



Large language model-driven knowledge graph learning for herb–macromolecule interaction prediction and functional analysis[☆]

Sisi Yuan^a, Huihui Wu^b, Jiapeng Yao^{b,*}, Xinze Lian^b, Linlin Zhuo^{b,*}, Keqin Li^c, Dongsheng Cao^d, Aiping Lu^a, Jianfu Xia^{e,*}

^a School of Chinese Medicine, Hong Kong Baptist University, Hong Kong, SAR, 999077, China

^b School of Data Science and Artificial Intelligence, Wenzhou University of Technology, Wenzhou, 325000, China

^c Department of Computer Science, State University of New York, 12561, New Paltz, New York, USA

^d Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, 410003, China

^e Department of General Surgery, Wenzhou Central Hospital, Affiliated to Wenzhou Medical University, Wenzhou, 325027, China

ARTICLE INFO

Keywords:

Large language model (LLM)

Knowledge graph learning

Macromolecular target function analysis

ABSTRACT

The rapid accumulation of biological macromolecule data has created an urgent need for computational methods capable of characterizing the functional roles of macromolecular targets in complex therapeutic systems. Traditional Chinese medicine represents a typical scenario, in which multiple herbal components may coordinately regulate diverse protein targets through multi-component and multi-target mechanisms. However, existing methods often fail to jointly integrate traditional medicine knowledge, modern biomedical associations, and the functional context of biological targets, limiting systematic identification and mechanistic interpretation of herb–macromolecular target interactions (HTIs). Here, we propose LLM-HTI, a large language model (LLM)-driven graph learning framework with adaptive parameter updating based on a Traditional Chinese and Western medicine knowledge graph (TWKG), for predicting potential interactions between herbs and biological macromolecular targets. LLM-HTI extracts latent semantic representations of herb–target associations using LLMs and integrates them with knowledge graph topology, enabling joint characterization of herbs, macromolecular targets, and their functional relationships. The framework further introduces a gradient-guided adaptive parameter updating mechanism to dynamically select important parameter subsets, improving training stability and optimization efficiency. In addition, gastrointestinal surgeons independently assessed the top-ranked candidate targets for disease relevance, biological plausibility, clinical actionability, and translational priority to provide a clinician-guided evaluation of their medical relevance. These benchmark and case-study results demonstrate that LLM-HTI supports macromolecular target function analysis and herb–target mechanism discovery, while clinician-guided assessment further highlights medically relevant candidates for translational follow-up. Our data and code are publicly available at <https://github.com/sisyuan/LLM-HTI>.

1. Introduction

Biological macromolecules, particularly proteins and RNAs, constitute major classes of therapeutic targets and play central roles in disease intervention. Protein targets remain the dominant basis of approved drug action, while previously “undruggable” proteins are increasingly being explored through emerging targeting strategies [1,2]. In parallel, RNAs, including mRNAs and non-coding RNAs, have gained increasing attention as druggable targets and therapeutic intervention points [3].

Herbal medicines have been used for millennia in traditional medical systems and still hold important therapeutic value in contemporary medical practice [4]. Unlike modern drugs, which typically follow a “single compound–single target” paradigm, traditional Chinese medicine is characterized by multi-component, multi-target, and multi-pathway coordination, with therapeutic effects often arising from the collective modulation of diverse biological macromolecular targets across distinct biological processes [5,6]. This mode of action provides a biological basis for the use of single herbs in multiple diseases and

[☆] This article is part of a Special issue entitled: ‘Deep learning’ published in International Journal of Biological Macromolecules.

* Corresponding authors.

E-mail addresses: yaojiapeng1983@163.com (J. Yao), 20210339@wzut.edu.cn (L. Zhuo), xia189687@163.com (J. Xia).

<https://doi.org/10.1016/j.ijbiomac.2026.154371>

Received 19 June 2026; Received in revised form 22 August 2026; Accepted 4 September 2026

Available online 7 September 2026

0141-8130/© 2026 Published by Elsevier B.V.

highlights the potential of traditional medicine in the management of chronic and complex disorders [4]. With continued advances in this field, the clinical application of traditional Chinese medicine has expanded from China to broader global settings [7]. Previous studies have further shown that many classical herbs exert clear pharmacological effects in disease intervention. For example, Astragalus has been associated with immunomodulatory and hepatoprotective activities, whereas Angelica sinensis has been extensively studied for its anti-inflammatory and hematopoietic properties [8]. These observations suggest that systematic modeling of herb–macromolecular target associations is important for understanding the functional roles of biological targets in traditional medicine-mediated therapeutic processes.

The high chemical complexity of herbs and the broad diversity of their targets make systematic dissection of herb–target mechanisms inherently challenging. Traditional target discovery has relied largely on wet-lab experiments and accumulated clinical experience; however, because most herbs contain dozens to thousands of chemical constituents, experimental identification requires not only distinguishing active compounds but also resolving their corresponding biological targets, substantially increasing both cost and complexity [9]. Clinical observation-based approaches are further constrained by subjectivity, high cost, and limited scalability, making them insufficient for large-scale and systematic investigation of HTIs [10,11]. As a result, under conditions of high compositional complexity and target heterogeneity, traditional experimental strategies alone are no longer adequate for efficient mechanistic analysis [12]. Importantly, the long-term accumulation of clinical and pharmacological data has enabled the development of high-quality databases. Resources such as SymMap integrate symptoms, diseases, herbs, and molecular mechanisms, thereby linking traditional Chinese medicine with modern medicine at the phenotypic level [13]. The continued expansion of such resources provides an essential foundation for systematic herb–target studies and the development of computational approaches.

The rapid development of bioinformatics methods and related technologies has driven growing interest in computational approaches for uncovering unknown HTIs. For example, HTP exploits information propagation in protein–protein interaction networks to rank potential targets, whereas heNetRW further introduces random walks on heterogeneous herb–target networks to improve predictive performance [14]. More recently, representation-learning approaches have been used to integrate multi-source heterogeneous information. HTINet constructs a heterogeneous network by incorporating symptoms, herbs, and targets, and learns low-dimensional embeddings for potential target prediction [15]. Although these methods have laid an important foundation for HTI prediction, their ability to capture complex semantic information and higher-order relational structure remains limited.

Among these computational approaches, deep learning has emerged as a major strategy for modeling drug–target interactions owing to its powerful representation-learning capability. With the rise of deep learning, an increasing number of approaches have been developed for drug–target and herb–target prediction. Some methods integrate graph neural networks, network embeddings, and multi-view feature learning to jointly model drug–target interaction patterns, whereas others combine network embedding with ranking learning for drug repositioning and indication discovery [16,17]. Meanwhile, global–local modeling, pretrained representations, and ensemble strategies have been introduced to further improve the accuracy and robustness of drug–protein interaction prediction [18,19]. In the context of HTI prediction, HTINet2 further combines knowledge graph embedding with a residual-like graph neural network to model and integrate target-related features, thereby improving the prediction of multi-target interactions [20]. Despite these advances, unified modeling of traditional medicine semantics and complex heterogeneous relations remains limited.

Continued advances in database construction and predictive algorithm development have provided an important foundation for accurate HTI prediction and have substantially promoted traditional medicine

research. However, representative HTI modeling methods still face two important limitations. First, existing methods mainly rely on graph topology, while the rich semantic information associated with herbs and targets remains underexploited. Second, multi-source HTI modeling introduces a nontrivial optimization challenge. When semantic and structural features are jointly fused, different parameter subsets may contribute unevenly during training, and conventional uniform gradient updates may propagate noisy or less informative signals throughout the entire model. Moreover, in biomedical AI prediction tasks, computational performance alone is insufficient to establish the medical reliability and potential translational value of model predictions. Whether candidate targets are involved in key pathophysiological processes, are consistent with existing medical and pharmacological evidence, and warrant further functional or translational investigation still requires expert medical judgment. Previous studies on medical LLMs have shown that, even when models achieve favorable automated evaluation scores [21,22], their outputs may still contain factual errors, missing information, or limited clinical relevance. Therefore, clinical expert assessment provides an important complement to automated evaluation of biomedical AI models [23].

Motivated by these limitations, we developed LLM-HTI, an LLM-driven computational framework built on TWKG, to identify potential HTIs and elucidate the multi-component and multi-target mechanisms of traditional medicine. At the representation level, LLM-HTI introduces LLM-generated semantic descriptions of herbs and targets and integrates them with TWKG-based structural embeddings, enabling joint semantic–structural representation learning of herbs and biological macromolecular targets. At the optimization level, LLM-HTI incorporates a gradient-guided adaptive parameter updating mechanism that selectively updates highly informative parameter subsets during training. Additionally, incorporating clinical experts beyond computational prediction and molecular-level validation may improve the reliability of medical interpretation and support prioritization of candidate targets for translational research. The main contributions of this study are as follows:

- (1) This study proposes an LLM-driven knowledge graph learning framework with adaptive parameter updating for predicting herb–macromolecular target interactions and supporting macromolecular target function analysis.
- (2) This study contributes a joint semantic–topological representation strategy that integrates LLM-derived latent semantics with TWKG topology to characterize herbs, biological macromolecular targets, and their functional associations.
- (3) This study develops a gradient-guided adaptive parameter updating strategy that improves training stability and optimization efficiency through dynamic selection of key parameter subsets.
- (4) Clinical expert-guided structured evaluation is introduced to enhance the medical interpretability and translational value of candidate targets in terms of disease relevance, biological plausibility, clinical actionability, and research priority.

2. Materials and methods

The proposed LLM-HTI is an LLM-driven adaptive graph learning framework based on TWKG for accurate HTI prediction. Compared with current representative methods, LLM-HTI differs mainly in two aspects: representation learning and optimization. First, at the representation-learning level, HTINet2 and related methods primarily rely on knowledge graph embeddings and graph-structure propagation, while the rich semantic information associated with herbs and targets, including efficacy, indications, medicinal properties, molecular functions, signaling pathways, and disease associations, remains insufficiently exploited. LLM-HTI does not simply introduce additional textual features; instead, it extends HTI modeling from topology-dominated relation prediction to

a representation-learning problem in which semantic and structural information are jointly used to characterize herbs and targets. Specifically, LLM-HTI uses LLMs to generate entity-level semantic descriptions, encodes them with BERT, and integrates the resulting semantic representations with TWKG-based structural embeddings to enhance herb–target relationship modeling. Second, at the optimization level, LLM-HTI introduces adaptive parameter updating to address the uneven contribution of parameter subsets and the noise introduced by uniform gradient updates during heterogeneous semantic–structural feature fusion. By dynamically selecting and updating important parameter subsets according to gradient importance, this mechanism improves training stability and optimization efficiency under multi-source information integration. In addition, this study incorporates clinical expert assessment for medical interpretation and translational evaluation. Representative top-ranked candidate targets were systematically evaluated in terms of gastrointestinal disease relevance, biological plausibility, clinical actionability in modern medicine, and priority for further investigation, thereby enhancing the medical interpretability and potential translational value of the model predictions.

2.1. Data preparation

The HTI data used in this study were obtained from SymMap [24]. Following previous work [20], we extracted all annotated HTI pairs and removed low-confidence associations, resulting in a benchmark dataset containing 38,002 reliable HTI pairs across 563 herbs and 2106 target molecules. We further adopted TWKG, a multi-source knowledge graph constructed by integrating SymMap, soFDA, STRING v11, KEGG and Gene Ontology (GO), which combines traditional medicine theory with modern biomedical knowledge for HTI modeling [25]. TWKG contains 15 entity types, 74,529 nodes, and 31 semantic relation types, yielding 192,415 high-quality triplets.

2.2. Model overview

The proposed LLM-HTI framework consists of three tightly integrated components as shown in Fig. 1. First, a TWKG is constructed from multiple databases, and structured prompts derived from node information are fed into an LLM to generate textual descriptions; these descriptions are then encoded by BERT and integrated with LINE

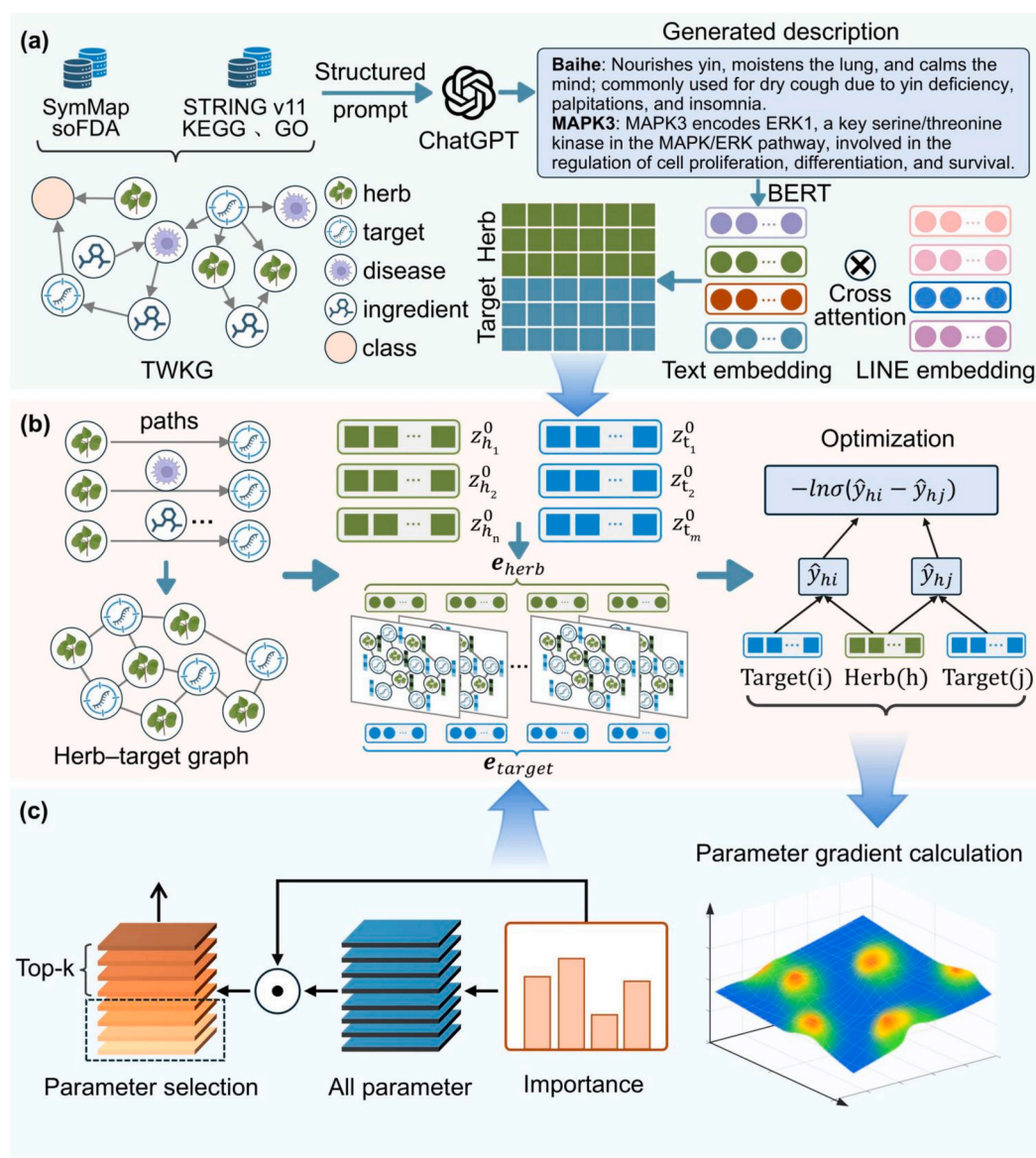


Fig. 1. The framework comprises three modules: (a) joint semantic–structural representation learning based on TWKG and a large language model; (b) herb–target graph representation learning using multi-layer GNNs; and (c) a gradient-guided adaptive parameter updating mechanism.

embeddings learned from TWKG to obtain the initial representations of herbs and targets (Fig. 1a). Second, a herb–target graph is built, upon which a multi-layer GNN is applied to learn node representations and capture potential interaction patterns (Fig. 1b). In this stage, heterogeneous associations among entities such as symptoms, diseases, and pathways are not explicitly introduced. Instead, these richer cross-entity relationships are indirectly incorporated through the structural embeddings learned from TWKG by LINE, which are used as part of the initial node features for subsequent graph representation learning. Third, an adaptive parameter updating mechanism is introduced to estimate parameter importance from gradients and dynamically update the most informative parameter subsets, thereby improving training stability and optimization efficiency (Fig. 1c).

2.3. TWKG-based knowledge embedding

We applied LINE to the TWKG to learn low-dimensional structural representations of graph nodes. As an efficient and scalable network embedding model, LINE maps entities such as herbs, targets, diseases, and symptoms into a continuous vector space by jointly preserving first-order and second-order proximities, thereby capturing both local connectivity and neighborhood structural characteristics. Specifically, first-order proximity reflects associations between directly connected nodes, whereas second-order proximity characterizes similarity in their neighborhood distributions. Based on this framework, a structural embedding vector h^{TWKG} was learned for each herb and target to represent its topological position and association patterns in the graph. The resulting embeddings were then standardized to reduce scale differences across dimensions and used as initial node features for the downstream graph neural network.

2.4. LLM-driven semantic embedding

LLMs can extract high-level semantic information from biomedical literature and databases, and have shown strong potential in tasks such as drug discovery, drug design, and disease association analysis [26–28]. In this study, an LLM was used to generate entity-level textual descriptions for each node in the herb–target graph, thereby enriching the biomedical semantics of herb and target node representations. Each entity description was generated using the node name and entity type, such as herb or target, as input through a structured prompt. The prompt template consists of three components: role definition, task constraints, and output constraints. The role definition instructs the model to act as an expert in traditional Chinese medicine or biomedicine. The task constraints restrict the generated content to entity type-specific information; for example, herb nodes focus on efficacy, indications, medicinal properties, and clinical applications, whereas target nodes focus on molecular functions, signaling pathways, and disease associations. The output constraints standardize the length and style of the generated descriptions. This structured design reduces irrelevant or redundant content and improves the consistency and comparability of semantic descriptions across different nodes.

For example, the LLM can generate the following descriptions:

“*Stemona japonica*”: “Relieves cough and directs lung qi downward, and kills parasites and lice. It is used for acute and chronic cough, pulmonary tuberculosis-related cough, pertussis, enterobiasis, and head and body lice.”

“*Lilium brownii*”: “Nourishes yin, moistens the lung, and calms the mind. It is used for dry cough due to yin deficiency, hemoptysis associated with consumptive disease, palpitations, insomnia, dream-disturbed sleep, and mental restlessness.”

“MAPK3”: “The MAPK3 gene encodes ERK1 (extracellular signal-regulated kinase 1), a core serine/threonine kinase in the MAPK/ERK pathway. Upon activation by the RAS–RAF–MEK cascade, it regulates cell proliferation, differentiation, and survival. MAPK3 is functionally redundant with ERK2 (MAPK1) and plays key roles in development and

cancer.”

The generated text effectively captures the core semantic characteristics of different entity types. For herb nodes, the descriptions mainly emphasize efficacy, indications, and clinical use, as illustrated by the cough-relieving effect of *Stemona japonica* and the calming properties of *Lilium brownii*. In contrast, for target nodes, the generated content focuses more on molecular functions, signaling pathways, and biological roles, exemplified by MAPK3 as a key kinase in the MAPK/ERK pathway. Overall, the LLM adaptively produces type-specific semantic representations, highlighting functional and clinical attributes for herbs and mechanistic and pathway-level information for targets, thereby providing more discriminative semantic features for downstream modeling.

In implementation, entity descriptions were generated by calling the OpenAI API using the script *Knowledge_embedding/gpt_LLM.py*. The model used in our experiments was gpt-5.3, with *temperature* = 0.3 and *max_tokens* = 300. All textual descriptions of herbs and targets were generated offline before model training and saved locally. These cached descriptions were then reused throughout downstream training and evaluation, rather than being dynamically regenerated at each training iteration. This caching strategy avoids stochastic variation caused by repeated LLM calls and improves experimental reproducibility. Failed API requests were handled with a retry mechanism. Before BERT encoding, the generated texts were normalized and checked to remove empty outputs, truncated descriptions, or malformed responses.

Subsequently, the textual descriptions of herbs and targets were encoded using the pre-trained language model BERT, which maps natural language into a continuous vector space [29]. Formally, for a node v with textual description T_v , its semantic representation is defined as:

$$h_v^{\text{LLM}} = \text{BERT}(T_v) \quad (1)$$

Then, mean pooling was applied to the token representations from the last hidden layer of BERT to construct sentence-level vectors, rather than using only the [CLS] token representation. This strategy better aggregates the global semantic information of each node description and provides stable input for subsequent joint semantic–structural representation learning. Through this process, the high-level semantic knowledge encoded by the LLM is explicitly introduced into the graph representation learning framework, enhancing the semantic expressiveness of node embeddings.

To further integrate semantic and structural information, we combine the LLM-driven semantic embedding h^{LLM} with the TWKG-based structural embedding h^{TWKG} . Specifically, the semantic vector h^{LLM} is used as the Query, whereas the $\hat{h}^{\text{TWKG}}$ obtained by linearly projecting the structural vector h^{TWKG} are used as the Key and Value, respectively. A multi-head attention mechanism is then employed for feature fusion:

$$h^f = \text{MultiHeadAttn}(h^{\text{LLM}}, \hat{h}^{\text{TWKG}}, \hat{h}^{\text{TWKG}}) \quad (2)$$

where h^f denotes the fused representation. Finally, h^f is nonlinearly transformed by a multilayer perceptron (MLP) and compressed into a unified 64-dimensional vector:

$$h^* = \text{MLP}(h^f) \quad (3)$$

The vector h^* is used as the final joint representation for downstream prediction tasks.

Before the graph representation learning module, herb nodes and target nodes undergo the same representation construction process, as shown in Eqs. (1)–(3). Specifically, for each herb or target node, BERT is first used to encode the LLM-generated textual description to obtain its semantic representation, while LINE is used to learn structural embeddings from TWKG to provide prior topological information. The semantic representation and structural embedding are then fused through

a shared multi-head attention module. Considering the heterogeneity between herb and target nodes in terms of semantic sources and biological attributes, the fused representations are further projected into a unified shared vector space with the same dimensionality using the same MLP. The resulting herb and target representations are then used as the initial node features for GCN.

2.5. Graph representation learning

Graph neural networks are well suited for modeling complex topological structures and have therefore been widely applied to bioinformatics tasks such as biological network analysis and molecular structure representation [20]. Motivated by this, we employ a graph convolutional network (GCN) to capture potential interaction patterns between herbs and targets. Specifically, a herb–target graph $G \leq V, A, X >$ is constructed from known HTI data, where V denotes the node set (including herbs and targets), A is the adjacency matrix encoding HTIs, and X represents the initial node features obtained by integrating LLM-derived semantic representations with TWKG-based structural embeddings. In the adjacency matrix, $A_{ij} = 1$ indicates that herb i interacts with target j , whereas $A_{ij} = 0$ indicates no interaction.

Let E^k denote the embedding matrix at layer k . The message-passing rule on the herb–target graph is defined as:

$$E^{k+1} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} E^k W^k \quad (4)$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops added, I is the identity matrix, $\tilde{D}$ is the degree matrix of $\tilde{A}$, and W^k denotes the learnable weight matrix at layer k . Because deep GCNs often suffer from over-smoothing, we further aggregate node representations across layers to preserve multi-hop neighborhood information and improve representation quality. Specifically, for herb node h and target node t , the final representations are obtained by summing all layer-wise embeddings together with the initial node features:

$$e_h = e_h^0 + e_h^1 + \dots e_h^k \quad (5)$$

$$e_t = e_t^0 + e_t^1 + \dots e_t^k \quad (6)$$

This cross-layer aggregation strategy makes fuller use of hierarchical feature outputs, improves information propagation across layers, and yields more accurate representations of herbs and targets.

2.6. Parameter update mechanism

As shown in Fig. 1a and b, the graph representation learning process of LLM-HTI integrates multiple heterogeneous information sources. Therefore, the overall graph representation learning process involves multiple information-modeling components, including BERT-based semantic encoding, TWKG-based structural embedding, and GCN-based graph propagation and aggregation, and can be regarded as an implicit joint learning process driven by multi-source information. Under this joint learning setting, different information sources may induce inconsistent or even conflicting parameter update directions at different training stages, which may affect training stability and convergence efficiency. Although standard optimizers such as Adam can adaptively adjust learning rates, they still perform dense and uniform gradient updates over all parameters, without explicitly distinguishing parameters that are currently more informative from those that may introduce noise or conflicts.

To address this issue, we design a parameter-adaptive updating mechanism that dynamically selects and reweights implicit subtasks at the parameter level, thereby reducing task interference and improving training stability. Specifically, the mechanism dynamically evaluates the contribution of different parameter subsets and prioritizes gradient updates on the most critical ones, enabling more effective task decoupling

and conflict mitigation while preserving cross-task knowledge sharing. It should be noted that the proposed method does not explicitly construct multiple task branches or define separate task-specific loss functions. Instead, it partitions the model parameters into multiple subsets, each treated as corresponding to an implicit task, and mitigates potential gradient conflicts among implicit tasks during joint learning through adaptive parameter updating.

Let M denote the total parameters of the model. According to task implicit relevance, M is divided into K parameter subsets, where the subset corresponding to the k -th task is denoted by M^k . The partition of parameter subsets is fixed at initialization, whereas the subsets selected and updated during training vary dynamically. A mask vector s_t^k is introduced to determine the parameter subset M_t^k activated during the t -th iteration, denoted by $M_t = \{M_t^k | s_t^k = 1, k \in \{1, 2, \dots, K\}\}$. Accordingly, the model parameters at iteration $t+1$ can be updated as:

$$M_{t+1} = M_t - \alpha s_t \odot \triangle \mathcal{L}(M_t) \quad (7)$$

where α denotes the learning rate, and $\triangle \mathcal{L}(M_t)$ represents the gradient of the loss function with respect to the model parameters. Under this mechanism, updates are applied only to the parameter subsets identified as important, resulting in an optimization process driven by sparse parameter selection.

To reduce the instability in parameter selection caused by fluctuations in single-step gradients, we further introduce a momentum-based mechanism for long-term importance tracking method. Specifically, the parameter importance of k -th task at t -th iteration, denoted by γ_t^k , is calculated as:

$$\gamma_t^k = \eta \gamma_{t-1}^k + (1 - \eta) \hat{\gamma}_t^k \quad (8)$$

where η represents the momentum parameter, and $\hat{\gamma}_t^k = (\triangle \mathcal{L}(M_t))^2$ represents the instantaneous importance estimated from the current gradient. This mechanism helps preserve the importance information of implicit tasks, reduces frequent subtask switching caused by instantaneous noise, and thereby improves training stability. Additionally, an advanced sampling strategy is employed to identify the most informative parameter subsets:

$$M_t = \left\{ M_t^k | q_t^k = \frac{\gamma_t^k}{\sum_i \gamma_t^i}, k \in \text{Top} - \rho(t) \right\} \quad (9)$$

where ρ denotes the predefined parameter update ratio, and q_t^k is the probability of selecting the subset M_t^k at t -th iteration. Finally, the update rule for the model parameters can be defined as:

$$M_t = M_{t-1} - \beta \sum_{k=1}^K s_t^k \odot M_t^k \odot \triangle \mathcal{L}(M_t) \quad (10)$$

When $s_t^k = 1$, the parameters associated with the corresponding implicit task are activated and updated; When $s_t^k = 0$, those parameters are excluded from updating in the current iteration.

2.7. Model training and optimization

The model was optimized using the Bayesian Personalized Ranking loss to enhance discrimination of HTI patterns. Let (h_i, t_j, t_n) denote a training triplet, where h_i is the i -th herb node and t_j and t_n are a related and an unrelated target, respectively. Based on the herb–target embeddings, the scores of positive and negative pairs, $r_{i,j}$ and $r_{i,n}$ are computed using the dot product. The model is trained to increase the score of positive pairs while decreasing that of negative pairs, thereby ranking true interactions higher than non-interactions:

$$\mathcal{L} = \sum_{i=0}^{H-1} \sum_{(h_i, t_j, t_n) \in C_i} -\ln(\sigma(r_{i,j} - r_{i,n})) + \varphi \|\Omega\|^2 \quad (11)$$

where σ denotes the sigmoid function, C_i represents the set of all triplets (related to i -th herb) used for ranking in the training set, Ω denotes the model parameters, φ is the regularization coefficient, and H is the total number of herb nodes.

At each training iteration, the model first computes the BPR ranking loss based on (h_i, t_j, t_n) triplets, and then performs backpropagation to obtain gradients for all trainable parameters. Based on these gradients, the adaptive parameter updating mechanism estimates the importance of different parameter subsets by combining the current gradients with their historical momentum information, and selects the most important subsets according to a predefined update ratio. For low-importance parameter subsets, gradient masks are applied so that only the update signals at the selected parameter positions are retained before the optimizer step.

2.8. Clinician-guided assessment

To further evaluate the medical relevance and potential translational value of the predicted targets, we incorporated clinician-guided assessment in addition to database, literature, and molecular docking validation. Two physicians with clinical experience in gastrointestinal surgery independently evaluated eight representative high-ranking candidate targets predicted for *Coptis chinensis* and *Artemisia annua*. The assessment covered four dimensions: (1) relevance to gastrointestinal diseases, including gastrointestinal tumors, chronic inflammation, mucosal immunity, intestinal barrier function, and metabolism-related disorders; (2) biological plausibility of the candidate targets; (3) clinical actionability in modern medicine, including human disease evidence, drug development potential, clinical trial information, and established biomarkers or therapeutic pathways; and (4) priority for further functional or translational investigation. The two evaluators independently completed the initial assessment and were blinded to the model-predicted rankings during scoring. Candidate targets with substantial disagreement were further adjudicated by a third clinical expert.

To further assess the medical reliability of LLM-generated semantic information, we randomly sampled descriptions of 100 biological macromolecular targets and invited two physicians with clinical medical backgrounds to independently review them. The evaluation covered molecular function accuracy, signaling pathway accuracy, disease-association plausibility, medical relevance, and the presence of unsupported inferences. Information containing clear factual errors or lacking support from authoritative databases and literature was recorded separately.

3. Results

3.1. Experimental settings

Following previous work [20], the 38,002 HTI records were randomly split into training, validation and test sets at a ratio of 8:1:1. Under this unified setting, LLM-HTI was compared with representative baseline methods, including local topology-based link prediction approaches (CN [24], Salton [30], Jaccard, HPI [31], LHN-1 [32], AA [33] and RA [33]), network propagation-based models (heNetRW [15] and Prince [34]), and the deep learning-based HTINet2 [20]. Topology-based methods infer potential HTIs from neighborhood similarity and related local structural features, but are limited in capturing higher-order relationships in complex biological networks. By contrast, heNetRW and Prince further exploit global structural information through network propagation, whereas HTINet2 integrates knowledge graph embedding with a residual-like graph neural network to model multi-relational information and target-specific features. These baselines together provide a comprehensive benchmark for evaluating the effectiveness of LLM-HTI.

We used hit ratio (HR) and normalized discounted cumulative gain

(NDCG) as evaluation metrics. For herb-side predictions, C_h is the set of ground-truth targets and $C_{h@N}$ denotes the set of top- N candidate targets predicted by the model. Then, $HR@N$ is defined as:

$$HR@N = \frac{|C_{h@N} \cap C_h|}{|C_h|} \quad (12)$$

$HR@N$ reflects the ability of the model to identify true targets within the top- N predictions, with higher values indicating better predictive performance. $NDCG@N$ is used to evaluate the ranking quality of the predicted targets in the recommendation list, and is defined as follows:

$$NDCG@N = \frac{DCG@N}{IDCG@N} \quad (13)$$

where $DCG@N$ denotes discounted cumulative gain, and $IDCG@N$ denotes ideal discounted cumulative gain.

3.2. Comparison with baseline methods

Table 1 compares the performance of LLM-HTI with multiple representative baselines under different Top- N metrics. Overall, LLM-HTI consistently outperformed all competing methods across all evaluation metrics, with the clearest advantage at Top-10, where $HR@10$ and $NDCG@10$ reached 0.4669 and 0.3999, respectively. Traditional link prediction methods based on local topology, such as CN and AA, performed reasonably well on some low-order metrics but were limited in overall performance because they could not capture higher-order structural relationships. Network propagation-based models, including heNetRW and PRINCE, benefited from global structural information but still underperformed relative to LLM-HTI, underscoring the importance of multi-source knowledge integration and deep graph modeling for HTI prediction. Although HTINet2 showed strong competitive performance, LLM-HTI achieved further consistent gains across all metrics, likely owing to the combined effects of parameter-adaptive updating and hierarchical multi-hop neighborhood modeling, which improved both heterogeneous information integration and complex relationship modeling.

3.3. Ablation study

To systematically evaluate the contribution of each key component, we conducted a series of ablation experiments by removing individual modules and measuring the resulting changes in HTI prediction performance. Specifically, “w/o GPT” denotes the removal of ChatGPT-derived textual semantic embeddings, “w/o ML” denotes the removal of multi-layer GCN aggregation while retaining only the final-layer output, and “w/o APUM” denotes the removal of the parameter-adaptive updating mechanism. And “w/o TWKG” denotes the removal of the prior structural embeddings learned from TWKG using LINE. In this setting, the model uses only the semantic features obtained by encoding the LLM-generated textual descriptions with BERT as the initial representations of herb and target nodes.

As shown in Table 2, all components contributed positively to overall performance. Among them, removing multi-hop neighborhood modeling (ML) caused the largest drop, with $HR@1$ decreasing from 0.3975 to 0.3075 and $HR@10$ to 0.3413, indicating that higher-order neighborhood information is critical for capturing potential HTI relationships. Removing the parameter-adaptive updating mechanism (APUM) also substantially degraded performance, reducing $HR@1$ to 0.3368, which highlights the importance of dynamic parameter selection for stable and robust information propagation. By comparison, textual semantic features (GPT) and TWKG structural information produced more moderate but still consistent gains across Top- N metrics, suggesting that language semantics and graph topology provide complementary information for node representation learning. Overall, the full model achieved the best performance across all metrics, validating

Table 1

Performance of LLM-HTI and representative baseline methods for HTI prediction.

Models/metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
CN	0.0274	0.0812	0.0799	0.1705	0.1206	0.2131	0.2001	0.2760
Salton	0.0097	0.0294	0.0284	0.0614	0.0438	0.0792	0.0805	0.1117
Jaccard	0.0113	0.0305	0.0335	0.0667	0.053	0.0850	0.0983	0.1255
HPI	0.0003	0.0006	0.003	0.0073	0.0067	0.0162	0.0201	0.0400
LHN-1	0.0003	0.0005	0.0012	0.0015	0.0022	0.0031	0.0062	0.0079
AA	0.0278	0.0831	0.0787	0.1669	0.1209	0.2148	0.1996	0.2756
RA	0.0270	0.0795	0.0763	0.1569	0.1170	0.2118	0.1929	0.2667
heNetRW	0.0080	0.0186	0.0254	0.0436	0.0455	0.0597	0.0915	0.0942
PRINCE	0.0032	0.0065	0.0134	0.0174	0.0269	0.0282	0.0495	0.0422
HTINet2	0.3536	0.3536	0.3448	0.3398	0.3706	0.3500	0.4458	0.3746
LLM-HTI	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999

Table 2

Ablation results for key components of LLM-HTI.

Models/metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
w/o GPT	0.384	0.384	0.3701	0.3612	0.3806	0.3597	0.4573	0.3907
w/o TWKG	0.3849	0.3849	0.3602	0.3552	0.3788	0.3573	0.4564	0.3869
w/o ML	0.3075	0.3075	0.2838	0.2847	0.2908	0.2822	0.3413	0.2943
w/o APUM	0.3368	0.3368	0.3152	0.3165	0.3386	0.3232	0.4117	0.3506
LLM-HTI	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999

the effectiveness of multi-source information fusion, structural modeling, and parameter-adaptive optimization.

3.4. Parameter experiments

To assess the robustness of LLM-HTI, we systematically evaluated three key hyperparameters: the number of GCN layers, the parameter subset update ratio ρ , and the momentum coefficient β . Overall, Tables 3–5 show that the strong performance of LLM-HTI depends not only on multi-hop neighborhood modeling and parameter-adaptive optimization, but also on appropriate hyperparameter settings. Specifically, a 4-layer GCN achieved the best balance between information propagation depth and the risk of over-smoothing; $\beta=0.7$ stabilized parameter updates while suppressing noise; and $\rho=0.3$ allowed the sparse updating strategy to operate most effectively. Collectively, these settings are critical for capturing complex HTI relationships, exploiting higher-order structural information, and maintaining training stability.

More specifically, the number of GCN layers had a pronounced impact on model performance. As the depth increased from 1 to 4 layers, the model consistently improved across Top-K metrics, indicating that progressive aggregation of multi-hop neighborhood information enhances the discriminative power of node representations. However, further increasing the depth led to performance degradation, suggesting over-smoothing in deeper graph convolutional networks. In contrast, the momentum coefficient β and the update ratio ρ mainly affected optimization stability and efficiency. The best performance was obtained at $\beta=0.7$, indicating that a moderate degree of historical gradient memory helps stabilize updates while preserving responsiveness to current task features. For the update ratio, $\rho=0.3$ yielded the optimal result, suggesting that updating only a subset of key parameters is sufficient to retain informative signals while improving computational efficiency;

Table 3

Performance of LLM-HTI under different numbers of GCN encoder layers.

Layers/metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
1	0.3849	0.3849	0.3532	0.3519	0.3628	0.3489	0.4451	0.3780
2	0.3932	0.3932	0.3620	0.3672	0.3856	0.3692	0.4179	0.3556
3	0.3954	0.3954	0.3712	0.3677	0.3864	0.3704	0.4658	0.3981
4	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999
5	0.3745	0.3745	0.3581	0.3563	0.3802	0.3623	0.464	0.3915
6	0.3716	0.3716	0.3384	0.3407	0.3758	0.3561	0.4577	0.3856

both overly large and overly small update ratios impaired performance. Together, these findings highlight the importance of proper hyperparameter selection in fully realizing the advantages of LLM-HTI.

Additionally, we compared the model performance under different embedding dimensions using the main evaluation metrics. As shown in Table 6, when the embedding dimension was below 64, model performance improved as the dimension increased. When the dimension exceeded 64, performance slightly decreased, suggesting that overly large dimensions may introduce redundant structural information or increase the optimization burden. Overall, LLM-HTI showed stable performance within the 16–128 dimensional range and was relatively insensitive to the LINE embedding dimension, indicating that this hyperparameter has a broad feasible range.

3.5. Case analysis

To assess the practical applicability of LLM-HTI, we conducted case studies on *Artemisia annua* and *Coptis chinensis* by validating their top 10 predicted targets. Three complementary strategies were used, including database validation, literature validation, and molecular docking. Database evidence was obtained from ETCM [35] and HERB [36], literature evidence was collected from PubMed, and molecular docking with AutoDock was performed to evaluate the potential interactions of artemisinin and berberine with the predicted targets. In Table 7, Targets shown in bold are supported by database, literature, or molecular docking evidence. These analyses are intended to provide supportive evidence for the biological plausibility of the predictions rather than direct proof of causal mechanisms.

3.5.1. Validation of targets for *Coptis chinensis*

Among the top 10 candidate targets predicted for *Coptis chinensis*,

Table 4
Performance of LLM-HTI under different momentum coefficients β .

β /metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
0.1	0.3782	0.3782	0.3597	0.3572	0.3777	0.3586	0.4534	0.3927
0.3	0.4015	0.4015	0.3687	0.3674	0.3855	0.3667	0.4648	0.3926
0.5	0.3923	0.3923	0.3731	0.3670	0.3866	0.3682	0.4679	0.3996
0.7	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999
0.9	0.3796	0.3787	0.3714	0.3663	0.3886	0.3706	0.4658	0.4021

Table 5
Performance of LLM-HTI under different parameter update ratios ρ .

ρ /metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
0.1	0.3814	0.3814	0.3673	0.3604	0.3852	0.3656	0.4612	0.3928
0.3	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999
0.5	0.3954	0.3954	0.3699	0.3648	0.3857	0.3664	0.4621	0.3933
0.7	0.3775	0.3774	0.3583	0.3522	0.3719	0.3548	0.4509	0.3821
0.9	0.3595	0.3603	0.3517	0.3462	0.3672	0.3515	0.4460	0.3849

Table 6
Performance of LLM-HTI under different LINE embedding dimensions.

Dimensions/ metrics	HR@1	NDCG@1	HR@3	NDCG@3	HR@5	NDCG@5	HR@10	NDCG@10
16	0.3915	0.3915	0.3692	0.3687	0.3865	0.3677	0.4627	0.3948
32	0.3943	0.3943	0.3716	0.3694	0.3871	0.3684	0.4632	0.3957
64	0.3975	0.3975	0.3745	0.3712	0.3892	0.3702	0.4669	0.3999
128	0.3924	0.3924	0.3682	0.3608	0.3816	0.3651	0.4628	0.3915
256	0.3894	0.3894	0.3671	0.3567	0.3784	0.3575	0.4584	0.3891

Table 7
Top 10 predicted candidate targets of *Artemisia annua* and *Coptis chinensis* by LLM-HTI.

Rank	<i>Coptis chinensis</i>		<i>Artemisia annua</i>	
	Predicted targets	Evidence	Predicted targets	Evidence
1	SLC2A4	Tang et al. [37]	TLR4	ETCM
2	CHRM3	Docking (BE = -7.5)	STAT3	Gao et al. [40]
3	SOD1	-	MC1R	Docking (BE = -8.7)
4	POR	Docking (BE = -6.4)	ADH1A	Docking (BE = -7.4)
5	NQO1	Shou et al. [38]	PLB1	-
6	CHRM2	HERB	TRPM8	ETCM
7	PLB1	-	CDH1	Xu et al. [41]
8	IL18	Li et al. [42]	EIF6	HERB
9	RHO	Tang et al. [43]	EGFR	Liu et al. [44]
10	AHSA1	Docking (BE = -6.0)	NOS3	Docking (BE = -7.7)

Note: Targets shown in bold are supported by database, literature, or molecular docking evidence.

eight were supported by external evidence, indicating high predictive reliability. CHRM2 has been recorded in HERB as a potential target, whereas previous studies showed that berberine may regulate SLC2A4 [37], promote NQO1 expression [38]. Molecular docking in Figs. 2a-c further supported several predictions, with CHRM3, POR and AHSA1 showing relatively strong binding to berberine; among them, CHRM3 exhibited the lowest binding energy of -7.5 kcal/mol and interacted with residues K228 and E197. Although PLB1 could not be directly evaluated because no experimentally resolved three-dimensional structure is available, it remains a biologically plausible candidate target given its role in phospholipid metabolism and inflammatory signaling, together with the known anti-inflammatory and lipid-regulating effects of berberine [39]. Together, these findings support the biological

plausibility of several top-ranked *Coptis chinensis*-target predictions and indicate that LLM-HTI can prioritize potential candidate targets while recovering known associations. However, these results should be interpreted as hypothesis-generating evidence rather than direct proof of causal mechanisms.

3.5.2. Validation of targets for *Artemisia annua*

Among the top 10 candidate targets predicted for *Artemisia annua*, nine were supported by database, literature, or molecular docking evidence, indicating high predictive reliability. TLR4 and TRPM8 have been recorded in ETCM [35], EIF6 can be retrieved from HERB [36], and published studies further support the relevance of STAT3, CDH1, and EGFR [40; 41; 44]. These findings support the biological plausibility of several top-ranked *Artemisia annua*-target predictions and indicate that LLM-HTI can prioritize candidate targets with potential research value while recovering known associations. Nevertheless, these predictions should be interpreted as hypothesis-generating evidence and require further experimental validation. Molecular docking in Figs. 3a-c further supported several predictions, with MC1R, ADH1A and NOS3 all showing relatively strong binding to artemisinin; among them, MC1R exhibited the strongest affinity, with a binding energy of -8.7 kcal/mol. By contrast, PLB1 could not be directly evaluated because no experimentally resolved three-dimensional structure is currently available. Nevertheless, given its role in phospholipid metabolism and membrane dynamics, together with the known anti-inflammatory activity of artemisinin, PLB1 remains a biologically plausible candidate target, although further experimental validation is required.

3.5.3. Clinician-guided interpretation of predicted targets in gastrointestinal diseases

To further assess the medical relevance and potential translational value of the top-ranked candidate targets, representative targets from Table 7 were subjected to independent structured evaluation by two gastrointestinal surgeons. As shown in Tables 8 and 9, among the eight representative candidate targets evaluated, seven were consistently rated by both clinicians as having medium-to-high relevance to

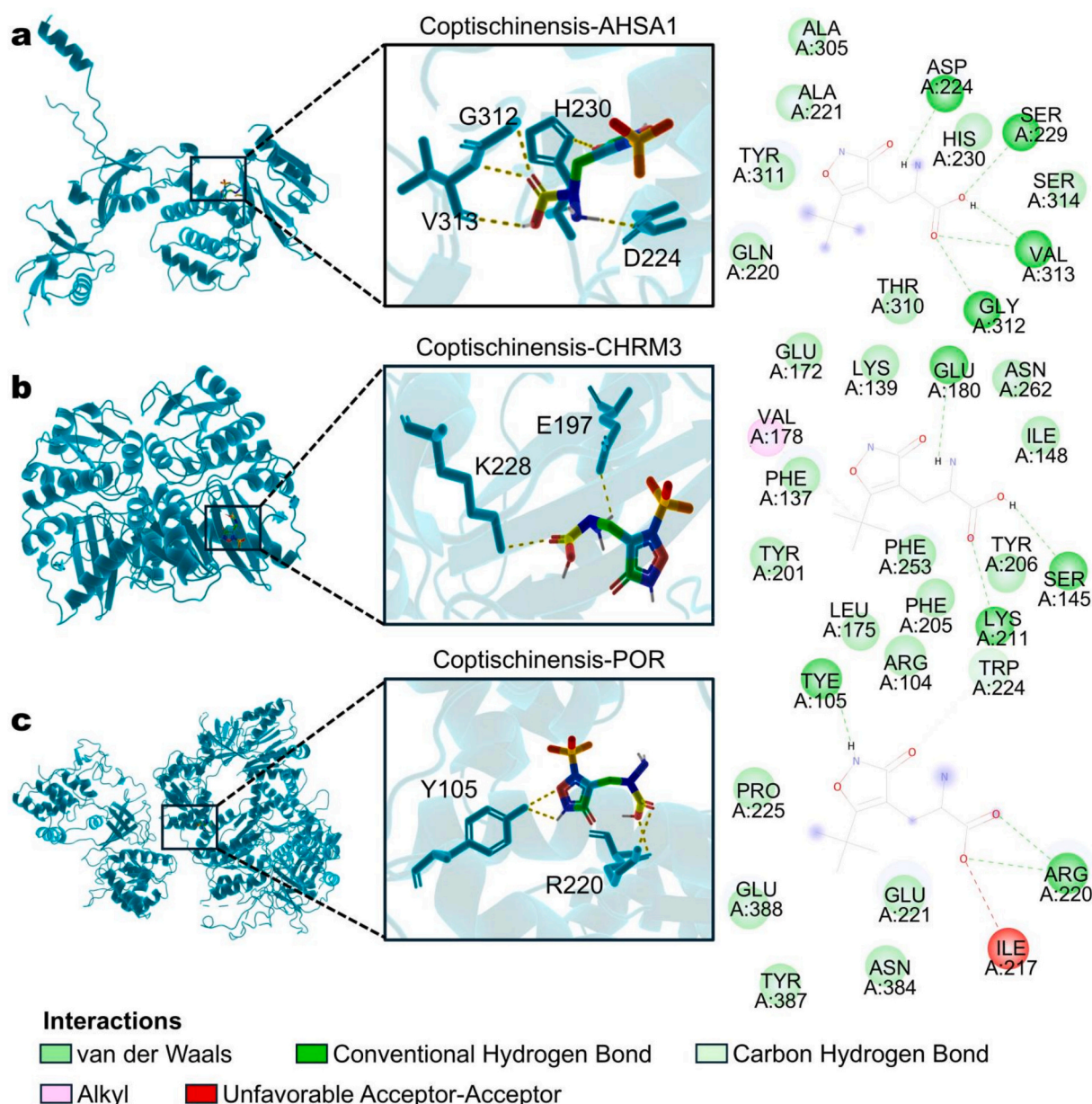


Fig. 2. Molecular docking visualization of *Coptis chinensis* with three candidate targets: (a) AHSA1, (b) CHRM3, and (c) POR.

gastrointestinal diseases. These targets mainly mapped to two disease-relevant mechanistic categories. The first category was associated with gastrointestinal tumor biology, including EGFR, STAT3, CDH1, and NQO1. The second category was related to inflammatory responses, mucosal immunity, and barrier regulation, including TLR4, IL18, and CHRM3. Most targets also showed medium-to-high clinical actionability and translational priority, suggesting that LLM-HTI predictions have not only statistical value but also medical interpretability and potential utility for downstream validation. In contrast, SLC2A4 showed relatively limited gastrointestinal disease specificity and direct translational value, with mild disagreement between the two clinicians, indicating that it may be better interpreted as a metabolism-related candidate rather than a high-priority gastrointestinal disease target.

From the perspective of modern gastrointestinal disease mechanisms, several top-ranked targets predicted by LLM-HTI intersect with key pathophysiological processes, including gastrointestinal tumorigenesis, chronic inflammation, mucosal immunity, epithelial barrier regulation, and metabolic control. For example, STAT3 links cytokine

signaling, chronic inflammation, and inflammation-associated tumorigenesis; EGFR has established disease relevance and clinical intervention value in gastrointestinal malignancies, particularly colorectal cancer; and CDH1 is closely related to epithelial adhesion, gastric mucosal barrier maintenance, and gastric cancer development. Therefore, gastrointestinal disease-oriented clinician assessment helps further prioritize candidate targets with disease relevance, mechanistic interpretability, and translational potential from computational predictions. Compared with algorithmic metrics such as HR and NDCG alone, this strategy integrates model prediction, database and molecular-level evidence, and clinical expert judgment, thereby establishing a multilayered evaluation path from computational prediction to biological evidence and clinical medical interpretation.

4. Discussions

Built upon the integration of traditional medicine theory and modern biomedical information, this study develops a large language model-

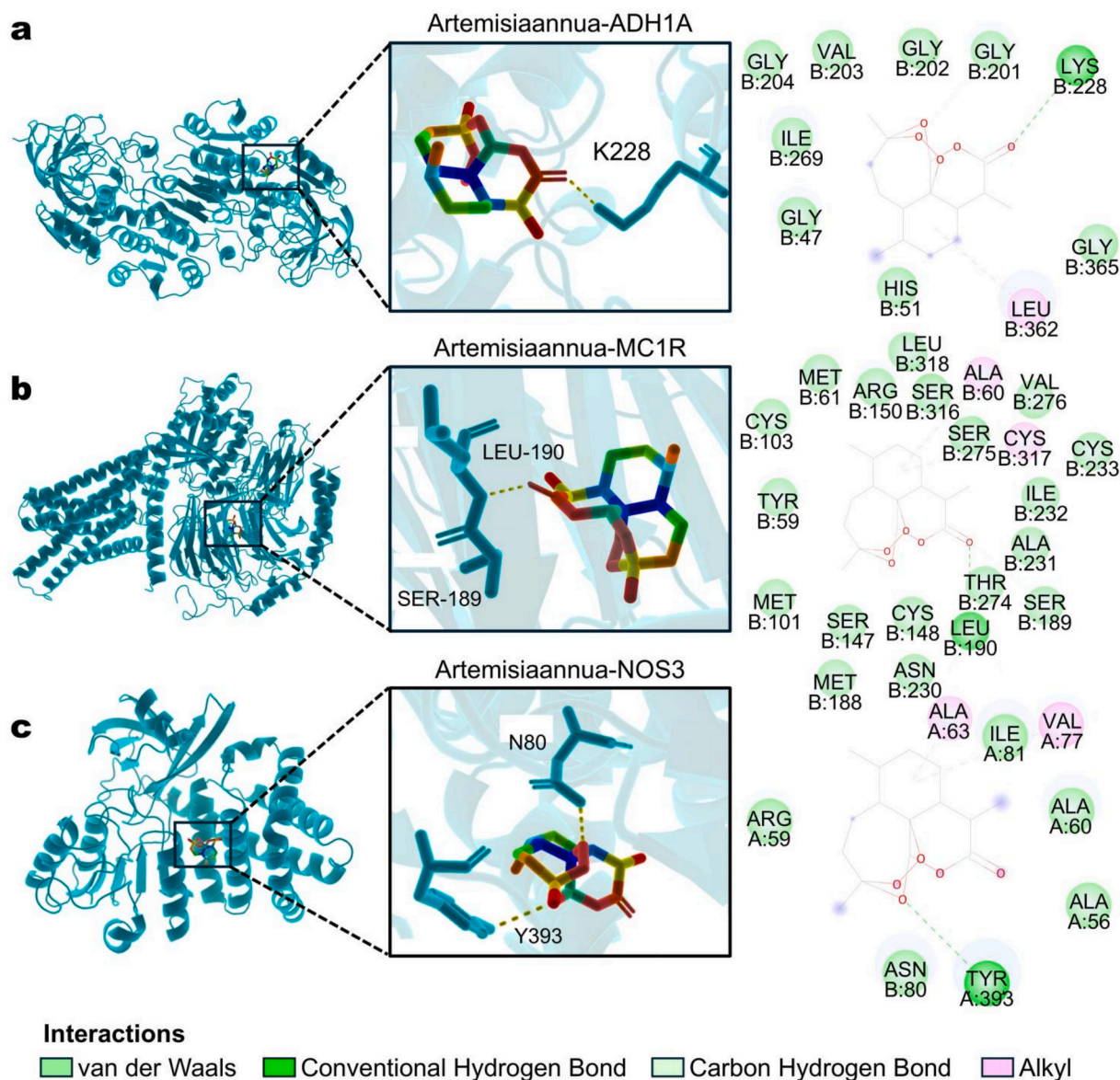


Fig. 3. Molecular docking visualization of *Artemisia annua* with three candidate targets: (a) ADH1A, (b) MC1R, and (c) NOS3.

driven framework for HTI modeling. The framework captures higher-order topological dependencies between herbs and targets through multi-layer graph convolution, and further improves training stability and optimization efficiency by dynamically selecting key parameter subsets via a parameter-adaptive updating strategy, without altering the model architecture. Experimental results show that LLM-HTI consistently outperforms traditional link prediction methods and existing HTI models, while ablation and hyperparameter analyses further indicate that multi-hop structural modeling and parameter-adaptive updating are the major sources of performance gain.

More importantly, the case studies provide biologically plausible and medically interpretable evidence for selected high-ranking predictions. After excluding known targets in the test set, multiple top-ranked predictions for *Artemisia annua* and *Coptis chinensis* were supported by databases, literature, or molecular docking. These findings suggest that LLM-HTI may help prioritize candidate targets, support active compound screening, and generate mechanistic hypotheses, but they should not be interpreted as direct evidence of causal mechanisms.

Nevertheless, several limitations remain in the current study. First, the semantic representations rely on LLM-generated entity descriptions,

whose quality may still be affected by prompt design, model behavior, and domain knowledge coverage. Second, the present validation is mainly based on benchmark evaluation together with post hoc evidence from databases, literature, and molecular docking, rather than direct functional experiments or mechanism-level explanation analyses. Third, the current framework is built on a relatively static herb-target interaction graph and does not yet incorporate richer multimodal information such as clinical phenotypes and omics perturbation signals. Therefore, future work will focus on improving the reproducibility and domain specificity of semantic generation, introducing more direct explanation and experimental validation strategies, and extending the framework to more realistic scenarios, including formula compatibility and complex disease intervention.

5. Conclusions

We propose LLM-HTI, an HTI prediction framework jointly driven by LLMs and an integrative Chinese–Western medicine knowledge graph, enabling unified modeling of traditional medicine knowledge, modern biomedical relations, and graph representation learning. By combining

Table 8
Clinician-guided gastrointestinal interpretation of top-ranked *Artemisia annua*-target candidate pairs.

Targets	Herbs	Evidence type	Gastrointestinal clinical relevance	Clinical actionability	Translational priority
EGFR	<i>Artemisia annua</i>	Literature	High/High. EGFR is associated with gastrointestinal epithelial repair and tumor signaling, with established disease relevance in colorectal cancer.	High/High. EGFR is a well-established clinically actionable druggable target, but its modulation by <i>Artemisia annua</i> requires validation.	High/High
STAT3	<i>Artemisia annua</i>	Literature	High/High. STAT3 is involved in cytokine signaling and inflammation-associated tumorigenesis, and is associated with the transition from chronic gastrointestinal inflammation to cancer.	High/High. The STAT3 pathway has clear pharmacological relevance, but direct clinical targeting remains context-dependent.	High/High
CDH1	<i>Artemisia annua</i>	Literature	High/High. CDH1 is involved in epithelial cell adhesion and gastric mucosal barrier maintenance, with established disease relevance in gastric cancer development and progression.	Medium/Medium. CDH1 has strong mechanistic and biomarker value but is not a conventional directly druggable target.	High/High
TLR4	<i>Artemisia annua</i>	ETCM database	High/High. TLR4 participates in microbial recognition and mucosal immune regulation, while its dysregulated activation is associated with gastrointestinal inflammation and barrier injury.	Medium/Medium. The TLR4 pathway is suitable for mechanistic validation in inflammation and barrier injury models, but direct clinical targeting remains immature.	High/High

Note: Gastrointestinal clinical relevance, clinical actionability, and translational priority were independently assessed by two gastrointestinal surgeons using a three-level scale: High, Medium, and Low. Ratings such as “High/Medium” indicate the independent scores assigned by the first and second clinician, respectively. The overall rationale was summarized based on both clinicians' ratings and medical interpretation.

Table 9
Clinician-guided gastrointestinal interpretation of top-ranked *Coptis chinensis*-target candidate pairs.

Targets	Herbs	Evidence type	Gastrointestinal clinical relevance	Clinical actionability	Translational priority
IL18	<i>Coptis chinensis</i>	Literature	High/High. IL18 is associated with inflammasome activation, mucosal immune regulation, and intestinal inflammation, and is involved in IBD-related mechanisms.	Medium/Medium. IL18 marks inflammation and mucosal immune regulation, but its therapeutic targetability remains uncertain.	High/Medium
CHRM3	<i>Coptis chinensis</i>	Molecular docking	Medium/Medium. CHRM3 is involved in gastrointestinal motility, secretion, and epithelial homeostasis, with potential neuroimmune relevance.	Medium/Medium. CHRM3 has some pharmacological tractability, but requires disease-specific validation.	Medium/Medium
NQO1	<i>Coptis chinensis</i>	Literature	Medium/Medium. NQO1 is associated with oxidative stress regulation and tumor redox homeostasis, and may indirectly contribute to gastrointestinal injury and tumor biology.	Medium/Medium. NQO1 may serve as a mechanistic marker of <i>Coptis chinensis</i> -related oxidative stress, with relatively indirect clinical translational value.	Medium/Medium
SLC2A4	<i>Coptis chinensis</i>	Literature	Medium/Low. SLC2A4 is mainly involved in glucose transport and metabolic regulation, while its direct association with gastrointestinal diseases remains unclear.	Medium/Low. SLC2A4 has mechanistic relevance to metabolic regulation, but its gastrointestinal disease specificity and direct translational value are limited.	Medium/Low

Note: Gastrointestinal clinical relevance, clinical actionability, and translational priority were independently assessed by two gastrointestinal surgeons using a three-level scale: High, Medium, and Low. Ratings such as “High/Medium” indicate the independent scores assigned by the first and second clinician, respectively. The overall rationale was summarized based on both clinicians' ratings and medical interpretation.

semantic enhancement, multi-layer graph aggregation, and parameter-adaptive updating, the framework achieves strong performance, stability, and interpretability. Experimental and case-study results show that LLM-HTI effectively identifies potential herb–target relationships and uncovers pharmacologically supported candidate targets. Additionally, the clinician-guided assessment of disease relevance and translational value by gastrointestinal surgeons further integrates model predictions with modern clinical medical knowledge, providing complementary support for improving the medical interpretability and translational potential of AI-assisted candidate target discovery.

CRedit authorship contribution statement

Sisi Yuan: Writing – original draft, Methodology, Data curation. **Huihui Wu:** Methodology, Data curation. **Jiapeng Yao:** Writing – review & editing, Supervision, Methodology. **Xinze Lian:** Methodology, Data curation. **Linlin Zhuo:** Methodology. **Keqin Li:** Writing – review & editing, Supervision. **Dongsheng Cao:** Writing – review & editing, Supervision. **Aiping Lu:** Writing – review & editing, Supervision, Methodology. **Jianfu Xia:** Writing – review & editing, Supervision,

Methodology.

Declaration of competing interest

We declare that we have no conflict of interest.

Acknowledgements

The authors thank Senzhe Xia from the Department of General Surgery, Wenzhou Central Hospital, Affiliated to Wenzhou Medical University, for assistance with clinician-guided assessment of candidate targets in gastrointestinal diseases. The authors also thank Mingming Qi for assistance with figure preparation. This work was supported in part by the Major Scientific and Technological Innovation Project of Wenzhou (No. ZG2024013) and the Key Project Jointly Constructed by the Ministry and the Province of Zhejiang Medical and Health Science and Technology Project (No. WKJ-ZJ-26088).

Data availability

Our data and code are publicly available at <https://github.com/sis-yuan/LLM-HTI>.

References

- [1] R. Santos, O. Ursu, A. Gaulton, et al., A comprehensive map of molecular drug targets, *Nat. Rev. Drug Discov.* 16 (1) (2017) 19–34.
- [2] X. Xie, T. Yu, X. Li, et al., Recent advances in targeting the “undruggable” proteins: from drug discovery to clinical trials, *Signal Transduct. Target. Ther.* 8 (1) (2023) 335.
- [3] C.M. Connelly, M.H. Moon, J.S. Schneekloth, The emerging role of RNA as a therapeutic target for small molecules, *Cell Chem. Biol.* 23 (9) (2016) 1077–1090.
- [4] L. Shao, B. Zhang, Traditional Chinese medicine network pharmacology: theory, methodology and application, *Chin. J. Nat. Med.* 11 (2) (2013) 110–120.
- [5] F.-S. Li, J.-K. Weng, Demystifying traditional herbal medicine with modern approach, *Nat. Plants* 3 (8) (2017) 17109.
- [6] S. Zhang, H. Shu, J. Zhou, et al., Single-cell polygenic risk scores dissect cellular and molecular heterogeneity of complex human diseases, *Nat. Biotechnol.* 44 (5) (2026) 845–861.
- [7] F. Cheung, TCM: made in China, *Nature* 480 (7378) (2011) S82–S83.
- [8] W.-L. Wei, R. Zeng, C.-M. Gu, et al., *Angelica sinensis* in China-A review of botanical profile, ethnopharmacology, phytochemistry and chemical analysis, *J. Ethnopharmacol.* 190 (2016) 116–141.
- [9] A.L. Hopkins, Network pharmacology: the next paradigm in drug discovery, *Nat. Chem. Biol.* 4 (11) (2008) 682–690.
- [10] B. Yang, Y. Liu, J. Wu, et al., Genndti: Drug-target interaction prediction using graph neural network enhanced by router nodes, *IEEE J. Biomed. Health Inform.* 28 (12) (2024) 7588–7598.
- [11] H. Hu, Y. Li, Z. Li, Nffgram: Nonlinear multi-feature fusion and gated recurrent self-attention mechanism for traditional chinese medicine formula recommendation, *IEEE J. Biomed. Health Inform.* 29 (5) (2025) 3698–3711.
- [12] A.E. Gururaj, R.H. Scheuermann, D. Lin, AI and immunology as a new research paradigm, *Nat. Immunol.* 25 (11) (2024) 1993–1996.
- [13] J. Long, J. Niu, H. Wang, et al., RTGN: robust traditional Chinese medicine graph networks for patient similarity learning, *IEEE J. Biomed. Health Inform.* 29 (3) (2024) 2185–2198.
- [14] K. Yang, X. Zhou, R. Zhang, et al., Integrating herb effect similarity for network-based herb target prediction, in: 2015 8th international conference on biomedical engineering and informatics (BMEI), 2015.
- [15] K. Yang, G. Liu, N. Wang, et al., Heterogeneous network propagation for herb target identification, *BMC Med. Inform. Decis. Mak.* 18 (Suppl. 1) (2018) 17.
- [16] S. Zhang, K. Yang, Z. Liu, et al., DrugAI: a multi-view deep learning model for predicting drug–target activating/inhibiting mechanisms, *Brief. Bioinform.* 24 (1) (2023) bbac526.
- [17] K. Yang, Y. Yang, S. Fan, et al., DRONet: effectiveness-driven drug repositioning framework using network embedding and ranking learning, *Brief. Bioinform.* 24 (1) (2023) bbac518.
- [18] Z. Zhou, Q. Liao, J. Wei, et al., Revisiting drug–protein interaction prediction: a novel global–local perspective, *Bioinformatics* 40 (5) (2024) btac271.
- [19] J. Wei, Y. Zhu, L. Zhuo, et al., Efficient deep model ensemble framework for drug-target interaction prediction, *The Journal of Physical Chemistry Letters* 15 (30) (2024) 7681–7693.
- [20] P. Duan, K. Yang, X. Su, et al., HTINet2: herb–target prediction via knowledge graph embedding and residual-like graph neural network, *Brief. Bioinform.* 25 (5) (2024) bbac414.
- [21] J. Wei, L. Zhuo, X. Fu, et al., DrugReAlign: a multisource prompt framework for drug repurposing based on large language models, *BMC Biol.* 22 (1) (2024) 226.
- [22] S. Yuan, Z. Zhou, X. Jin, et al., Enhancing herbal medicine-drug interaction prediction using large language models, *IEEE J. Biomed. Health Inform.* 29 (10) (2025) 6971–6981.
- [23] T.Y.C. Tam, S. Sivarajkumar, S. Kapoor, et al., A framework for human evaluation of large language models in healthcare derived from literature review, *NPJ digital medicine* 7 (1) (2024) 258.
- [24] Y. Wu, F. Zhang, K. Yang, et al., SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping, *Nucleic Acids Res.* 47 (D1) (2019) D1110–D1117.
- [25] X. Su, P. Hu, H. Yi, et al., Predicting drug-target interactions over heterogeneous information network, *IEEE J. Biomed. Health Inform.* 27 (1) (2022) 562–572.
- [26] Y. Zheng, H.Y. Koh, J. Ju, et al., Large language models for drug discovery and development, *Patterns* 6 (10) (2025).
- [27] J. Lu, K. Choi, M. Ereemeev, et al., Large language models and their applications in drug discovery and development: a primer, *Clin. Transl. Sci.* 18 (4) (2025) e70205.
- [28] H. Zhou, F. Liu, J. Wu, et al., A collaborative large language model for drug analysis, *Nat. Biomed. Eng.* 10 (5) (2026) 870–881.
- [29] J. Devlin, M.-W. Chang, K. Lee, et al., Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019.
- [30] M. Davenport, Introduction to modern information retrieval, *Journal of the Medical Library Association: JMLA* 100 (1) (2012) 75.
- [31] E. Ravasz, A.L. Somera, D.A. Mongru, et al., Hierarchical organization of modularity in metabolic networks, *Science* 297 (5586) (2002) 1551–1555.
- [32] E.A. Leicht, P. Holme, M.E. Newman, Vertex similarity in networks, *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 73 (2) (2006) 026120.
- [33] L.A. Adamic, E. Adar, Friends and neighbors on the web, *Soc. Networks* 25 (3) (2003) 211–230.
- [34] O. Vanunu, O. Magger, E. Ruppin, et al., Associating genes and protein complexes with disease via network propagation, *PLoS Comput. Biol.* 6 (1) (2010) e1000641.
- [35] Y. Zhang, X. Li, Y. Shi, et al., ETCM v2. 0: an update with comprehensive resource and rich annotations for traditional Chinese medicine, *Acta Pharm. Sin. B* 13 (6) (2023) 2559–2571.
- [36] S. Fang, L. Dong, L. Liu, et al., HERB: a high-throughput experiment-and reference-guided database of traditional Chinese medicine, *Nucleic Acids Res.* 49 (D1) (2021) D1197–D1206.
- [37] W.-K. Tang, Q.-Y. Li, J. Liu, et al., Research progress on signaling pathway of berberine in relieving insulin resistance, *Drugs & Clinic* 37 (6) (2022) 1409–1413.
- [38] J.-W. Shou, X.-X. Li, Y.-S. Tang, et al., Novel mechanistic insight on the neuroprotective effect of berberine: The role of PPAR δ for antioxidant action, *Free Radic. Biol. Med.* 181 (2022) 62–71.
- [39] J.E. Burke, E.A. Dennis, Phospholipase A2 structure/function, mechanism, and signaling1, *J. Lipid Res.* 50 (2009) S237–S242.
- [40] P. Gao, L.-I. Wang, J. Liu, et al., Dihydroartemisinin inhibits endothelial cell tube formation by suppression of the STAT3 signaling pathway, *Life Sci.* 242 (2020) 117221.
- [41] N. Xu, X. Zhou, S. Wang, et al., Artesunate induces SKM-1 cells apoptosis by inhibiting hyperactive β -catenin signaling pathway, *Int. J. Med. Sci.* 12 (6) (2015) 524.
- [42] C. Li, F. Yang, Z. Wu, et al., Coptisine alleviates intestinal barrier dysfunction and inflammation in TNBS-induced colitis rats and LPS-stimulated human intestinal Caco-2 cells, *Int. Immunopharmacol.* 167 (2025) 115697.
- [43] F. Tang, D. Wang, C. Duan, et al., Berberine inhibits metastasis of nasopharyngeal carcinoma 5-8F cells by targeting Rho kinase-mediated Ezrin phosphorylation at threonine 567, *J. Biol. Chem.* 284 (40) (2009) 27456–27466.
- [44] Q. Liu, Z. Yang, Progress of anti-tumor activities of artemisinin and its derivatives, *Chin. Bull. Life Sci.* 32 (2020) 62–69.