Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

HerbGL: a network propagation and graph regularization-based framework for herb pairs prediction in traditional Chinese medicine

Weixiang Liu[a], Qian Yuan[a], Junjie Zhang[a], Xinliang Sun[b], Kongfa Hu[a, c*], Tao Yang[a*]

a. School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

b. School of Computer Science and Engineering, Central South University, Changsha, Hunan 410083, China

c. Jiangsu Engineering Research Center for Smart Traditional Chinese Medicine Health Services, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

ARTICLE INFO

Article history
Received 23 November 2025
Accepted 23 March 2026
Available online 25 June 2026

Keywords
Traditional Chinese medicine
Herb pair
Network propagation
Graph regularization
Network pharmacology

ABSTRACT

Objective To address the challenges of systematically identifying herb pairs in traditional Chinese medicine (TCM), we proposed HerbGL, a framework for predicting potential herb pairs that integrates network propagation and graph regularization.

Methods Based on the assumption that herbal actions induce subtle perturbations in biological systems, a framework named HerbGL was proposed. Random walk with restart (RWR) was first applied to the protein-protein interaction (PPI) network to reconstruct herb-specific perturbation effects and generate weighted subnetworks. Then, to quantify affinity between herb pairs, two network-proximity metrics, Closeness and PageRank, were computed from the weighted subnetworks to construct herb-pair affinity matrices. Finally, these matrices, together with known herb pairs (derived from co-occurrence analysis of TCM formulas with a threshold determined from the inflection point of the frequency distribution), were incorporated into a graph regularization model to predict potential herb pairs. Model performance was assessed through baseline comparison, ablation and robustness experiment under different ratios of positive and negative samples, using the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), accuracy, and precision as evaluation metrics. Furthermore, the predicted herb pairs were validated through both literature evidence and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses.

Results The weighted subnetworks constructed by RWR provided a refined simulation of herb-specific perturbation effects, which formed the basis for subsequent affinity modeling and prediction. Analysis of herb pair co-occurrence frequencies revealed a marked change around 150, which was selected as the threshold to distinguish herb pairs from non-herb pairs. HerbGL exhibited superior predictive performance compared with baseline models (AUROC = 0.970 5, AUPRC = 0.955 5, accuracy = 0.726 6, precision = 0.970 6). Ablation results showed that removing the Closeness and PageRank metrics substantially degraded performance (AUROC = 0.819 1, AUPRC = 0.876 8), confirming their necessity. Robustness evaluation under an imbalanced positive-to-negative sample ratio of 1 : 5 yielded AUROC = 0.969 6 and AUPRC = 0.840 4, indicating stable predictive ability. Moreover, multiple case studies further validated the rationality of the predicted herb pairs, such as Fangfeng (Saposhnikoviae Radix) and Qingpi (Citri Reticulatae Pericarpium Viride) which are recorded in *Liangpeng Huiji* (《良朋汇集》, *Collection of Excellent Recipes*) Vol. 3: Fangfeng Shengma Tang (防风升麻汤). Additionally, pathway enrichment analysis of the Renshen (Ginseng Radix et Rhizoma)

*Corresponding author: Tao Yang, E-mail: yangtao@njucm.edu.cn. Kongfa Hu, E-mail: kfhu@njucm.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2026.05.009

Citation: LIU WX, YUAN Q, ZHANG JJ, et al. HerbGL: a network propagation and graph regularization-based framework for herb pairs prediction in traditional Chinese medicine. Digital Chinese Medicine, 2026, 9(2): 265-277.

and Lianqiao (Forsythiae Fructus) pair further supported the biological plausibility of their compatibility.

Conclusion HerbGL offers an effective and biologically informed framework for identifying herb pairs in TCM. Beyond improving herb pair prediction, the framework also provides data support for research on herb compounds and mechanisms, thereby supporting data-driven exploration of TCM compatibility.

1 Introduction

Traditional Chinese medicine (TCM) formulas play a crucial role in disease treatment by leveraging the synergistic or complementary effects of multiple herbs to achieve optimal therapeutic outcomes. In particular, herb pairs, the fundamental unit of TCM formulas, have attracted growing attention owing to their straightforward compatibility and notable therapeutic effects. The rationale underlying herb pairs is that their rational combination can produce complementary or synergistic effects while counteracting each other's undesirable properties, serving as the building block for more complex formulas. For example, Huanglian (Coptidis Rhizoma) and Wuzhuyu (Evodiae Fructus) are combined in the formula Zuojin Wan (左金丸), a basic herb pair widely prescribed for gastric diseases [1]. Suxiao Jiuxin Wan (速效救心丸), composed of Chuanxiong (Chuanxiong Rhizoma) and Bingpian (Borneolum Syntheticum), is a commonly used Chinese patent medicine for coronary heart disease and angina pectoris [2]. However, systematically identifying herb pairs remains challenging due to the inherent complexity of herbs originating from their multi-component nature.

Recently, large-scale databases focusing on comprehensive molecular and pharmacological characterizations of herbs are rapidly growing and becoming increasingly accessible. Some examples include the Traditional Chinese Medicine Information Database (TCMID) [3], the Encyclopedia of Traditional Chinese Medicine (ETCM) [4], the Symptom Mapping (SymMap) [5], and the Traditional Chinese Medicine Bank (TCMBank) [6]. The wealth of herb-related data offers great opportunities to elucidate novel mechanisms of herbal actions and devise computational strategies that expedite the integration of TCM well into clinical practice. Several studies have introduced graph-based deep learning models to explore the complex relationships in TCM systems. For example, LIU et al. [7] combined graph convolutional networks with herbal properties to predict the efficacy of TCM formulas. To capture the correlation between TCM formulas and their targets, a Knowledge Graph Relational Path Network for Target Prediction of TCM Prescriptions (KGRN) was developed for this very purpose [8]. Other studies have focused on the application of interpretable artificial intelligence in TCM and treatment-recommendation systems. For instance, LIU et al. [9] offered a unique framework for disease treatment based on traditional herbal theories. JIN et al. [10] leveraged the meta-path guided attention network to provide the explainable herb recommendations, and ZENG et al. [11] presented an interpretable graph artificial intelligence framework to quantify such mechanisms in TCM. Despite these advances in artificial intelligence modeling and treatment-recommendation systems for TCM, most of them focus only on formula-level prediction, herb recommendation, or herb-disease associations, while systematic investigations specifically targeting herb pairs remain limited.

Unlike predicting drug combinations based on a single compound in modern medicine, computational approaches targeted at herb pairs have been designed according to their characteristics as a holistic system with multiple compounds. To address this complexity, recent studies have increasingly adopted network-based pharmacology and network-based medicinal frameworks to investigate the interactions among herbs, compounds, and biological targets [2, 3]. These approaches typically quantify the functional relationships between herbal components and disease-related targets by analyzing their topological proximity in biological networks. Previous studies have suggested that network-based pharmacology models of the drug-target interactions underlying herb combinations may provide novel insights into the mechanism of actions of herb pairs, which are critical for phenotypic-based TCM medication discovery [12, 13]. For example, WANG et al. [14] proposed five distance metrics based on the network proximity method to identify potential herb pairs. GAN et al. [15] established a network-based medicinal framework for TCM by leveraging the human protein-protein interaction (PPI) network to infer novel network relationships among TCM components. However, these approaches mainly rely on the statistical or topological properties of biological networks and do not explicitly model the perturbation effects induced by herbs on the network structure.

To address this limitation, we proposed a network propagation-based graph regularization framework, HerbGL, to identify herb pairs. Based on the assumption that herbal actions may induce subtle perturbations in biological systems, this study aimed to predict potential

herb pairs by leveraging network propagation and graph regularization, combined with known co-occurrence patterns from TCM formulas.

2 Data and methods

Based on the assumption that herbal actions induce subtle perturbations in biological systems, we proposed a framework named HerbGL. As shown in Figure 1. First, we collected herbs and their targets from ETCM 2.0, and the human PPI network was obtained from Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (version 11.5). Based on the known herb-target mappings and the PPI network, connected herb-specific target subnetworks were reconstructed to capture their potential biological effects. And then, a RWR-based network propagation algorithm was applied to model the diffusion of herb-induced propagation in the PPI network, from which network proximity between herbs was quantified based on two centrality measures (Closeness and PageRank). Finally, the proximity scores were converted into a herb-pair affinity matrix and integrated with known herb pairs in a graph regularization framework to predict potential herb pairs.

2.1 Data collection and preprocessing

To comprehensively retrieve herb-related data, we sourced TCM formulas and herb-related targets from the ETCM 2.0 database [16]. This platform provides standardized, comprehensive information on commonly used TCM herbs and formulas, along with their bioactive constituents and molecular targets. Notably, we filtered out TCM formulas containing fewer than two herbs. For the human PPI network, we retrieved interaction networks from the STRING database [17]. Accounting for network noises, we retained only interactions with confidence scores > 0.7.

Inspired by previous study [18], we also derived herb pairs from the collected TCM formulas through pairwise co-occurrence analysis. Based on the assumption that frequently co-occurring herbs are more likely to constitute herb pairs, by visually inspecting the histogram of herb-pair co-occurrence frequencies and identifying a natural cut-off point where the frequency distribution markedly declined, a threshold was determined for distinguishing herb pairs from non-herb pairs, thereby constructing the known herb pairs dataset $\hat{Y}$.

2.2 Herb-target network construction

Since each herb contains multiple compounds, characterizing herbs poses a more challenging task than that of chemical drugs. We therefore employed a multi-step mapping approach that employed targets as a bridge. This strategy allowed us to represent herbs as a herb-specific target network. By doing so, we effectively translated the intricate relationships between herbs and their therapeutic effects into a comprehensible and analyzable framework. Specifically, for each herb i, we collected its curated target proteins from ETCM 2.0 [16] and denoted the set as T_i. We then selected the minimum number of neighbors required for each target t within the PPI network. Specifically, for each herb, we constructed a PPI

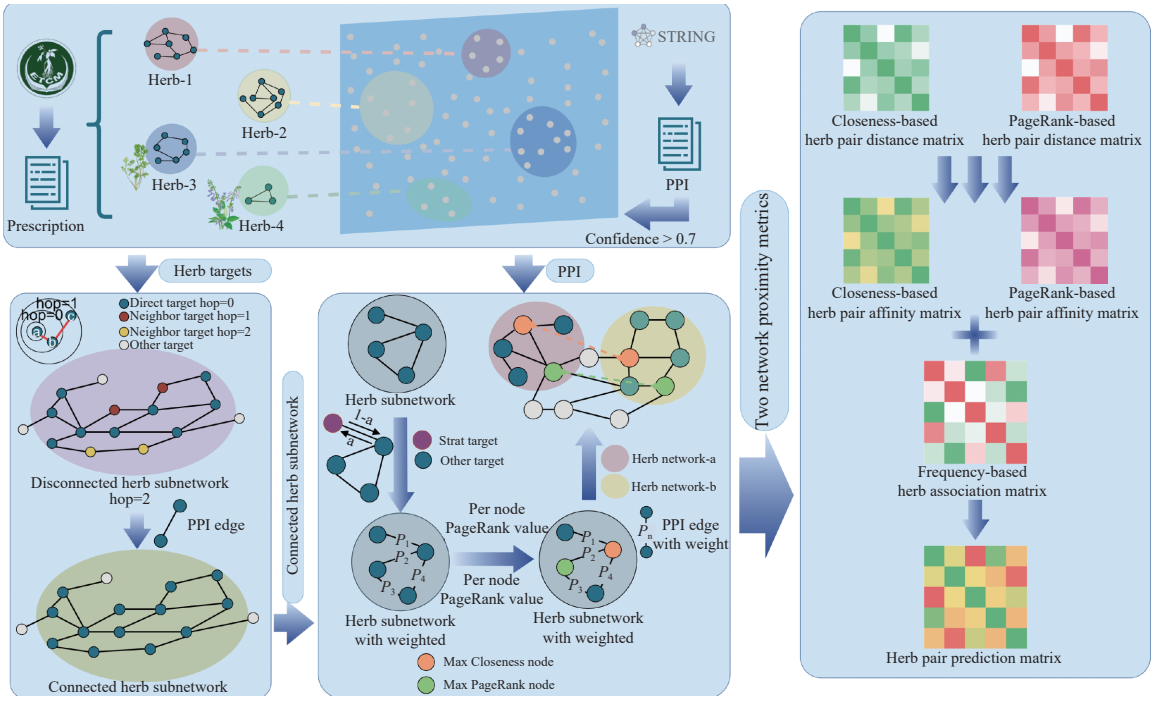

Figure 1 The architecture of HerbGL

subnetwork by taking all nodes within k-hop neighborhoods of its mapped targets. We tested $k \in \{1,2\}$ and found that $k=2$ markedly increased the size of the subnetworks, which not only imposed a significant computational burden but also potentially compromised the accuracy of herb-specific subnetwork representations. For example, the Anxixiang (Benzoinum) subnetwork contained 1 423 targets under $k=2$, compared to only 653 targets under $k=1$. Thus, we selected $k=1$. When the subnetwork was disconnected, we employed the largest connected component as the final herb subgraph. This selection was critical to ensure that the defined target set T_i could form a connected graph, which allowed us to depict herbs with a target graph G_i.

2.3 Network propagation of the herb-target based on RWR

To simulate the propagation of herb perturbation effects on the PPI network, we applied a network propagation method called RWR [19] on the herb-specific target subgraph $G_i=(V_i,E_i)$. Specifically, each target on herb-specific target subgraph G_i was used as the starting seed node, with the assignment of 1 as the starting value. It then propagated through the network, assigning higher weights to targets closely connected to the seed herb. During the propagation process, each step involved a probabilistic choice: randomly moving to a neighboring node or restarting from the seed, as defined by:

$$p^{t+1}=(1-r)Wp^{t}+rp^{0} \tag{1}$$

where r denotes the restarting probability, with the use of an optimal value ($r=0.7$) [19]; $1-r$ denotes the probability of moving to a neighboring node; W, which was defined as W_i, denotes adjacency matrix of the herb-specific target subgraph G_i; specifically, both the rows and columns of W_i correspond to the target proteins in T_i; an entry $W_i(p,q)=1$ indicates a direct PPI interaction between the corresponding targets. p^0 represents initial vector, in which seed target nodes were assigned value 1 and non-seed nodes value 0. In this way, we modeled the perturbation relationships between targets under the intervention of herbs and quantified the importance of targets in herbs. As a result, each herb-specific target subgraph G_i was transformed into a weighted subnetwork, where the network topology remained unchanged and each target node was assigned a steady-state weight given by the RWR score r_i.

2.4 Modeling the relationships between each herb pair based on network-proximity metrics

Subsequently, Closeness and PageRank [20] were applied to quantify herb-pair affinity, exploiting network proximity between their target interactomes. Closeness captures the efficiency of information dissemination for each node, while the PageRank reflects the overall significance of nodes. We denoted the RWR resulting weighted herb subnetwork as (G_i,r_i), where $G_i=(V_i,E_i)$ and r_i is the steady-state probability vector obtained by RWR, Closeness was defined by:

$$C_{\text{close}}(v)=\frac{|V_i|-1}{\sum_{u\in V_i,u\neq v} d_{G_i}(v,u)},\quad v\in V_i \tag{2}$$

in which $v_i^C=\arg\max_{v\in V} C_{\text{close}}(v)$. $d_{G_i}(v,u)$ represents the shortest path length between target v and u in the weighted herb subnetwork G_i. PageRank was defined by:

$$\text{PR}(v)=(1-\alpha)\cdot\frac{1}{|V_i|}+\alpha\sum_{u\in N-(v)}\frac{\text{PR}(u)}{\deg(u)},\quad v\in V_i \tag{3}$$

Initially, $\text{PR}(u)=\frac{1}{n}$ initializes all nodes evenly, α is the damping factor, $N-(v)$ denotes the set of in-neighbors of node v in G_i, $\deg(u)$ represents the out-degree of node u in G_i, V_i denotes the set of target nodes in the herb-specific subnetwork G_i. Subsequently, the targets with the maximum values were identified to characterize the topological position of this herb. And then, two types of distances were then computed as shortest path lengths on the global PPI network. Finally, we used the means of min-max normalization to transform distances to the affinity matrix between each herb pair, set as $S^{\text{Closeness}}$ and S^{PageRank}.

2.5 Predicting the herb pairs based on graph regularization

Some studies have demonstrated that computational methods based on graph regularization can effectively predict diversified relationships among various biological entities [21-23]. Thus, to jointly learn herb synergy and complementarity from both the herb-target subnetwork and pharmacological patterns, we introduced a graph co-regularization approach that effectively extended traditional graph regularization for predicting potential herb pairs. Assuming that the known herb pairs dataset is $\hat{Y}$, the known herb pairs matrix for training is S^{com} which was constructed based on $\hat{Y}$, the Closeness-based affinity matrix is $S^{\text{Closeness}}$, and the PageRank-based affinity matrix is S^{PageRank}, specifically, the Closeness/PageRank affinity matrix that was label-independent was computed solely from the PPI-derived subnetworks. During cross-validation, only training-fold labels in Y, which was initialized based on known herb pairs, were used in the supervised objective, while labels in the test fold were masked. The proposed graph co-regularization simultaneously preserved the geometric structures of the herb-pair network, the herb Closeness-based similarity network, and the herb PageRank-based similarity network through the

following objective function:

$$F_{GC}(Y)=\frac{1}{2}\left(\lambda^{com}\sum_{i,j=1}^{n}L\left(Y_i,Y_j\right)S_{ij}^{com}+\lambda^{Closeness}\sum_{i,j=1}^{n}L\left(Y_i,Y_j\right)S_{ij}^{Closeness}+\lambda^{PageRank}\sum_{i,j=1}^{n}L\left(Y_i,Y_j\right)S_{ij}^{PageRank}\right) \quad (4)$$

where $\lambda^{com}\in(0,1)$, $\lambda^{Closeness}\in(0,1)$, and $\lambda^{PageRank}\in(0,1)$ are three regularization hyperparameters, implying the importance of the herb pair network, herb Closeness-based similarity network, and herb PageRank-based similarity network, respectively. After grid search on the training folds, the final values were selected ($\lambda^{com}=0.1$, $\lambda^{Closeness}=0.3$, $\lambda^{PageRank}=0.3$). Thus, in the objective function, the cost terms $\sum_{i,j=1}^{n}L(Y_i,Y_j)S_{ij}^{com}$, $\sum_{i,j=1}^{n}L(Y_i,Y_j)S_{ij}^{Closeness}$, and $\sum_{i,j=1}^{n}L(Y_i,Y_j)S_{ij}^{PageRank}$ were used to apply herb subnetwork information, herb Closeness-based similarity network information, and herb PageRank-based similarity network information to predict potential herb pairs. Similar to graph regularization, the loss function is defined as:

$$L(F,Y)=\|M\odot(F-Y)\|_F^2 \quad (5)$$

In the above formulation, F and Y denote the predicted and ground-truth herb-pair matrices, respectively; M is a binary mask for training samples, $\odot$ denotes element-wise multiplication, and $\|\cdot\|_F$ is the Frobenius norm. In addition, we applied a normalization technique in graph-based learning to suppress hub effects, which has been widely validated in previous study [24]. Through analytical derivation and algebraic transformation, the final objective function is as follows:

$$F=Y\times(\lambda^{com}\overline{L}^{com}+\lambda^{Closeness}\overline{L}^{Closeness}+\lambda^{PageRank}\overline{L}^{PageRank}+I)^{-1} \quad (6)$$

where I is the identity matrix of size $n\times n$, the L^{com}, $L^{Closeness}$, and $L^{PageRank}$ denote the graph Laplacian matrices constructed from the known herb-pair network, the Closeness-based affinity network, and the PageRank-based affinity network, respectively. The normalized Laplacian is defined as $\overline{\overline{L}}=D^{-1/2}LD^{-1/2}$, where D is the corresponding degree matrix. In this way, their normalized forms, $\overline{L}^{com}$, $\overline{L}^{Closeness}$, and $\overline{L}^{PageRank}$, were obtained to suppress hub effects and ensure numerical stability, and the matrix $\lambda^{com}\overline{L}+\lambda^{Closeness}\overline{L}+\lambda^{PageRank}\overline{L}+I$ is a positive definite and invertible. F represents the scores of predicted herb pairs which we set as the F value.

2.6 Experimental setup and validation framework

2.6.1 Threshold determination for herb pair classification By examining the histogram of herb-pair co-occurrence frequencies, a threshold was established based on an observed inflection point in the frequency distribution, which served to differentiate herb pairs from non-herb pairs. Model generalization was assessed using five-fold cross-validation. Specifically, the positive herb pairs (co-occurrence frequency ⩾ threshold) were randomly divided into five folds. In each round, one fold served as the test set, with its positive labels masked during co-regularization to ensure their invisibility. Each test set contained positive samples and an equal number of negatives randomly selected from pairs with co-occurrence frequency < threshold with the use of a fixed random seed for reproducibility.

2.6.2 Baseline comparison, ablation, and robustness evaluation To evaluate HerbGL's predictive performance, we employed four metrics: the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), accuracy, and precision. Using these metrics, we compared HerbGL against several baselines, including k-nearest neighbors (KNN, $k=3$), logistic regression (L2-regularized), a multi-layer perceptron (MLP, two hidden layers of 64 and 32 neurons, ReLU activation), random forest (10 trees), and gradient boosting. All models received identical feature representations derived from the $S^{Closeness}$ and $S^{PageRank}$ affinity matrices to ensure the fairness of the comparison. Ablation experiments were conducted to assess the contribution of multi-source affinity matrices by comparing HerbGL with and without incorporating both similarity measures. Robustness was further evaluated by varying the positive-to-negative sample ratio from 1 : 1 to 1 : 5, testing model stability under imbalanced conditions.

2.6.3 Pathway level validation based on Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses To validate the biological relevance of predicted herb pairs, a multi-level evaluation strategy was adopted in this study. First, high-scoring candidates were cross-referenced with classical TCM literature and modern pharmacological studies to identify historical or clinical support. And then, GO and KEGG enrichment analyses were performed on the target sets of individual herbs and their combinations. Enrichment analyses were conducted using the gseapy package (Enrichr platform) with the KEGG and GO gene set libraries. An adjusted P value (Benjamini-Hochberg false discovery rate, BH-FDR) threshold of 0.05 was considered statistically significant. For both of GO and KEGG enrichment analyses, we defined synergy as identifying the pathways which were significantly enriched in the combination of herbs A + B, but not in A or B alone, and where the intensity of the combination enrichment was significantly higher than the average single-herb enrichment. Similarly, the complementarity analysis aimed to discern pathways that demonstrated significant enrichment

solely in A or B, yet also in the combination of A + B. This suggested that the combination might integrate the distinct mechanisms of action inherent in single herb, thereby yielding complementary effects.

3 Results

3.1 The co-occurrence analysis of herb pairs

With 46 057 TCM formulas and 1 810 herbs obtained from ETCM 2.0 [16], we counted the co-occurrence frequencies of all herb pairs in the same formulas. As frequently co-occurring combinations are more likely to represent true herb pairs [25, 26], this assumption is further supported by Figure 2A, where the high-frequency combination Renshen (Ginseng Radix et Rhizoma) and Gancao (Glycyrrhizae Radix et Rhizoma) is commonly recognized as a classical herb pair. We also constructed a histogram of co-occurrence frequencies. Figure 2B revealed a marked change around a frequency of 150, which was selected as the threshold to distinguish herb pairs from non-herb pairs, as indicated by the dashed line in the figure. In network pharmacology, TCM network distances are typically classified as either dissociated or intersecting [27]. In Figure 2C, the comparison of shared targets among high-frequency herb pairs showed substantial overlap, supporting their biological relevance. Building on this, we applied HerbGL to compute *F* values, reflecting the likelihood of each combination forming a true herb pair. To validate that *F* values distinguish high-frequency herb pairs from low-frequency ones, we extracted the top 500, 750, 1 000, 1 250, 1 500, and 1 750 herb pairs ranked by co-occurrence frequencies and calculated their mean *F* values. As a baseline, we randomly sampled the same number from negative samples and computed their mean *F* values. The results shown in Figure 2D demonstrated that *F* values effectively discriminated high-frequency herb pairs from those below the 150-frequency threshold.

3.2 Performance comparison of HerbGL with other baseline models

AUROC, AUPRC, precision, and accuracy were used to evaluate the performance of HerbGL. The results found that although the neural network achieved the highest accuracy (0.824 7), HerbGL exhibited an overall optimal performance in predicting potential herb pairs with AUROC = 0.970 5, AUPRC = 0.955 5, and precision = 0.970 6 (Figure 3 and Table 1). We further compared the performance of HerbGL with and without incorporating multi-source affinity matrices. The results showed that HerbGL with multi-source affinity matrices achieved superior

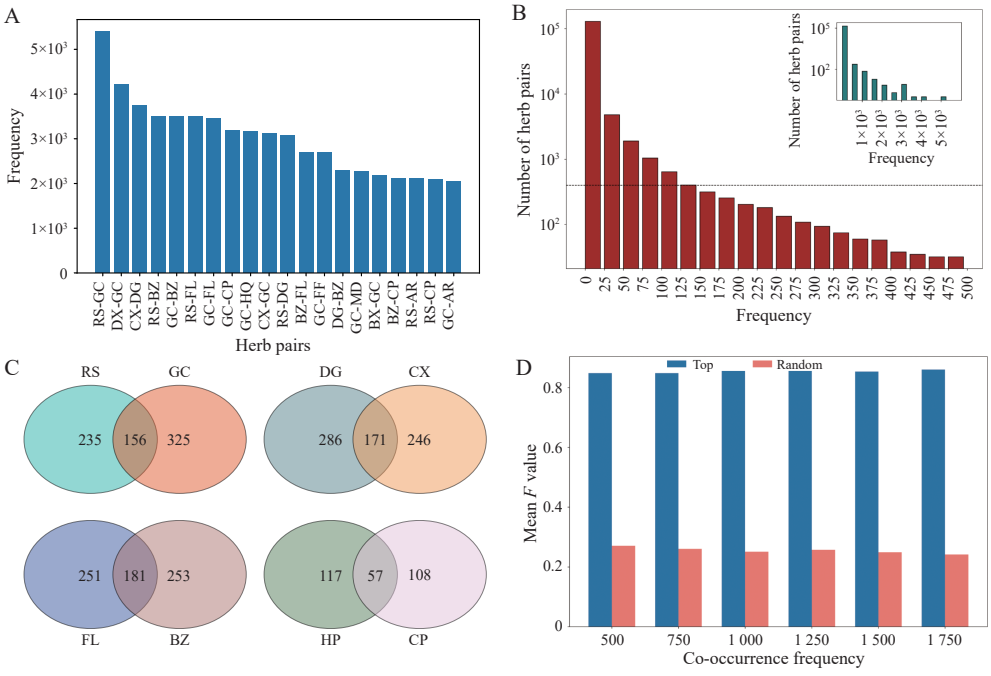

Figure 2 Analyses of herb pairs based on frequencies and targets in co-occurrence relationships

A, the top 20 frequent herb pairs. RS: Renshen (Ginseng Radix et Rhizoma), GC: Gancao (Glycyrrhizae Radix et Rhizoma), DG: Danggui (Angelicae Sinensis Radix), CX: Chuanxiong (Chuanxiong Rhizoma), BZ: Baizhu (Atractylodis Macrocephalae Rhizoma), CP: Chenpi (Citri Reticulatae Pericarpium), HQ: Huangqin (Scutellariae Radix), FF: Fangfeng (Saposhnikoviae Radix), MD: Maidong (Ophiopogonis Radix), BX: Banxia (Pinelliae Rhizoma), AR: Huangqi (Astragali Radix), FL: Fuling (Poria), HP: Houpu (Magnoliae Officinalis Cortex). B, histogram of herb pair co-occurrence frequencies. C, shared targets between herbal combinations of the top four high-frequency herb pairs. D, herb pairs of top frequency and random herb pairs. The baseline were randomly sampled with co-occurrence frequencies below 150.

predictive performance compared with the model without them (AUROC = 0.891 9, AUPRC = 0.876 8). Additionally, robustness experiments were further conducted to evaluate the resistance of HerbGL under varying positive-to-negative sample ratios. Specifically, the proportion of positive-to-negative samples was adjusted from 1 : 1 to 1 : 5. The results demonstrated that HerbGL maintained robust performance even under imbalanced conditions (Table 2), thereby confirming its high stability.

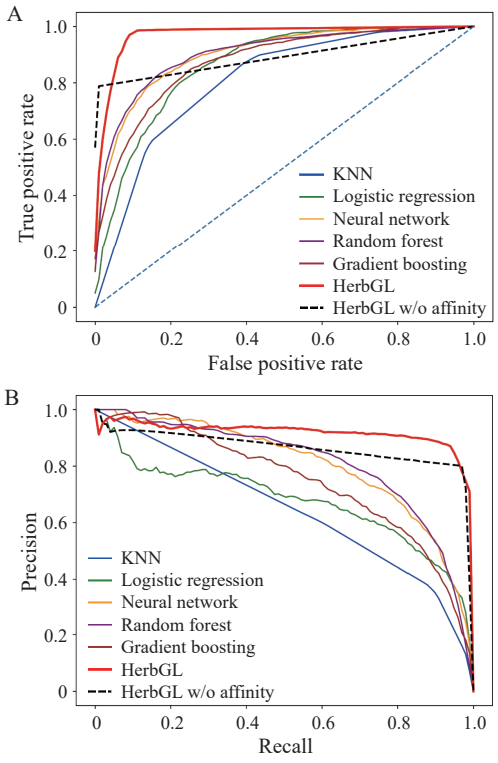

Figure 3 Comparison of HerbGL with baseline models in predicting herb pairs

A, the AUROC results for each model. B, the AUPRC results for each model.

Table 1 Comparison of the performance of HerbGL with other baseline models

Model	AUROC	AUPRC	Accuracy	Precision
HerbGL	0.970 5	0.955 5	0.726 6	0.970 6
HerbGL w/o affinity	0.891 9	0.876 8	0.631 9	0.964 3
KNN	0.805 5	0.741 6	0.739 9	0.683 6
Logistic regression	0.860 2	0.855 9	0.772 0	0.762 7
Neural network	0.905 9	0.907 9	0.824 7	0.812 8
Random forest	0.831 5	0.814 4	0.758 7	0.793 3
Gradient boosting	0.777 8	0.779 1	0.702 7	0.730 6

3.3 Prediction and validation of the potential herb pairs

Based on the F values of each herb pair obtained from HerbGL according to Equations (4) and (5), several potential herb pairs were identified, which were initially input into the model as non-herb pairs. To validate these potential pairs, we screened out the top four potential herb pairs with the highest F values ($F > 0.7$).

3.3.1 Shared targets and literature evidence validation of potential herb pairs First, we calculated the number of shared targets among them. As shown in Table 3, a substantial overlap of targets was observed, suggesting their potential pharmacological relevance.

Then, among the potential herb pairs, three groups of potential herb pairs, Fangfeng (Saposhnikoviae Radix) and Qingpi (Citri Reticulatae Pericarpium Viride), Chuanxiong (Chuanxiong Rhizoma) and Zexie (Alismatis Rhizoma), and Danggui (Angelicae Sinensis Radix) and Baifuzi (Typhonii Rhizoma), caught our attention for their combined use had been reported and documented in previous literature and clinical records. The herb pair of Fangfeng (Saposhnikoviae Radix) and Qingpi (Citri Reticulatae Pericarpium Viride) is recorded in *Liangpeng Huiji* (《良朋汇集》, *Collection of Excellent Recipes*) Vol. 3: Fangfeng Shengma Tang (防风升麻汤) for treating toothache. The herb pair of Chuanxiong (Chuanxiong Rhizoma) and Zexie (Alismatis Rhizoma) has been reported in a study on their combined treatment for hypertension [28]. The herb pair of Danggui (Angelicae Sinensis Radix) and Baifuzi (Typhonii Rhizoma) has precedent application in migraine treatment [29].

Table 2 Robust experimental results of HerbGL

Positive-to-negative ratio	AUROC	AUPRC
1 : 1	0.970 5	0.955 5
1 : 2	0.969 5	0.917 8
1 : 3	0.969 1	0.887 8
1 : 4	0.971 8	0.875 3
1 : 5	0.969 6	0.840 4

Table 3 The number of shared targets between potential herb pairs

Herb 1	Herb 2	Herb 1 target	Herb 2 target	Shared target
Chuanxiong (Chuanxiong Rhizoma)	Zexie (Alismatis Rhizoma)	246	149	124
Renshen (Ginseng Radix et Rhizoma)	Lianqiao (Forsythiae Fructus)	235	236	161
Fangfeng (Saposhnikoviae Radix)	Qingpi (Citri Reticulatae Pericarpium Viride)	205	92	62
Danggui (Angelicae Sinensis Radix)	Baifuzi (Typhonii Rhizoma)	286	123	69

3.3.2 GO and KEGG enrichment analysis of Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus) Of the four candidate herb pairs, Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus) showed the largest number of shared targets (n = 161) and were thus selected as a representative pair for subsequent GO and KEGG pathway enrichment analysis. In the ancient book *Shennong Bencao Jing* (《神农本草经》, *Shennong's Classic of the Materia Medica*), it is recorded that Renshen (Ginseng Radix et Rhizoma) can replenish the vital energy, and mainly treats deficiency and low immunity. Meanwhile, Lianqiao (Forsythiae Fructus) is recorded in *Bencao Gangmu* (《本草纲目》, *Compendium of Materia Medica*) as clearing away heat and removing toxins, as well as treating heat and inflammation. The combination of the two reflects the TCM principle of strengthening the body's resistance to pathogenic factors and eliminating them: Renshen (Ginseng Radix et Rhizoma) reinforces Qi and consolidates the root [Fuzheng (扶正, supporting the righteous)], while Lianqiao (Forsythiae Fructus) clears heat and purges fire [Quxie (驱邪, eliminating the pathogenic heat)]. Together, they regulate the pathological state of Qi deficiency and the excess of pathogenic factors.

Therefore, we can tell from Figure 4 that the two herbs were enriched in metabolic and immune-related pathways, such as protein phosphorylation, which plays an important role in immune cell signaling and the activation of immune cells, is often accompanied by a series of protein phosphorylation events that mediate signal transduction. Furthermore, by observing the results in Figure 5, the two exhibited a high degree of overlap in KEGG pathways, reaching up to 238, and we found that in the complementary KEGG pathways diagram, these two herbs mainly complement each other in "other glycan degradation" "steroid biosynthesis" and "terpenoid backbone biosynthesis", which were found to be mainly involved in substance metabolism and synthesis, affecting the metabolic homeostasis of the organism. Abnormal metabolism may be associated with metabolic diseases (e.g. diabetes and hyperlipidemia). There were also "asthma" "allograft rejection" "primary immunodeficiency" "graft-versus-host disease" and "type I diabetes mellitus pathways", which were still directly associated with immune and inflammatory related diseases.

Through the above analyses, it is easy to find that the results of both synergistic and complementary analyses of KEGG and GO indicate that this herb pair mainly focuses

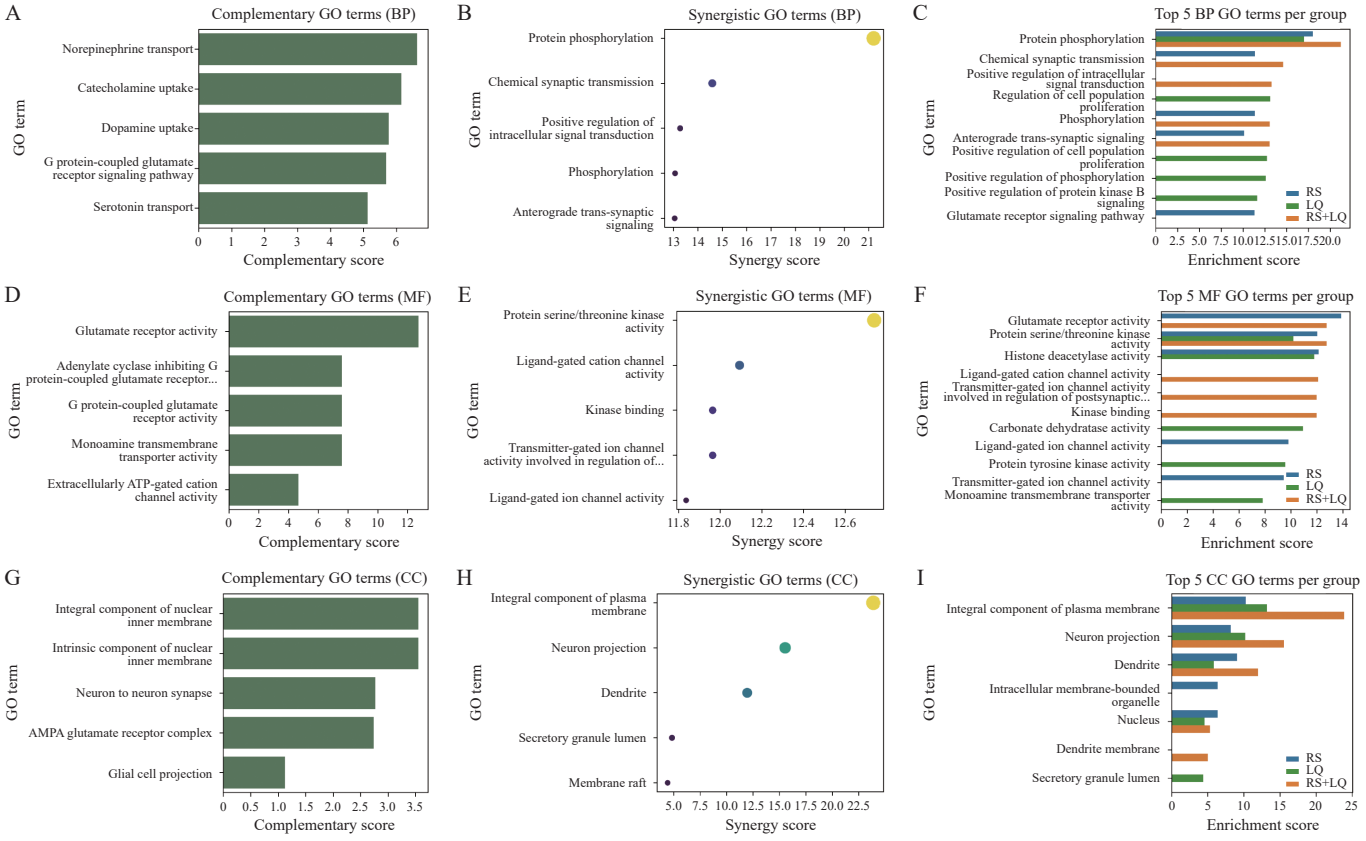

Figure 4 GO enrichment analysis of Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus)

A – C, the top five enriched biological process (BP) terms identified from the complementary, synergistic, and comparison enrichment analyses, respectively. D – F, the top five enriched molecular function (MF) terms identified from the complementary, synergistic, and comparison enrichment analyses, respectively. G – I, the top five enriched cellular component (CC) terms identified from the complementary, synergistic, and comparison enrichment analyses, respectively.

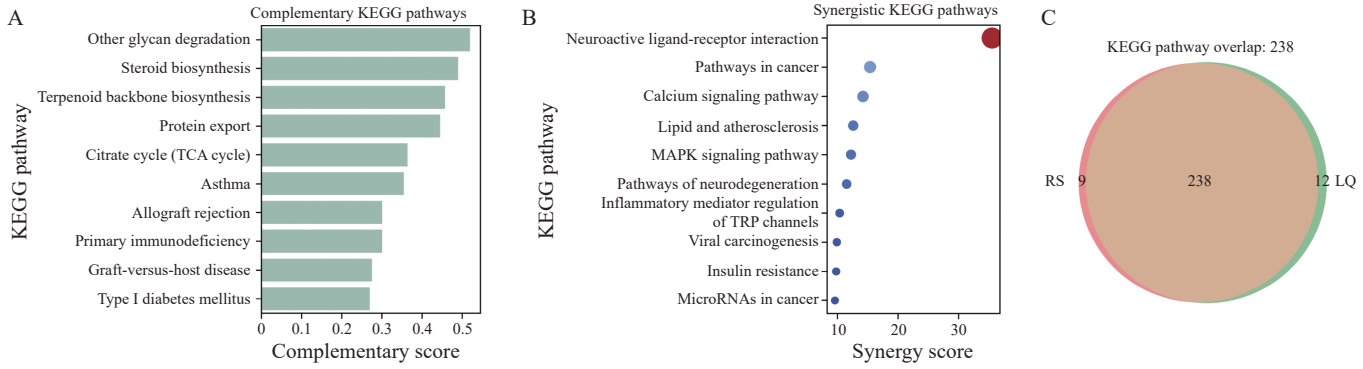

Figure 5 KEGG enrichment analysis of Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus)

A, the top 10 complementary enrichment analysis between Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus). B, the top 10 synergistic enrichment analysis between Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus). C, the overlap KEGG pathway number between Renshen (Ginseng Radix et Rhizoma) and Lianqiao (Forsythiae Fructus).

on immunomodulatory and anti-inflammatory, which is also consistent with TCM classics that writes the combination of the two can clear pathogenic heat and replenish Qi.

4 Discussion

4.1 Mechanistic insights into compound and target interactions underlying Danggui (Angelicae Sinensis Radix) and Baifuzi (Typhonii Rhizoma)

Through HerbGL, weighted herbal subnetworks were constructed to better represent compound-mediated effects on potential targets, motivating a supplementary analysis. Therefore, Danggui (Angelicae Sinensis Radix) and Baifuzi (Typhonii Rhizoma) were selected. Target screening was performed using a network centrality-driven strategy on RWR-derived weighted subnetworks, where nodes with the highest Closeness and PageRank were selected as seed nodes, and only the top 5% strongest edge-weighted interactions with their first-order neighbors were retained. Screening results showed that several key targets were consistent with previously reported compound and target associations in both herbs, supporting the reliability of the network framework.

As shown in Figure 6, for Danggui (Angelicae Sinensis Radix), ligustilide has been reported to activate EGFR via GPR30, promoting bone formation and exhibiting antioxidant and anti-apoptotic activities [30]. It has also been investigated in combination with EGFR for osteoporosis treatment and bone protection, and shown to possess anti-inflammatory, neuroprotective, and anti-cancer effects [31]. Phenolic acids are among its major active components, where ferulic acid has been reported to bind EGFR and TP53, thereby affecting the PI3K/AKT signaling pathway, while caffeic acid exerts anti-tumor effects through PI3K/AKT and JAK/STAT pathways and also shows anti-inflammatory and metabolic regulatory activities [32]. In addition, polysaccharides exhibit multi-target pharmacological effects, with reported efficacy in anemia, metabolic syndrome, tumors, cardiovascular diseases, and radiation protection [33].

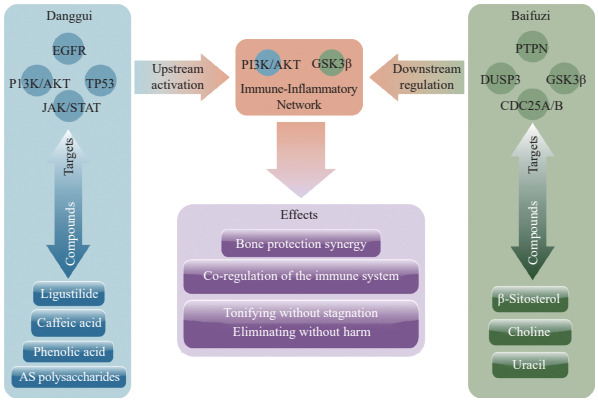

Figure 6 Mechanistic analysis of Danggui (Angelicae Sinensis Radix) and Baifuzi (Typhonii Rhizoma) based on compounds and targets

EGFR, epidermal growth factor receptor. PI3K/AKT, phosphatidylinositol 3-kinase/protein kinase B. TP53, tumor protein p53. JAK/STAT, janus kinase/signal transducer and activator of transcription. GSK3β, glycogen synthase kinase 3 beta. PTPN, protein tyrosine phosphatase, non-receptor type. DUSP3, dual specificity phosphatase 3. CDC25A/B, cell division cycle 25 A/B.

In contrast, key targets of Baifuzi (Typhonii Rhizoma), including CDC25A, CDC25B, DUSP3, GSK3B, PLA2G1B, and PTPN family members (PTPN1, PTPN2, PTPN6, PTPN7, and PTPRC), are mainly involved in cell cycle regulation, immune signaling, and inflammatory responses [34, 35]. These targets are partially associated with bioactive compounds such as β-sitosterol and choline [36]. In particular, β-sitosterol has been reported to regulate PI3K/AKT/GSK3B signaling and exert anti-inflammatory and immunomodulatory effects, suggesting its involvement in GSK3B- and PTPN-related pathways [37]. Choline and uracil are mainly involved in metabolic and cholinergic pathways, and may contribute to the overall

pharmacological effects of Baifuzi (Typhonii Rhizoma) [38].

Overall, Danggui (Angelicae Sinensis Radix) mainly regulates upstream signaling pathways, including EGFR and PI3K/AKT, to promote osteogenesis and exert antioxidant and multi-system regulatory effects. In contrast, Baifuzi (Typhonii Rhizoma) is primarily associated with cell cycle and phosphorylation-related processes, thereby influencing inflammatory responses and immune homeostasis. Both herbs converge on PI3K/AKT and GSK3B signaling, indicating a shared mechanistic basis. Importantly, they act at different regulatory levels, forming a complementary pattern characterized by upstream signaling modulation versus downstream immune and inflammatory regulation, which supports their potential synergistic relationship as a herb pair.

4.2 Mechanistic rationale and methodological advantages of HerbGL

Herb pairs constitute the fundamental building blocks of TCM prescriptions, yet their systematic identification remains challenging due to the multi-component and multi-target nature of herbal systems [39]. To address this issue, HerbGL integrates a network propagation-based model with graph regularization to explicitly simulate herb-induced perturbations on the PPI network. Unlike conventional approaches that rely on static topology, co-occurrence statistics, or simple network proximity metrics, HerbGL incorporates RWR to construct weighted herb-specific subnetworks, enabling node importance to reflect both direct and indirect regulatory effects. This propagation-aware representation captures dynamic network behavior and provides a mechanistic perspective linking molecular interactions with herbal compatibility principles. Furthermore, by integrating multiple affinity matrices through graph co-regularization, HerbGL leverages complementary topological information and achieves superior predictive performance compared with traditional similarity-based and machine learning models, as reflected by improved AUROC and AUPRC values.

4.3 Multi-level validation of herb pair compatibility

Beyond predictive accuracy, the results in this study highlighted the interpretability of HerbGL in relation to established TCM knowledge. Several high-scoring herb pairs identified by the model were supported by TCM classics and modern pharmacological studies [28, 29], suggesting that HerbGL reflected the long-recognized compatibility rules. Enrichment analyses further revealed that predicted herb combinations tended to converge on immune-regulatory and metabolic pathways, providing a biological basis for their synergistic or complementary effects. These findings indicated that the proposed HerbGL not only predicted plausible herb pairs but also facilitated the generation of hypothesis regarding their underlying mechanisms of action. In addition, the propagation-based weighting of herb-pair subnetworks enabled systematic exploration at both the compounds and target levels. This multi-level consistency provides further support for the rationality of herb compatibility and highlights the potential of HerbGL to uncover the molecular mechanisms underlying herb pair formation.

4.4 Limitations and future prospects

Despite its advantages, this study has several limitations. For example, the proposed framework relies heavily on the completeness and reliability of current herb-target associations and PPI databases, which may introduce biases and restrict data coverage, especially for less-studied herbs. Future research should focus on integrating more comprehensive and high-quality data sources, such as multi-omics datasets and experimentally validated interactions, to enhance model robustness and biological fidelity. In addition, improving the accuracy of target annotations and expanding the herb-target database will help mitigate data bias. Further extensions of HerbGL to incorporate multi-herb formulations and dosage-dependent effects may provide deeper insights into complex therapeutic mechanisms. These improvements will contribute to a more reliable and mechanistically grounded framework, facilitating the modernization of TCM and enabling more precise, data-driven strategies of herbal formulation.

5 Conclusion

This study proposes HerbGL, a network propagation-based and interpretable framework for identifying herb pairs. The results indicate that HerbGL can effectively uncover compatibility patterns consistent with both molecular evidence and TCM knowledge, with predicted herb pairs supported by enrichment analysis and biological plausibility. In addition, by integrating target-level and component-level information, HerbGL provides mechanistic insights into herb pair compatibility and offers a scalable approach for data-driven herbal formulation. These findings contribute to the modernization of TCM and suggest promising directions for future research, including multi-herb compatibility modeling and clinical applications.

Fundings

National Natural Science Foundation of China (82575255), Open Research Project of Jiangsu Provincial Research Institute of Chinese Medicine Schools (JSZYLP2024060), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX24_2156).

Acknowledgements

We sincerely thank all colleagues and collaborators for their support and insightful discussions during the course of this research. We would also like to express our gratitude to those who contributed directly or indirectly to this work.

Author contributions

Weixiang Liu: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization, and funding acquisition. Qian Yuan: investigation, resources, and data curation. Junjie Zhang: investigation and data curation. Xinliang Sun: conceptualization, methodology, formal analysis, writing – review & editing, supervision, and project administration. Kongfa Hu: writing – review & editing, supervision, project administration, and funding acquisition. Tao Yang: conceptualization, methodology, resources, writing – review & editing, supervision, project administration, and funding acquisition. All authors approved the submission and take responsibility for this manuscript.

Competing interests

The authors declare no conflict of interest.

References

[1] PAN JL, XU YX, SONG HY, et al. Extracts of Zuojin Wan, a traditional Chinese medicine, phenocopies 5-HTR1D antagonist in attenuating Wnt/β-catenin signaling in colorectal cancer cells. BMC Complementary and Alternative Medicine, 2017, 17(1): 506.

[2] HU XY, LIU YY, FRANCIS OB, et al. Efficacy, safety, and pharmacological mechanism of Suxiao Jiuxin Pill in acute coronary syndrome: a systematic review and network pharmacology analysis. European Journal of Integrative Medicine, 2025, 78: 102524.

[3] XUE RC, FANG Z, ZHANG MX, et al. TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Research, 2012, 41(D1): D1089–D1095.

[4] XU HY, ZHANG YQ, LIU ZM, et al. ETCM: an encyclopaedia of traditional Chinese medicine. Nucleic Acids Research, 2019, 47(D1): D976–D982.

[5] WU Y, ZHANG FL, YANG K, et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. Nucleic Acids Research, 2019, 47(D1): D1110–D1117.

[6] LV QJ, CHEN GX, HE HH, et al. TCMBank: bridges between the largest herbal medicines, chemical ingredients, target proteins, and associated diseases with intelligence text mining. Chemical Science, 2023, 14(39): 10684–10701.

[7] LIU JJ, HUANG QF, YANG XY, et al. HPE-GCN: predicting efficacy of tonic formulae via graph convolutional networks integrating traditionally defined herbal properties. Methods, 2022, 204: 101–109.

[8] GONG Z, ZHANG NX, HE JY. KGRN: knowledge graph relational path network for target prediction of TCM prescriptions. Intelligent Computing Theories and Application. Cham: Springer International Publishing, 2021: 148–161.

[9] LIU R, YAO JQ, SUI YH, et al. Understanding the molecular basis of herbal medicines for cough variant asthma under the guidance of traditional herbal theories. Frontiers in Pharmacology, 2025, 16: 1594308.

[10] JIN YY, JI WD, SHI Y, et al. Meta-path guided graph attention network for explainable herb recommendation. Health Information Science and Systems, 2023, 11(1): 5.

[11] ZENG JQ, JIA XB. Quantifying compatibility mechanisms in traditional Chinese medicine with interpretable graph neural networks. Journal of Pharmaceutical Analysis, 2025, 15(8): 101342.

[12] KEISER MJ, SETOLA V, IRWIN JJ, et al. Predicting new molecular targets for known drugs. Nature, 2009, 462(7270): 175–181.

[13] MOHD FAUZI F, KOUTSOUKAS A, LOWE R, et al. Chemogenomics approaches to rationalizing the mode-of-action of traditional Chinese and ayurvedic medicines. Journal of Chemical Information and Modeling, 2013, 53(3): 661–673.

[14] WANG YY, YANG HB, CHEN LX, et al. Network-based modeling of herb combinations in traditional Chinese medicine. Briefings in Bioinformatics, 2021, 22(5): bbab106.

[15] GAN X, SHU ZX, WANG XY, et al. Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine. Science Advances, 2023, 9(43): eadh0215.

[16] ZHANG YQ, LI X, SHI YL, et al. ETCM v2.0: an update with comprehensive resource and rich annotations for traditional Chinese medicine. Acta Pharmaceutica Sinica B, 2023, 13(6): 2559–2571.

[17] SZKLARCZYK D, KIRSCH R, KOUTROULI M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research, 2023, 51(D1): D638–D646.

[18] WU M, MA TJ, ZHU Y, et al. The evolution of traditional Chinese medicine as recombinant inventions. Proceedings of the National Academy of Sciences of the United States of America, 2024, 121(46): e2400812121.

[19] TONG HH, FALOUTSOS C, PAN JY. Random walk with restart: fast solutions and applications. Knowledge and Information Systems, 2008, 14(3): 327–346.

[20] SAXENA A, IYENGAR S. Centrality measures in complex networks: a survey. arXiv, 2020. doi: 10.48550/arXiv.2011.07190.

[21] LUO JW, DING PJ, LIANG C, et al. Semi-supervised prediction of human miRNA-disease association based on graph regularization framework in heterogeneous networks. Neurocomputing, 2018, 294: 29–38.

[22] WANG F, HUANG ZA, CHEN X, et al. LRLSHMDA: Laplacian regularized least squares for human microbe-disease association prediction. Scientific Reports, 2017, 7: 7601.

[23] XIAO Q, LUO JW, LIANG C, et al. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. Bioinformatics, 2018, 34(2): 239–248.

[24] JI M, SUN YZ, DANILEVSKY M, et al. Graph regularized transductive classification on heterogeneous information networks. Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer, 2010: 570–586.

[25] CHEN CW, TSAI CF, TSAI YH, et al. Association rule mining for the ordered placement of traditional Chinese medicine containers: an experimental study. Medicine, 2020, 99(18): e20090.

[26] LEEM J, JUNG W, KIM Y, et al. Exploring the combination and modular characteristics of herbs for alopecia treatment in traditional Chinese medicine: an association rule mining and network analysis study. BMC Complementary and Alternative Medicine, 2018, 18(1): 204.

[27] CHENG FX, KOVÁCS IA, BARABÁSI AL. Network-based prediction of drug combinations. Nature Communications, 2019, 10: 1197.

[28] ZHU Z. Clinical efficacy of herbal patch combined with Chuanxiong Zexie Decoction in the treatment of hypertension. China Journal of Health Care and Medicine, 2018, 30(15): 85–86.

[29] XING PY. A traditional Chinese medicine for treating headache: China, CN101780207B. 2011–08–10.

[30] YANG F, LIN ZW, HUANG TY, et al. Ligustilide, a major bioactive component of *Angelica sinensis*, promotes bone formation via the GPR30/EGFR pathway. Scientific Reports, 2019, 9: 6991.

[31] LI XX, YANG P, ZHANG YH, et al. Research progress on mechanisms of ligustilide in treatment of nervous system diseases. China Journal of Chinese Materia Medica, 2024, 49(9): 2290–2298.

[32] LI ZY, JIA CC, ZHOU YQ, et al. Efficacy and mechanisms of *Angelica sinensis* in treating endometrial cancer: an integrated study. Discover Oncology, 2025, 16(1): 904.

[33] REN CZ, LUO YL, LI XJ, et al. Pharmacological action of *Angelica sinensis* polysaccharides: a review. Frontiers in Pharmacology, 2025, 15: 1510976.

[34] TANG XL, QI CM, ZHOU HH, et al. Critical roles of PTPN family members regulated by non-coding RNAs in tumorigenesis and immunotherapy. Frontiers in Oncology, 2022, 12: 972906.

[35] KUANG WB, WANG X, DING JY, et al. PTPN2, a key predictor of prognosis for pancreatic adenocarcinoma, significantly regulates cell cycles, apoptosis, and metastasis. Frontiers in Immunology, 2022, 13: 805311.

[36] ADHIMOOLAM K, SURESHBABU A, SMIRNOVA E, et al. β-Sitosterol: dietary sources and role in cancer and diabetes management. Food Science & Nutrition, 2024, 12(11): 8870–8886.

[37] CAO Z, WANG X, LV L, et al. β-Sitosterol and gemcitabine exhibit synergistic anti-pancreatic cancer activity by modulating apoptosis and inhibiting epithelial-mesenchymal transition by deactivating Akt/GSK-3β signaling. Frontiers in Pharmacology, 2019, 9: 1525.

[38] BAUMEL BS, DORAISWAMY PM, SABBAGH M, et al. Potential neuroregenerative and neuroprotective effects of uridine/choline-enriched multinutrient dietary intervention for mild cognitive impairment: a narrative review. Neurology and Therapy, 2021, 10(1): 43–60.

[39] ZHOU W, WANG JN, WU ZY, et al. Systems pharmacology exploration of botanic drug pairs reveals the mechanism for treating different diseases. Scientific Reports, 2016, 6: 36985.

(Editor-in-Charge Jie Deng)

HerbGL：一种基于网络传播与图正则化的中药药对预测框架

刘为相[a], 袁倩[a], 张峻杰[a], 孙鑫亮[b], 胡孔法[a,c*], 杨涛[a*]

a. 南京中医药大学人工智能与信息技术学院, 江苏 南京 210023, 中国
b. 中南大学计算机学院, 湖南 长沙 410083, 中国
c. 南京中医药大学江苏省智慧中医药健康服务工程研究中心, 江苏 南京 210023, 中国

【摘要】目的 为解决传统中医药系统性方法识别配伍药对的问题，本文提出一种融合网络传播与图正则化的潜在药对预测框架 HerbGL。**方法** 基于“中药的药理作用会在生物系统中引发细微扰动”的假设，本文提出 HerbGL 框架。首先，在蛋白质互作（PPI）网络上应用重启随机游走（RWR）算法，以重构各中药多靶点蛋白的特异性扰动并生成中药加权子网络。随后，为量化药对间的亲和度，计算每个中药加权子网络上的两种网络邻近度指标（Closeness 和 PageRank），从而构建药对亲和矩阵。最后，将其与已知药对信息（基于中药方剂中药对共现频次分析，并通过频率分布拐点确定阈值筛选获得）一起整合入图正则化模型，实现潜在药对的预测。同时，通过基线模型对比、消融实验及不同正负样本比例下的鲁棒性实验对模型性能进行系统评估，评价指标包括受试者工作特征曲线下面积（AUROC）、精确度-召回率曲线下面积（AUPRC）、准确率和精确率。**结果** RWR 构建的加权子网络能够更细致地模拟中药特异性扰动效应，为后续的亲和度建模与预测提供了基础表示。药对共现频次的分析结果，显示在频次为 150 处，药对数量发生显著变化，因此该值被选作区分药对与非药对的阈值。此外，HerbGL 的预测性能优于所有对比的基线模型（AUROC = 0.970 5，AUPRC = 0.955 5，准确率 = 0.726 6，精确率 = 0.970 6）。消融实验结果表明，去除 Closeness 与 PageRank 指标会显著降低模型性能（AUROC = 0.819 1，AUPRC = 0.876 8），这验证了两类网络特征的必要性。鲁棒性实验表明，即使在正负样本比例为 1∶5 的情况下，模型仍保持稳定表现（AUROC = 0.969 6，AUPRC = 0.840 4）。此外，多项现有案例亦验证了模型所识别药对的合理性，例如防风与青皮在《良朋汇集》（第三卷：防风升麻汤）中就有合用于治疗牙疼的记载。通路富集分析进一步验证了药对人参和连翘的生物学配伍合理性。**结论** HerbGL 提供了一种有效且具有生物学依据的中药药对识别框架，不仅能提升药对预测能力，还能为中药成分作用及机制研究提供数据支持，从而助力基于数据的中药配伍规律探索。

【关键词】中医药；药对；网络传播；图正则化；网络药理学