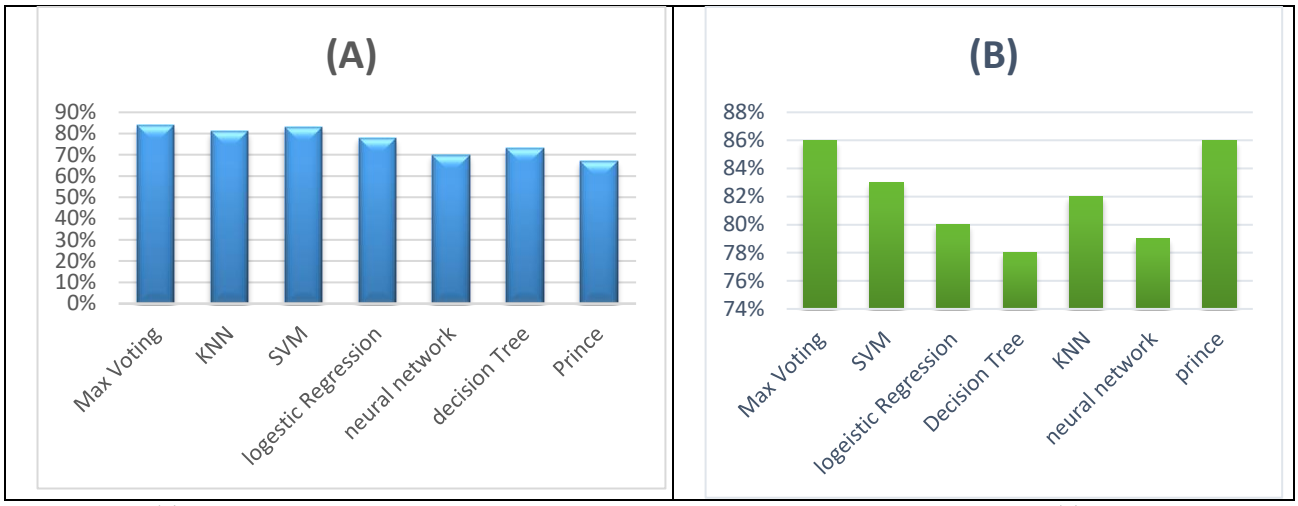

Figure 4. The performance of PRINCE and the proposed method was assessed on the HIT dataset, which comprises 310 herbs, 1065 targets, and 8933 herb-target links [57]. (A) AUROC. (B) AUPR.

Table 3. The performance of herb-target interaction prediction on dataset HIT2 [49]

Model	AUROC	AUPR
Max voting	90%	89%
KNN	84%	86%
SVM	85%	83%
Decision Tree	78%	77%
Neural network	74%	72%
Logistic Regression	81%	80%

4-Discussion

Network pharmacology has emerged as a modern avenue for exploring drug mechanisms and innovating new drug designs. Computational methodologies stand out as primary tools for forecasting the targets of herbal medicines. However, the intricate and diverse chemical structures of herbs pose limitations on the accuracy of predicting herb targets solely based on chemical similarity. This constraint arises from the insufficiency of a comprehensive and exhaustive database, compounded by the complexity of herb structural features. Moreover, the underlying mechanisms of most herbal formulations remain elusive for various reasons. Despite significant advancements in computational techniques and network pharmacology, unraveling the complete mechanisms behind the majority of herbal preparations still poses a challenge. This opacity underscores the ongoing complexity and depth of research required in the realms of herbal medicine and drug discovery. Many medicinal herbs contain numerous ingredients, complicating the distinction between effective and ineffective components. Identifying active compounds is challenging, and discovering their pharmacological targets is essential. The proposed research plan significantly mitigates these challenges by integrating diverse datasets. This integration process paves the way for understanding the molecular mechanisms of herbs through their clinical effects. A computational approach is employed to forecast herb targets, incorporating heterogeneous information on herbal medicines, symptoms, targets, and their interactions. The study utilizes a heterogeneous network to predict herb targets, utilizing lower-dimensional feature vectors for edges and applying machine learning methods for target prediction. This comprehensive approach marks a significant stride in advancing research on herb-target interactions and unraveling the molecular mechanisms behind the clinical effects of medicinal herbs. In the proposed method, seven key steps are involved. These steps include integrating databases, creating a heterogeneous network, extracting features, weighting the network, generating a new database, detecting protein complexes, and learning a model for goal prediction. Each of these sequential steps plays a crucial role in the overall methodology, contributing to the process of data integration, analysis, and prediction within the framework of the study. Following the establishment of a heterogeneous network using the dataset, a new dataset is derived based on diverse layers of information and molecular data concerning herbs and targets. Subsequently, prominent features utilized in chemical and herbal drug prediction contribute to weighting the network. The weighted heterogeneous network is then employed to construct a new dataset focusing on inter-layer communication. Traditional techniques (random walk, information diffusion, electrical resistance) enhance information spread [60-62]. Proposed method handles noisy, incomplete biological data by combining structural and functional features. This amalgamation aims to facilitate training and extraction of informative node representations within the network [63], underscoring a comprehensive and innovative approach to processing large-scale biological data. The results demonstrate the method's commendable performance compared to state-of-the-art approaches. Noteworthy advancements have been achieved when comparing its performance with the PRINCE algorithm in the heterogeneous network. Moving forward, our plan involves incorporating additional relevant data into our model, focusing on chemical drugs and diseases. Furthermore, we aim to explore symptom-symptom similarity measures [52] to enhance the depth and efficacy of our methodology. By expanding the scope of data integration and exploring new similarity measures, we anticipate further refining our approach and its outcomes in future research endeavors.

References

1- qaDongna Li, Jing Hu, Lin Zhang, Lili Li, Qingsheng Yin, Jiangwei Shi, Hong Guo, Yanjun Zhang, Pengwei Zhuang. Deep learning and machine intelligence: New computational modeling techniques for discovery of the combination rules and pharmacodynamic characteristics of Traditional Chinese Medicine. European Journal of Pharmacology. 2022; 933(): 175260. 10.1016/j.ejphar.2022.175260.

2- Xin Luan, Li-Jun Zhang, Xiao-Qin Li, Khalid Rahman, Hong Zhang, Hong-Zhuan Chen, Wei-Dong Zhang. Compound-based Chinese medicine formula: From discovery to compatibility mechanism. Journal of Ethnopharmacology. 2020; 254():112687.10.1016/j.jep.2020.112687.

3- Chan HCS, Shan H, Dahoun T, et al. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci. 2019;40:592–604. 10.1016/j.tips.2019.06.004

4- Ali MA, Rizvi S, Syed BA. Trends in the market for antihypertensive drugs. Nat Rev Drug Discov. 2017;16:309–10. 10.1038/nrd.2016.262

5- Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med. 2011; 3:96ra76. 10.1126/scitranslmed.3002648

6- Momtahen, S., Taajobian, M., Jahanian, A. Drug Discovery Acceleration Using Digital Microfluidic Biochip Architecture and Computer-aided-design Flow. International Journal of Engineering. 2019; 32(8): 1169-1176. 10.5829/ije.2019.32.08b.13

7- Wang J, Wong YK, Liao F. What has traditional Chinese medicine delivered for modern medicine? Expert Rev Mol Med. 2018;20:e4. 10.1017/erm.2018.3

8- Tang J, Tero A. Network pharmacology strategies toward multi-target anticancer therapies: from computational models to experimental design principles. Curr Pharm Des. 2014;20:23–36. 10.2174/13816128113199990470

9- Menche J, Sharma A, Kitsak M, et al. Uncovering disease–disease relationships through the incomplete interactome. Science. 2015;347:1257601. 10.1126/science.1257601

10- Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. Nat Commun. 2016; 7: 10331. 10.1038/ncomms10331

11- Ding P, Luo J, Liang C, et al. Discovering synergistic drug combination from a computational perspective. Curr Top Med Chem. 2018; 18(12):965–74. 10.2174/1568026618666180330141804

12- Liu H, Zhang W, Zou B, et al. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Res. 2019;48:D871–81. 10.1093/nar/gkz1007

13- Huang L, David B, Clifford S, et al. Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction. Bioinformatics. 2019;35:3709–17. 10.1093/bioinformatics/btz109

14- Woo JH, Shimoni Y, Wan SY, et al. Elucidating compound mechanism of action by network perturbation analysis. Cell. 2015;162:441–51. 10.1016/j.cell.2015.05.056

15- Nogales C, Mamdouh ZM, List M, et al. Network pharmacology: curing causal mechanisms instead of treating symptoms. Trends Pharmacol Sci. 2021;43:136–50. 10.1016/j.tips.2021.11.004

16- Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34:i457–66. 10.1093/bioinformatics/bty294

17- Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol Syst Biol. 2008;4:189. 10.1038/msb.2008.27

18- Zhao S, Shao L. Network-based relating pharmacological and genomic spaces for drug target identification. PloS One. 2010;5:e11764. 10.1371/journal.pone.0011764

19- Sun Y, Sheng Z, Ma C, et al. Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. Nat Commun. 2015;6:8481. 10.1038/ncomms9481

20- Qiu T, Yang Y, Qiu J, et al. CE-BLAST makes it possible to compute antigenic similarity for newly emerging pathogens. Nat Commun. 2018;9:1772. 10.1038/s41467-018-04171-2

21- Wang Y, Zheng C, Chao H, et al. Systems pharmacology dissecting holistic medicine for treatment of complex diseases: an example using cardiocerebrovascular diseases treated by TCM. Evid Based Complement Alternat Med. 2015;2015:980190. 10.1155/2015/980190

22- Chen X, Zheng C, Wang C, et al. Systems-mapping of herbal effects on complex diseases using the network-perturbation signatures. Front Pharmacol. 2018;9:1174. 10.3389/fphar.2018.01174

23- Xu H, Zhang Y, Wang P, et al. A comprehensive review of integrative pharmacology-based investigation: a paradigm shift in traditional Chinese medicine. Acta Pharm Sin B. 2021;11:1379–99. 10.1016/j.apsb.2021.03.024

24- Lv C, Wu X, Wang X, et al. The gene expression profiles in response to 102 traditional Chinese medicine (TCM) components: a general template for research on TCMs. Sci Rep. 2017;7:352. 10.1038/s41598-017-00535-8

25- Li B, Wang Y, Gu H, et al. Modular screening reveals driver induced additive mechanisms of baicalin and jasminoidin on cerebral ischemia therapy. Front Cardiovasc Med. 2022;9:813983. 10.3389/fcvm.2022.813983

26- Gu P, Chen H. Modern bioinformatics meets traditional Chinese medicine. Brief Bioinform. 2014;15:984–1003. 10.1093/bib/bbt063

27- Quan Y, Wang ZY, Xiong M, et al. Dissecting traditional Chinese medicines by omics and bioinformatics. Nat Prod Commun. 2014;9:1391–6. 10.1177/1934578X1400900942

28- Qiuxiao Zhan, Junnan Zhao, Lu Liu, Biqing Wang, Jiaqi Hui, Quan Lin, Yuxuan Qin, Bing Xue, Fengqin Xu. Integrated network pharmacology and molecular docking analyses of the mechanisms underlying the antihypertensive effects of lotusine. European Journal of Pharmacology.2023; 945: 175622. 10.1016/j.ejphar.2023.175622.
29- Wang X, Wang ZY, Zheng JH, et al. TCM network pharmacology: a new trend towards combining computational, experimental and clinical approaches. Chin J Nat Med. 2021;19(1):1–11. 10.1016/S1875-5364(21)60001-8
30- Li S, Zhang B. Traditional Chinese medicine network pharmacology: theory, methodology and application. Chin J Nat Med. 2015;11(2):110–20.
31- Wu WY, Hou JJ, Long HL, et al. TCM-based new drug discovery and development in China. Chin J Nat Med. 2014;12(4):241–50. 10.1016/S1875-5364(14)60050-9
32- Tan YQ, Chen HW, Li J. Astragaloside IV: an effective drug for the treatment of cardiovascular diseases. Drug Des Devel Ther. 2020;14:3731–46. 10.2147/DDDT.S272355
33- Yao C, Zhang J, Li J, et al. Traditional Chinese medicine (TCM) as a source of new anticancer drugs. Nat Prod Rep. 2021;38(9):1618–33. 10.1039/D0NP00057D
34- Liu T, Qin M, Xiong X, et al. Multi-omics approaches for deciphering the complexity of traditional Chinese medicine syndromes in stroke: a systematic review. Front Pharmacol. 2022;13:980650. 10.3389/fphar.2022.980650
35- Yin LW, Gao YC, Li ZP, et al. Analysis of Chinese herbal formulae recommended for COVID-19 in different schemes in China: a data mining approach. Comb Chem High Throughput Screen. 2021;24(7):957–67. 10.2174/1386207323666201001114101
36- Lee WY, Lee Y, Lee S, et al. A machine learning approach for recommending herbal formulae with enhanced interpretability and applicability. Biomolecules 2022;12(11):1604. 10.3390/biom12111604
37- Huang Z, Miao J, Chen J, et al. A traditional Chinese medicine syndrome classification model based on cross-feature generation by convolution neural network: model development and validation. JMIR Med Inform. 2022;10(4):e29290. 10.2196/29290
38- Dong X, Zheng Y, Shu Z, et al. TCMPR: TCM Prescription recommendation based on subnetwork term mapping and deep learning. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Houston, TX: IEEE. 2021; 3776–83.
39- Zhao W, Lu W, Li Z, et al. TCM herbal prescription recommendation model based on multi-graph convolutional network. J Ethnopharmacol. 2022;297:115109. 10.1016/j.jep.2022.115109
40- Li W, Yang Z. Exploration on generating traditional Chinese medicine prescriptions from symptoms with an end-to-end approach. In: CF International Conference on Natural Language Processing and Chinese Computing. Springer, Cham. 2019, 486–98.
41- Cheng N, Chen Y, Gao W, et al. An improved deep learning model: S-TextBLCNN for traditional Chinese medicine formula classification. Front Genet. 2021;12:807825. 10.3389/fgene.2021.807825
42- Zhou W, Yang K, Zeng J, et al. FordNet: recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. Pharmacol Res. 2021;173:105752. 10.1016/j.phrs.2021.105752
43- Wang Z, Li L, Song M, et al. Evaluating the traditional chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning. J Ethnopharmacol. 2021;272:113957. 10.1016/j.jep.2021.113957
44- Yao L, Zhang Y, Wei B, et al. A topic modeling approach for traditional Chinese medicine prescriptions. IEEE Trans Knowl Data Eng. 2018;30(6):1007–21. 10.1109/TKDE.2017.2787158
45- Bu D, Xia Y, Zhang JY, et al. FangNet: mining herb hidden knowledge from TCM clinical effective formulas using structure network algorithm. Comput Struct Biotechnol J. 2021;19:62–71. 10.1016/j.csbj.2020.11.036
46- Jin Y, Zhang W, He X, et al. Syndrome-aware herb recommendation with multi-graph convolution network[C]//2020 IEEE 36th International Conference on Data Engineering (ICDE). Dallas, TX: IEEE. 2020; 145–56.
47- Qikai Niu, Hongtao Li, Lin Tong, Sihong Liu, Wenjing Zong, Siqi Zhang, SiWei Tian, Jingai Wang, Jun Liu, Bing Li, Zhong Wang, Huamin Zhang. TCMFP: a novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms. Briefings in Bioinformatics. 2023; 24(3): bbad102. 10.1093/bib/bbad102
48- Yu Y, Kong D. Protein complexes detection based on node local properties and gene expression in PPI weighted networks. *BMC Bioinformatics*. 2022; 23(24). 10.1186/s12859-021-04543-4
49- Yan D, Zheng G, Wang C, Chen Z, Mao T, Gao J, et al. HIT 2.0: an enhanced platform for Herbal Ingredients' Targets. Nucleic Acids Res. 2022; 50:D1238–D1243. 10.1093/nar/gkab1011
50- Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing - interpreting hypernymic propositions in biomedical text. J Biomed Inform 2003;36:462–77. 10.1016/j.jbi.2003. 11.003.
51- Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindflesch TC. SemMedDB: a PubMedscale repository of biomedical semantic predications. Bioinformatics 2012;28: 3158–60. 10.1093/bioinformatics/bts591.
52- Fisher RA. On the Interpretation of χ 2 from contingency tables, and the calculation of P. J Royal Stat Soc 1922;85:87–94. 10.2307/2340521.
53- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447–52. 10.1093/nar/gku1003.
54- Jiang X, Zhu R, Ji P, Li S. Co-Embedding of Nodes and Edges With Graph Neural Networks. IEEE Trans Pattern Anal Mach Intell. 2023 Jun;45(6):7075-7086. 10.1109/TPAMI.2020.3029762.

55- Qiu J, Dong Y, Ma H, Li J, Wang K, Tang J. Network embedding as matrix factorization: unifying DeepWalk, LINE, PTE, and node2vec. Wsdm 2018:459–67. 10.1145/3159652.3159706.
56- Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, et al. Deep learning-based drug–target interaction prediction. J Proteome Res 2017;16:1401–9. 10.1021/acs. jproteome.6b00618.
57- e H, Ye L, Kang H, Zhang D, Tao L, Tang K, et al. HIT: linking herbal active ingredients to targets. Nucleic Acids Res 2011;39:D1055–9. 10.1093/nar/gkq1165.
58- Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 2009;25:2397–403. 10.1093/bioinformatics/btp433
59- van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics 2011;27:3036–43. 10.1093/bioinformatics/btr500
60- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006;34:D354–7. 10.1093/nar/gkj102.
61- Ben-Avraham D, Havlin S. Diffusion and reactions in fractals and disordered systems. Cambridge University Press; 2000. 10.1017/CBO9780511605826.
62- Doyle PG, Snell JL. Random walks and electric networks. Math Assoc Am 1984;22: 595–9. 10.5948/UPO9781614440222.
63- Grover A, Leskovec J. Node2vec: scalable feature learning for networks. Kdd 2016; 2016:855–64. 10.1145/2939672.2939754.