KNOWACU: A KNOWLEDGE-GUIDED HETEROGENEOUS MULTI-GRAPH LEARNING FRAMEWORK FOR ACUPOINT PRESCRIPTION

Anonymous authors
Paper under double-blind review

ABSTRACT

As an indispensable component of Traditional Chinese Medicine (TCM), acupuncture therapy plays a pivotal role in both clinical treatment and preventive healthcare. However, generating appropriate acupoint prescriptions—structured treatment schemes tailored to complex symptoms remains a formidable challenge in automated medical reasoning. Existing studies struggle to bridge the semantic knowledge gap between diverse symptom manifestations and rigorous TCM laws, often failing to capture the inherent structural heterogeneity of symptom-acupoint interactions.

In this paper, we propose KnowAcu, a **Knowledge-Guided** framework for precise **Acupoint prescription** that synergizes TCM-driven heterogeneous multi-graph structures with the semantic expertise of Large Language Models (LLMs). Specifically, we construct five distinct graphs to explicitly model the multi-faceted relational dependencies defined by TCM principles and historical prescriptions. These structures are integrated through a Multi-Graph Joint Learning (MGJL) module, which extracts multi-granular representations of symptoms and acupoints. To align the model with professional clinical logic, we leverage LLMs for auxiliary supervision and knowledge distillation, incorporate the functional hierarchies of acupoints and employ an acupoint-symptom reconstruction task for semantic consistency verification. Extensive experiments on a curated real-world clinical dataset and a large-scale simulated benchmark demonstrate that KnowAcu significantly outperforms state-of-the-art methods in both clinical applicability and model robustness. We will release our source code and the optimized synthetic benchmark upon the acceptance of this paper.

1 INTRODUCTION

Acupuncture is a therapeutic technique originating from Traditional Chinese Medicine (TCM) that has gained global recognition for its clinical efficacy NIH (1998); Bonafede et al. (2008); Goldman et al. (2010). Guided by syndrome differentiation and treatment, acupuncture prescriptions require personalized acupoint selection according to individual patient conditions. In practice, physicians formulate treatment plans based largely on symptoms and accumulated clinical experience Yang et al. (2022). However, such expertise often remains fragmented as individual "knowledge islands" Zhou et al. (2021), limiting the systematic dissemination of TCM knowledge and the generation of consistent, clinically informed acupoint prescriptions.

Deep learning has been increasingly applied to medical recommendation, particularly for herbal and drug prescription generation Mao et al. (2022); Jin et al. (2023); Zhang et al. (2022); Tan et al. (2022). Knowledge-graph-based methods model complex relations among symptoms, diseases, herbs, and attributes Yang et al. (2026); Lin et al. (2026); Yang et al. (2022); Jin et al. (2023); Zhang et al. (2023c), but rely on predefined structures and expert-curated knowledge, making them vulnerable to incomplete knowledge and limited scalability. Generative, multi-graph, and attention-based approaches improve recommendation flexibility and relational modeling Rong et al. (2022); Zhang et al. (2023c); Yang & Ding (2023); Li et al. (2024), yet often lack explicit clinical constraints or fail to capture syndrome-level differentiation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Recent advances in generative AI (GAI) and large language models (LLMs) offer new opportunities for TCM clinical decision support and personalized treatment Wang et al. (2023c); Zhu et al. (2024); Tan et al. (2024). However, TCM-specific concepts, including zang–fu theory, eight-principle syndrome differentiation, and holistic diagnosis, present substantial challenges for LLM-based reasoning Cyranoski (2018); Zhang et al. (2024); Tan et al. (2024); Liu et al. (2024). Despite growing efforts to apply LLMs to TCM diagnosis and treatment Wang et al. (2023c); Zhu et al. (2024); Mondal et al. (2024), effectively integrating TCM knowledge, syndrome differentiation, and clinical experience into accurate and personalized acupoint recommendation remains an open challenge.

To bridge the gap between empirical clinical patterns and rigorous theoretical principles, we propose **KnowAcu**, a pioneering **knowledge-guided** framework for automated **acupoint** prescription. Unlike traditional recommendation engines that treat symptoms as isolated tokens, KnowAcu explicitly transforms the fragmented "knowledge islands" of TCM into a unified, computable heterogeneous multi-graph. By synergizing five distinct relational dimensions—spanning from historical prescription data to zang-fu organ and meridian theories—the framework captures the high-order relational dependencies essential for medical reasoning. Furthermore, by distilling high-level semantic expertise from Large Language Models (LLMs) into the graph embedding space, KnowAcu ensures that the generated treatment schemes are not only statistically accurate but also clinically consistent with professional medical logic.

Our contributions are summarized as follows:

- We propose KnowAcu, a novel knowledge-guided paradigm that redefines acupoint selection as a structured medical reasoning task, addressing the semantic gap in automated TCM prescription.
- We design a Multi-Graph Joint Learning (MGJL) module to fuse five types of relational dependencies (e.g., symptom-symptom, zang-fu organ correlations, and meridians). By explicitly modeling these interactions, KnowAcu effectively integrates empirical prescription patterns with structured TCM theoretical frameworks to capture high-order medical associations.
- We introduce a specialized distillation objective incorporating functional hierarchies and semantic consistency checks, injecting domain-specific linguistic expertise into the graph-based representations.
- We conduct extensive experiments on a curated real-world clinical dataset and a large-scale synthetic benchmark derived from public TCM resources. The results demonstrate that KnowAcu significantly outperforms state-of-the-art baselines, specifically validating its efficacy in multi-graph denoising and its superior capacity to model complex, theory-driven TCM knowledge.

2 RELATED WORKS

Clinical Recommendation for TCM: TCM recommendation methods primarily focus on herbal prescriptions by integrating textbook knowledge and expert experience. Current approaches leverage knowledge graphs and deep generative frameworks, utilizing Graph Neural Networks (GNNs) to fuse hierarchical symptom-herb relations, herbal properties, and pharmacological expertise for accurate and explainable results Zhang et al. (2023c); Yang & Ding (2023). In parallel, deep generative models, including SeqGAN- and CGAN-based frameworks, formulate prescription recommendation as a data-driven generation process, enabling both unconditional and condition-specific herbal prescription synthesis Rong et al. (2022). Recent studies Lin et al. (2026); Yang et al. (2026); Li et al. (2024) have further extended graph-based recommendation to acupoint selection by incorporating acupoint attribute embeddings and attention-enhanced GCNs, these methods are typically limited to single-disease settings.

Graph-based Medical Representation Learning: GNN-based methods have become widely adopted for recommendation, but their performance heavily depends on graph construction Chen et al. (2025); Kong et al. (2022); Long et al. (2021). Early approaches exploited high-order connectivity but suffered from noise and over-smoothing, while subsequent studies introduced attention, temporal modeling, and heterogeneous or multi-graph architectures to capture adaptive neighbor importance and diverse relational information Wang et al. (2019); He et al. (2020); Liu et al. (2021); Cen et al. (2019); Fan et al. (2022); Wang et al. (2018; 2022); Zhang et al. (2023a); Zhou et al. (2023). Despite these advances, most multi-graph GNNs learn representations independently from

different relational graphs and fuse them through shallow operations such as concatenation or simple aggregation Zhang et al. (2019), limiting cross-relation interaction and information propagation. Consequently, latent structural dependencies across relations remain insufficiently exploited. In contrast, unified graph convolution enables joint multi-hop message passing across multiple relations, providing more expressive representations while avoiding the computational and architectural complexity of separately modeling each graph.

Knowledge-Integrated LLMs in TCM: LLMs have shown strong potential for intelligent TCM applications, with recent studies adapting general-purpose models through domain-specific pretraining, instruction tuning, and reinforcement learning using TCM corpora Wang et al. (2023b); Zhang et al. (2023b); Yang et al. (2024a); Zhang et al. (2024); Yang et al. (2024b); Wang et al. (2023a). These approaches improve performance in TCM diagnosis, consultation, and medical dialogue, but primarily emphasize linguistic generation and remain limited in modeling the structural logic and heterogeneous relationships underlying clinical treatment. As a result, they often lack explicit integration of TCM theories with empirical prescription knowledge. In contrast, our KnowAcu framework combines the semantic capabilities of LLMs with a knowledge-guided multi-graph architecture, enabling more structured, interpretable, and theoretically consistent acupoint recommendation.

3 METHODOLOGY

The pipeline of KnowAcu is illustrated in the Figure 1. Our framework follows a systematic workflow that translates raw clinical observations into expert-level acupoint prescriptions through three synergistic stages. First, the Graph Construction stage transforms prescription data and TCM domain knowledge into five heterogeneous relational graphs. Second, these graphs are fed into the Multi-Graph Joint Learning (MGJL) module to perform deep feature fusion and representation learning. Finally, the generated embeddings are refined through an LLM-guided recommendation process to produce the final treatment scheme. Prior to graph construction, we initially perform data preprocessing:

Let $S = \{s_1, s_2, \ldots, s_m\}$ and $A = \{a_1, a_2, \ldots, a_n\}$ denote the sets of all symptoms and acupoints, respectively. The objective of the acupoint recommendation task is to generate an appropriate treatment scheme based on a given set of patient symptoms $sc = \{s_1, s_2, \ldots\}$. The treatment scheme is represented as a set of acupoints $ac = \{a_1, a_2, \ldots\}$.

A symptom set and its corresponding acupoint set together constitute a prescription $p = \langle sc, ac \rangle$. The symptom vector $sc \in \mathbb{R}^m$ is encoded using a multi-hot representation, where the i-th entry equals 1 if symptom s_i is present in sc, and 0 otherwise. Similarly, the acupoint vector $ac \in \mathbb{R}^n$ is also encoded in a multi-hot representation. For clarity of exposition, node indices in each graph relation are uniformly denoted by i, j, and k.

Therefore, we define the input and output of the acupoint recommendation task as follows:

- Input: prescriptions P, sc.
- Output: A learned function $f(sc, P; \theta)$, which generates the probability vector $\hat{ac}$ which represents the probability that each acupoint is effective for sc.

P indicates all prescriptions and θ is the trainable parameter of function f.

3.1 GRAPH CONSTRUCTION MODULE

3.1.1 SYMPTOM-SYMPTOM GRAPH CONSTRUCTION

We construct the symptom–symptom graph SS^p based on prescription, where each node represents a symptom, and an edge is established between s_i and s_j when they co-occur within the same prescription. We model the symptom-symptom graph as an undirected graph, which is formulated as follows:

$$SS^p_{i,j},\ SS^p_{j,i} = \begin{cases} 1, & \text{if } s_i, s_j \text{ co-occur in the same prescriptions;} \\ 0, & \text{otherwise} \end{cases}. \tag{1}$$

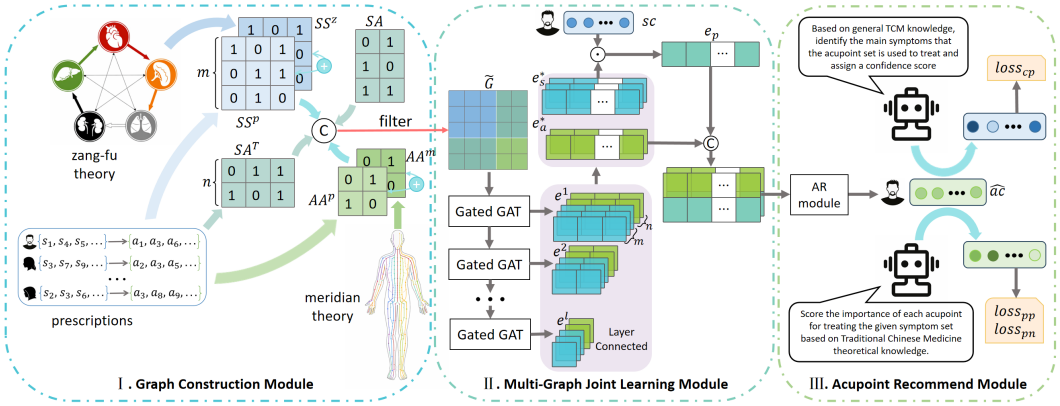


Figure 1: Overview of our proposed KnowAcu framework. KnowAcu consists of three parts: I. Graph Construction Module, II. Multi-Graph Joint Learning Module and III. Acupoint Recommend Module.

Subsequently, we construct the symptom–symptom graph SS^z based on zang-fu theory. Zang–fu theory posits that the manifestation of external symptoms originates from abnormalities in one or more zang–fu organs, the relevant explanations of this theory are found in Appendix, and based on this theory we construct a graph to model the relationships among symptoms. The formulation is as follows:

$$SS^z_{i,j},\ SS^z_{j,i} = \begin{cases} 1, & \text{if } s_i, s_j \text{ co-occur in the same zang-fu organ;} \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

We first fuse the two graphs SS^p and SS^z to generate an integrated symptom–symptom graph SS containing information from both:

$$SS = SS^p + \lambda_s \cdot SS^z, \tag{3}$$

where λ_s denotes weight parameter of SS^z.

3.1.2 Acupoint-Acupoint Graph Construction

Similarly, we construct the acupoint–acupoint graph AA^p based on prescription, where each node represents an acupoint, and edges are formed between acupoints that co-occur within the same prescription. The formulation is as follows:

$$AA^p_{i,j},\ AA^p_{j,i} = \begin{cases} 1, & \text{if } a_i, a_j \text{ co-occur in the same prescriptions;} \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Subsequently, we construct the acupoint–acupoint graph AA^m based on meridian theory, which posits that acupoints are intrinsically correlated through meridian pathways, the relevant explanations of this theory are found in Appendix, and this graph models the relational structure induced by meridian connections. The formulation is as follows:

$$AA^m_{i,j},\ AA^m_{j,i} = \begin{cases} 1, & \text{if } a_i, a_j \text{ co-occur in the same meridian;} \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

Similarly, we fuse the two graph AA^p and AA^m into AA:

$$AA = AA^p + \lambda_a \cdot AA^m, \tag{6}$$

where λ_a denotes weight parameter of AA^m.

3.1.3 SYMPTOM-ACUPOINT GRAPH CONSTRUCTION

Moreover, we investigate the associations between symptoms and acupoints to construct the bipartite graph SA. The formulation is as follows:

$$SA_{i,j},\, SA_{j,i} = \begin{cases} 1, & \text{if } s_i, a_j \text{ co-occur in the same prescriptions;} \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

3.2 MULTI-GRAPH JOINT LEARNING MODULE

3.2.1 JOINT MULTI-GRAPH FILTERING

We can integrate multiple graphs to construct a unified relational graph $G \in R^{(m+n)\times(m+n)}$, which is defined as follows:

$$G = \begin{pmatrix} SS & SA \\ SA^T & AA \end{pmatrix}, \tag{8}$$

where the graph $(SS \quad SA)$ corresponds to symptom nodes, and the edges indicate the relationships between a given symptom node and other symptoms as well as acupoints. Similarly, the graph $\begin{pmatrix} SA^T & AA \end{pmatrix}$ corresponds to acupoint nodes, and the edges represent the relationships between a given acupoint node and other symptoms and acupoints. To facilitate subsequent processing, we integrate these components into a single unified graph structure.

Given the complexity of the graph relations, we prune the graphs based on the topological similarity between nodes. Given the graph G, we first compute the cosine similarity matrix C:

$$C_{ij} = \frac{g_i^T g_j}{|g_i||g_j|}, \tag{9}$$

where g_i and g_j denote neighborhood vectors of nodes i and j. We then apply threshold filtering to C:

$$\widetilde{C}_{ij} = C_{ij} \cdot \mathbb{I}\{C_{ij} \geq \delta\}, \tag{10}$$

where δ denotes the predefined threshold. Finally, similarity weights are injected into G, enabling the graph representation to jointly capture structural and feature information:

$$\widetilde{G} = G \odot \widetilde{C}, \tag{11}$$

where $\odot$ denotes element-wise multiplication.

3.2.2 GATED AND LAYER-CONNECTED GRAPH ATTENTION MODULE

We incorporate a gating mechanism into the original GAT Velickovic et al. (2017) architecture to compute gated scores according to the association strength between nodes. The first-layer original GAT formulation is given by:

$$\alpha_{ij}^1 = \frac{\exp\big(\text{LeakyReLU}(\mathbf{a}_1^T[W_i^1\mathbf{e}_i^0 \| W_j^1\mathbf{e}_j^0])\big)}{\sum_{k\in\mathcal{N}_i}\exp\big(\text{LeakyReLU}(\mathbf{a}_1^T[W_i^1\mathbf{e}_i^0 \| W_k^1\mathbf{e}_k^0])\big)}, \tag{12}$$

where $\cdot^T$ denotes transposition, $\mathbf{a}_1$ denotes a first-layer learnable vector, e_i^0, e_j^0 and e_k^0 are the initial embedding of node i, j and k, W_i^1 and W_j^1 are the weight matrix applied to e_i^0 and e_j^0, $\|$ indicates concatenation, and $\mathcal{N}_i$ denotes the neighborhood of node i from graph $\widetilde{G}$.

We introduce the gated score γ_{ij} to explicitly modulate message propagation based on the intrinsic association strength between node features, which the original GAT attention alone may not fully capture. The first-layer gated score γ_{ij}^1 is formulated as follows:

$$\gamma_{ij}^1 = \begin{cases} \text{Sigmoid}(W_i^1\mathbf{e}_i^0 \cdot (W_j^1\mathbf{e}_j^0)^T, & \text{if } j \in \mathcal{N}_i; \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

$$\mathbf{e}_i^1 = \sigma\left(\sum_{j\in\mathcal{N}_i} \gamma_{ij}^1 \alpha_{ij}^1 W_j^1 \mathbf{e}_j^0\right), \tag{14}$$

where σ denotes ELU activation function. We can further generalize the one-hop propagation rule to a multi-layer setting. Specifically, the representation of node i at the l-th layer is recursively defined as follows:

$$\mathbf{e}_i^l = \sigma\left(\sum_{j\in\mathcal{N}_i} \gamma_{ij}^l \alpha_{ij}^l \, W_j^l \mathbf{e}_j^{l-1}\right), \tag{15}$$

where $e^l = \begin{pmatrix} e_s^l \\ e_a^l \end{pmatrix}$, e_s^l and e_a^l denote the l-order embeddings of symptoms and acupoints, respectively.

Concatenating layer-wise embeddings is intended to preserve and integrate multi-order relational information captured at different propagation depths. Specifically, lower-layer embeddings encode local, fine-grained interactions (e.g., direct symptom–acupoint associations), while higher-layer embeddings capture more abstract and global dependencies through multi-hop message passing. We concatenate all layer-wise node embeddings to obtain the final representations $e_s^* = [e_s^1 \| e_s^2 \| \dots \| e_s^l]$ and $e_a^* = [e_a^1 \| e_a^2 \| \dots \| e_a^l]$.

3.3 Acupoint Recommend Module

We compute the predicted acupoint activation by explicitly modeling the interaction between symptom representations and acupoint embeddings. Specifically, the multi-hot symptom vector sc is first used to aggregate the symptom embeddings e_s^*, yielding a sample-level symptom representation.

$$e_p = e_s^* \cdot sc. \tag{16}$$

We concatenate the embedding $(e_a^*)^i$ of node i with e_p respectively:

$$(e_p^*)^i = [(e_a^*)^i || e_p], \tag{17}$$

where $(e_p^*)^i \in e_p^*$, which is then fed into the Transformer module:

$$H = \text{softmax}\left(\frac{(e_p^*)^i W_q \cdot ((e_p^*)^i W_k)^\top}{\sqrt{d_h}}\right) \cdot ((e_p^*)^i W_v), \tag{18}$$

where W_q, W_k and W_v are the learnable projection matrices that transform the input embeddings into the query, key, and value representations, and $\sqrt{d_h}$ is used for scaling to stabilize gradients.

A residual connection is incorporated, followed by acupoint recommendation:

$$\hat{a}_i = \text{LN}\big(\text{LN}((e_p^*)^i + H) + \text{FFN}(\text{LN}((e_p^*)^i + H))\big), \tag{19}$$

where $\hat{a}_i$ denotes the output probability of acupoint a_i, that is, $\hat{ac} = \{\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n\}$, LN denotes layer normalization and FFN denotes the feed-forward network.

3.4 Loss Function

The KnowAcu model outputs $\hat{ac} = \{\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n\}$. We define P_a as the set of acupoints with label 1 in $\hat{ac}$ and N_a as the set of acupoints with label 0. For each prescription, we use LLMs to identify the primary therapeutic acupoints for treating sc according to the principal–adjunct acupoint theory, and obtain an acupoint score τ from ac. We first design a loss function $loss_{pn}$ to encourage the predicted probabilities of P_a to be higher than those of N_a.

$$loss_{pn} = \frac{1}{|P_a|} \sum_{\substack{\hat{a}_i \in P_a \\ \hat{a}_j \in N_a}} \tau_i \ln(1 + e^{-\hat{a}_i + \hat{a}_j + \kappa \cdot \tau_i}), \tag{20}$$

where κ controls the separation margin between $\hat{a}_i$ and $\hat{a}_j$. We further encourage acupoints in P_a with higher scores to be prioritized over those with lower scores during recommendation.

$$loss_{pp} = \frac{1}{|P_a|} \sum_{\substack{(\hat{a}_i, \hat{a}_k) \in P_a \\ w_i > w_k}} (\tau_i - \tau_k) \ln(1 + e^{-\hat{a}_i + \hat{a}_k + \kappa \cdot (\tau_i - \tau_k)}). \tag{21}$$

Table 1: Statistics of the evaluation data sets

Data	Prescriptions	Symptoms	Acupoints
real dataset	546	2174	258
synthetic data	26,360	372	325

Subsequently, we generate counter-prediction samples $\hat{x} = \text{LLMs}(\hat{ac})$ using LLMs, yielding the counter-prediction loss.

$$loss_{cp} = -\frac{1}{n}\sum_{i=1}^{n}\left[x_i \log(\hat{x}_i) + (1 - x_i)\log(1 - \hat{x}_i)\right]. \tag{22}$$

Therefore the total loss function is as follows:

$$\mathcal{L} = loss_{pn} + loss_{pp} + \lambda_{cp} loss_{cp} + \lambda_\theta |\theta|_2^2, \tag{23}$$

where $|\theta|_2^2$ denotes the regularization term of learned function $f(sc, P; \theta)$, λ_{cp} and λ_θ denote weight parameters.

4 Experiments

4.1 Dataset

Our dataset was collected from classical acupuncture literature and real-world clinical records, comprising a total of 546 prescriptions, $2,174$ symptoms and 258 acupoints. Each prescription consists of a set of symptoms and a set of acupoints, and we rank the acupoints in the set by their importance. Moreover, we apply data augmentation techniques to generate a synthetic data by Sunsimiao to avoid data leakage, thereby alleviating potential model overfitting caused by the limited size of the real dataset. Synthetic dataset includes $26,360$ prescriptions, 372 symptoms and 325 acupoints. The statistical characteristics of both datasets are reported in Table 1.

4.2 Implementation Details

We train our model with PyTorch, utilizing a single NVIDIA RTX 4090D GPU for the experiments. During training, we employ grid search to optimize the parameters λ_s, λ_a, δ, κ, λ_{cp} and λ_θ. Specifically, learning rate lr is varied in $\{10^{-3}, 10^{-4}, 10^{-5}\}$, Gated and Layer-connected Graph Attention Module number of head is varied in $\{1, 2, 4\}$, Transformer number of layer is searched in $\{1, 2, 4\}$, and the dropout rate is searched in $\{0, 0.1, \ldots, 0.5\}$, the batch size was set to 400, the optimizer was Adam Kinga et al. (2015). Through grid search, our parameter settings are $\lambda_s = 0.3$, $\lambda_a = 0.3$, $\delta = 0.01$, $\kappa = 0.5$, $\lambda_{cp} = 0.9$, and $\lambda_\theta = 0.1$. The datasets were split into training and test sets at a 4:1 ratio, and 5-fold cross-validation was employed to demonstrate the robustness of our model.

4.3 Overall Performance Comparison

To evaluate our approach, we adopt three commonly used evaluation metrics in recommender systems: Precision@K, Recall@K and NDCG@K Wang et al. (2013), K is the cutoff threshold indicating how many top-ranked items are considered. We compare our method with several baseline models, including NFM He & Chua (2017), BPRMF Rendle et al. (2012), NGCF Wang et al. (2019), NCF He et al. (2017), GTC Sun et al. (2025), DGNN Wang et al. (2024), PEGCN Li et al. (2024), 4SDrug Tan et al. (2022), RippleNet Wang et al. (2018), CPGNN Lin et al. (2026) and HARNN Yang et al. (2026). For all baseline methods, the hyperparameter settings follow those reported in their original papers. The comparative results on the real dataset are presented in Table 2 and several observations can be made.

KnowAcu consistently outperforms all competing methods across the evaluation metrics, demonstrating its overall effectiveness. Compared with traditional collaborative filtering models, KnowAcu achieves substantial gains by modeling complex symptom–acupoint associations beyond interaction

Table 2: Performance comparison of different methods on the real dataset (n@K means ndcg@K)

Method	p@5	p@10	p@15	p@20	r@5	r@10	r@15	r@20	n@5	n@10	n@15	n@20
NFM	0.30	0.24	0.20	0.17	0.19	0.30	0.37	0.42	0.37	0.46	0.52	0.56
BPRMF	0.12	0.11	0.10	0.09	0.07	0.12	0.16	0.19	0.21	0.29	0.34	0.37
SMGCN	0.31	0.27	0.25	0.22	0.19	0.31	0.43	0.50	0.31	0.40	0.48	0.53
NGCF	0.09	0.08	0.07	0.06	0.05	0.09	0.11	0.14	0.18	0.24	0.28	0.31
4SDrug	0.35	0.30	0.26	0.22	0.20	0.34	0.43	0.50	0.41	0.42	0.44	0.47
NCF	0.17	0.15	0.13	0.11	0.10	0.17	0.23	0.26	0.25	0.34	0.40	0.44
GTC	0.41	0.34	0.30	0.27	0.19	0.31	0.40	0.48	0.43	0.42	0.44	0.46
DGNN	0.18	0.17	0.15	0.14	0.11	0.20	0.27	0.33	0.19	0.21	0.23	0.26
RippleNet	0.35	0.29	0.25	0.22	0.19	0.31	0.39	0.46	0.36	0.36	0.38	0.40
PEGCN	0.53	0.41	0.34	0.29	0.35	0.51	0.60	0.68	0.57	0.57	0.60	0.63
CPGNN	0.49	0.38	0.32	0.27	0.31	0.46	0.55	0.62	0.60	0.67	0.72	0.77
HARNN	0.25	0.22	0.19	0.16	0.15	0.26	0.33	0.38	0.33	0.42	0.49	0.53
KnowAcu	**0.69**	**0.48**	**0.38**	**0.31**	**0.46**	**0.61**	**0.69**	**0.75**	**0.75**	**0.78**	**0.82**	**0.86**

signals. It also surpasses graph-based methods, particularly at larger K, indicating the effectiveness of the proposed MGJL module in multi-hop information aggregation and noise suppression. Moreover, KnowAcu achieves notable improvements in ranking quality, suggesting that LLM-based acupoint scoring facilitates more accurate prioritization of clinically relevant acupoints and better reflects practical treatment requirements.

We also compared it with the latest large model of traditional Chinese medicine, including Sunsimiao Xin Yan (2023), HuatuoGPT Zhang et al. (2023b) and QWen3 Yang et al. (2025a), and the comparison results are shown in the Table 3 below.

Table 3: Performance comparison of different LLMs on the real dataset (n@K means ndcg@K)

Model	p@5	p@10	p@15	p@20	r@5	r@10	r@15	r@20	n@5	n@10	n@15	n@20
Sunsimiao	0.36	0.22	0.15	0.12	0.22	0.26	0.27	0.27	0.60	0.64	0.65	0.65
HuatuoGPT	0.26	0.14	0.10	0.08	0.15	0.16	0.17	0.17	0.43	0.44	0.45	0.45
QWen3	0.50	0.39	0.29	0.22	0.33	0.50	0.55	0.57	0.62	0.74	0.77	0.78
KnowAcu	**0.69**	**0.48**	**0.38**	**0.31**	**0.46**	**0.61**	**0.69**	**0.75**	**0.75**	**0.78**	**0.82**	**0.86**

KnowAcu consistently achieves the best performance across all evaluation metrics and cutoff levels, demonstrating superior retrieval and ranking effectiveness. Among the baselines, QWen3 performs competitively, particularly at higher recall and NDCG cutoffs, and is therefore adopted as the prompt model due to its strong performance and compatibility with prompt-based tuning. KnowAcu further achieves substantial improvements over QWen3, validating the effectiveness of the proposed MGJL modules in enhancing domain-specific semantic understanding and integrating clinical knowledge for accurate acupuncture recommendation.

Moreover, to verify the robustness of our model, we conducted experiments on a large-scale synthetic data, with the results reported in Table 4, several conclusions can be drawn.

KnowAcu consistently achieves the best Precision, Recall, and NDCG across different K values and both synthetic and real datasets, demonstrating robust and generalizable performance across data distributions. Synthetic data generally yields higher absolute performance for most methods, likely due to its lower noise and clearer structural patterns, whereas the lower Recall@5 can be attributed to the larger number of therapeutic acupoints per case. Notably, KnowAcu shows stronger relative advantages on the real dataset, indicating that its knowledge guidance and multi-graph modeling are particularly effective in more complex clinical settings. Consistent NDCG improvements across both datasets further demonstrate its ability to enhance recommendation ranking quality and maintain robustness under varying data conditions.

4.4 Ablation Study

In this section, we demonstrate the effectiveness of our model by ablating several of its components, with the ablation results summarized in Table 5.

Table 4: Performance comparison of different methods in synthetic dataset (n@K means ndcg@K)

Method	p@5	p@10	p@15	p@20	r@5	r@10	r@15	r@20	n@5	n@10	n@15	n@20
NFM	0.61	0.51	0.44	0.37	0.28	0.47	0.60	0.69	0.66	0.68	0.77	0.83
BPRMF	0.28	0.24	0.21	0.19	0.13	0.22	0.29	0.35	0.35	0.45	0.53	0.59
SMGCN	0.75	0.63	0.51	0.43	0.35	0.59	0.71	0.79	0.75	0.74	0.82	0.87
NGCF	0.22	0.19	0.17	0.16	0.10	0.18	0.24	0.29	0.30	0.41	0.48	0.54
4SDrug	0.80	0.64	0.52	0.43	0.37	0.59	0.71	0.79	0.83	0.72	0.75	0.79
NCF	0.37	0.30	0.25	0.22	0.17	0.27	0.35	0.41	0.46	0.55	0.63	0.68
GTC	0.57	0.45	0.38	0.33	0.26	0.42	0.52	0.61	0.59	0.51	0.54	0.59
DGNN	0.57	0.47	0.39	0.34	0.26	0.43	0.53	0.62	0.61	0.53	0.56	0.61
RippleNet	0.81	0.70	0.68	0.66	0.16	0.25	0.36	0.46	0.76	0.71	0.71	0.72
PEGCN	0.75	0.61	0.50	0.42	0.35	0.56	0.69	0.77	0.78	0.68	0.71	0.76
CPGNN	0.80	0.64	0.52	0.44	0.37	0.60	0.73	0.81	0.82	0.78	0.86	0.91
HARNN	0.76	0.62	0.50	0.42	0.36	0.58	0.70	0.78	0.80	0.77	0.85	0.90
KnowAcu	**0.83**	**0.67**	**0.54**	**0.45**	**0.39**	**0.62**	**0.75**	**0.83**	**0.86**	**0.80**	**0.88**	**0.92**

Table 5: Ablation study results (n@K means ndcg@K)

Method	p@5	p@10	p@15	p@20	r@5	r@10	r@15	r@20	n@5	n@10	n@15	n@20
KnowAcu	**0.69**	**0.48**	**0.38**	**0.31**	**0.46**	**0.61**	**0.69**	**0.75**	**0.75**	**0.78**	**0.82**	**0.86**
w/o Theory Graph	0.66	0.45	0.36	0.29	0.43	0.57	0.64	0.70	0.73	0.76	0.80	0.84
w/o MGJL	0.60	0.43	0.35	0.28	0.39	0.54	0.63	0.69	0.67	0.72	0.77	0.81
w/o LLM loss	0.62	0.45	0.35	0.29	0.41	0.56	0.66	0.71	0.69	0.73	0.78	0.82
w/o all components	0.52	0.36	0.29	0.24	0.33	0.47	0.54	0.58	0.59	0.66	0.71	0.74

In Table 5, "w/o Theory Graph" denotes "without SS^z and AA^m graph", "w/o LLM loss" denotes "without $loss_{pp}$, $loss_{pn}$ and $loss_{cp}$". More ablation results, such as "w/o SS^z", "w/o AA^m", and "w/o $loss_{pp}$", are reported in Appendix. Based on the ablation results, the following conclusions can be drawn.

The complete KnowAcu model consistently achieves the best Precision, Recall, and NDCG across different K values, confirming the effectiveness of its integrated components. Removing the Theory Graph, MGJL, or LLM loss consistently degrades performance, demonstrating their respective contributions to TCM theory modeling, multi-granularity knowledge fusion, and knowledge-constrained representation learning. In particular, removing the LLM loss substantially affects NDCG, highlighting its role in distinguishing primary from auxiliary acupoints and improving recommendation ranking. When all key components are removed, performance drops markedly, further validating the necessity and complementary effects of the overall framework.

5 Conclusion

This paper introduces KnowAcu, a knowledge-guided framework that redefines acupoint recommendation as a structured medical reasoning task. By synergizing TCM theoretical graphs with an MGJL module, KnowAcu effectively captures multi-granular clinical features while mitigating over-smoothing. Furthermore, we integrate LLM-distilled expertise into the optimization process to prioritize principal acupoints with high interpretability. Both quantitative benchmarks and qualitative case studies demonstrate that KnowAcu significantly outperforms state-of-the-art baselines, offering a robust, privacy-preserving, and clinically-consistent solution for digitalized TCM.

References

Machaon Bonafede, Andrew Dick, Katia Noyes, Jonathan D Klein, and Theodore Brown. The effect of acupuncture utilization on healthcare utilization. *Medical care*, 46(1):41–48, 2008.

Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1358–1368, 2019.

Guoxuan Chen, Lianghao Xia, and Chao Huang. Lightgnn: Simple graph neural network for recommendation. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pp. 549–558, 2025.

David Cyranoski. Why chinese medicine is heading for clinics around the world. *Nature*, 561(7724): 448–448, 2018.

Wenqi Fan, Xiaorui Liu, Wei Jin, Xiangyu Zhao, Jiliang Tang, and Qing Li. Graph trend filtering networks for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pp. 112–121, 2022.

Nanna Goldman, Michael Chen, Takumi Fujita, Qiwu Xu, Weiguo Peng, Wei Liu, Tina K Jensen, Yong Pei, Fushun Wang, Xiaoning Han, et al. Adenosine a1 receptors mediate local anti-nociceptive effects of acupuncture. *Nature neuroscience*, 13(7):883–888, 2010.

Xiangnan He and Tat-Seng Chua. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 355–364, 2017.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pp. 173–182, 2017.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 639–648, 2020.

Yan Jiang and Jun Zou. Analysis of the tcm theory of traditional chinese health exercise. *Journal of Sport and Health Science*, 2(4):204–208, 2013.

Yuanyuan Jin, Wendi Ji, Yao Shi, Xiaoling Wang, and Xiaochun Yang. Meta-path guided graph attention network for explainable herb recommendation. *Health Information Science and Systems*, 11(1):5, 2023.

Diederik Kinga, Jimmy Ba Adam, et al. A method for stochastic optimization. In *International conference on learning representations (ICLR)*, volume 5. California;, 2015.

Taeyong Kong, Taeri Kim, Jinsung Jeon, Jeongwhan Choi, Yeon-Chang Lee, Noseong Park, and Sang-Wook Kim. Linear, or non-linear, that is the question! In *Proceedings of the fifteenth ACM international conference on web search and data mining*, pp. 517–525, 2022.

Hiu Yu Cherie Leung, Pou Kuan Leong, Jihang Chen, Kam Ming Ko, et al. Inter-organ relationships among gut, lung and skin beyond the pathogenesis of allergies: relevance to the zang-fu theory in chinese medicine. *Chinese Medicine*, 8(03):73, 2017.

Ruiling Li, Song Wu, Jinyu Tu, Limei Peng, and Li Ma. An enhanced graph convolutional network with property fusion for acupoint recommendation. *Applied Intelligence*, 54(22):11536–11546, 2024.

Wei Lin, Jiaqi Chen, Zhuo Chen, Tao Yin, Jing Guo, and Fang Zeng. Modeling symptom-acupoint interactions via a heterogeneous graph learning framework for intelligent acupoint recommendation. *Chinese Medicine*, 21(1):195, 2026.

Can Liu, Kaijie Sun, Qingqing Zhou, Yuchen Duan, Jianhua Shu, Hongxing Kan, Zongyun Gu, and Jili Hu. Cpmi-chatglm: parameter-efficient fine-tuning chatglm with chinese patent medicine instructions. *Scientific reports*, 14(1):6403, 2024.

Fan Liu, Zhiyong Cheng, Lei Zhu, Zan Gao, and Liqiang Nie. Interest-aware message-passing gcn for recommendation. In *Proceedings of the web conference 2021*, pp. 1296–1305, 2021.

Qingqing Long, Lingjun Xu, Zheng Fang, and Guojie Song. Hgk-gnn: Heterogeneous graph kernel based graph neural networks. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 1129–1138, 2021.

Chengsheng Mao, Liang Yao, and Yuan Luo. Medgcn: Medication recommendation and lab test imputation via graph convolutional networks. *Journal of Biomedical Informatics*, 127:104000, 2022.

Himel Mondal, Satyalakshmi Komarraju, Shrikanth Muralidharan, et al. Assessing the capability of large language models in naturopathy consultation. *Cureus*, 16(5), 2024.

NIH NIH. Consensus conference. acupuncture. *Jama*, 280(17):1518–24, 1998.

Dario Paladini, Gustavo Malinger, Gianluigi Pilu, Ilan Timor-Trisch, and Paolo Volpe. The meridian trial: caution is needed. *The Lancet*, 389(10084):2103, 2017.

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.

Zezeng Rong, Mingxu Cheng, Shupeng Lin, Jiyong Lin, and Huiling Chen. Chinese medicine prescription recommendation using generative adversarial network. In *International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 1–6. IEEE, 2022.

Yundong Sun, Dongjie Zhu, Yansong Wang, Yansheng Fu, and Zhaoshuo Tian. Gtc: Gnn-transformer co-contrastive learning for self-supervised heterogeneous graph representation. *Neural Networks*, 181:106645, 2025.

Yanchao Tan, Chengjun Kong, Leisheng Yu, Pan Li, Chaochao Chen, Xiaolin Zheng, Vicki S Hertzberg, and Carl Yang. 4sdrug: Symptom-based set-to-set small and safe drug recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3970–3980, 2022.

Yang Tan, Zhixing Zhang, Mingchen Li, Fei Pan, Hao Duan, Zijie Huang, Hua Deng, Zhuohang Yu, Chen Yang, Guoyang Shen, et al. Medchatzh: A tuning llm for traditional chinese medicine consultations. *Computers in Biology and Medicine*, 172:108290, 2024.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

Fei Wang, Yansheng Li, Yongjun Zhang, and Dong Wei. Klgcn: Knowledge graph-aware light graph convolutional network for recommender systems. *Expert Systems with Applications*, 195: 116513, 2022.

Guang-Jun Wang, M Hossein Ayati, and Wei-Bo Zhang. Meridian studies in china: a systematic review. *Innovations in Acupuncture and Medicine*, 3(1):1–9, 2010.

Haochun Wang, Chi Liu, Nuwa Xi, Zewen Qiang, Sendong Zhao, Bing Qin, and Ting Liu. Huatuo: Tuning llama model with chinese medical knowledge. *arXiv preprint arXiv:2304.06975*, 2023a.

Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 417–426, 2018.

Jinlu Wang, Jipeng Guo, Yanfeng Sun, Junbin Gao, Shaofan Wang, Yachao Yang, and Baocai Yin. Dgnn: Decoupled graph neural networks with structural consistency between attribute and graph embedding representations. *IEEE Transactions on Big Data*, 2024.

Rongsheng Wang, Yaofei Duan, ChanTong Lam, Jiexin Chen, Jiangsheng Xu, Haoming Chen, Xiaohong Liu, Patrick Cheong-Iao Pang, and Tao Tan. Ivygpt: Interactive chinese pathway language model in medical domain. In *CAAI International Conference on Artificial Intelligence*, pp. 378–382. Springer, 2023b.

Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 165–174, 2019.

Yining Wang, Liwei Wang, Yuanzhi Li, Di He, and Tie-Yan Liu. A theoretical analysis of ndcg type ranking measures. In *Conference on learning theory*, pp. 25–54. PMLR, 2013.

Zhe Wang, Keqian Li, Quanying Ren, Keyu Yao, and Yan Zhu. Traditional chinese medicine formula classification using large language models. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 4647–4654. IEEE, 2023c.

Dong Xue* Xin Yan. Sunsimiao: Chinese medicine llm. `https://github.com/thomas-yanxin/Sunsimiao`, 2023.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025a.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025b.

Ding Yang and Yufang Ding. Smrgat: A traditional chinese herb recommendation model based on a multi-graph residual attention network and semantic knowledge fusion. *Journal of Ethnopharmacology*, 315:116693, 2023.

Guoxing Yang, Xiaohong Liu, Jianyu Shi, Zan Wang, and Guangyu Wang. Tcm-gpt: efficient pretraining of large language models for domain adaptation in traditional chinese medicine. *Computer Methods and Programs in Biomedicine Update*, 6:100158, 2024a.

Hang Yang, Ren Wu, Mitsuru Nakata, and Qi-wei Ge. A hierarchical and interpretable machine learning model for acupoint determination. *Journal of Integrative Medicine*, 2026.

Songhua Yang, Hanjie Zhao, Senbin Zhu, Guangyu Zhou, Hongfei Xu, Yuxiang Jia, and Hongying Zan. Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 19368–19376, 2024b.

Yun Yang, Yulong Rao, Minghao Yu, and Yan Kang. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neural Networks*, 146:1–10, 2022.

Chenyan Zhang, Shan Xue, Jing Li, Jia Wu, Bo Du, Donghua Liu, and Jun Chang. Multi-aspect enhanced graph neural networks for recommendation. *Neural Networks*, 157:90–102, 2023a.

Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 793–803, 2019.

Heyi Zhang, Xin Wang, Zhaopeng Meng, Zhe Chen, Pengwei Zhuang, Yongzhe Jia, Dawei Xu, and Wenbin Guo. Qibo: A large language model for traditional chinese medicine. *arXiv preprint arXiv:2403.16056*, 2024.

Hong Zhang, Jiajun Zhang, Wandong Ni, Youlin Jiang, Kunjing Liu, Daying Sun, and Jing Li. Transformer-and generative adversarial network–based inpatient traditional chinese medicine prescription recommendation: Development study. *JMIR Medical Informatics*, 10(5):e35239, 2022.

Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Guiming Chen, Jianquan Li, Xiangbo Wu, Zhang Zhiyi, Qingying Xiao, et al. Huatuogpt, towards taming language model to be a doctor. In *Findings of the association for computational linguistics: EMNLP 2023*, pp. 10859–10885, 2023b.

Yingying Zhang, Xian Wu, Quan Fang, Shengsheng Qian, and Changsheng Xu. Knowledge-enhanced attributed multi-task learning for medicine recommendation. *ACM Transactions on Information Systems*, 41(1):1–24, 2023c.

Mingqiang Zhou, Kunpeng Li, Kailang Dai, and Quanwang Wu. Hin-based rating prediction in recommender systems via gcn and meta-learning. *Applied Intelligence*, 53(20):23271–23286, 2023.

Wuai Zhou, Kuo Yang, Jianyang Zeng, Xinxing Lai, Xin Wang, Chaofan Ji, Yan Li, Peng Zhang, and Shao Li. Fordnet: recommending traditional chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacological research*, 173:105752, 2021.

Lingxuan Zhu, Weiming Mou, Yancheng Lai, Junda Lin, and Peng Luo. Language and cultural bias in ai: comparing the performance of large language models developed in different countries on traditional chinese medicine highlights the need for localized models. *Journal of translational medicine*, 22(1):319, 2024.

A Appendix

B Policy Statement

We adhered to the principle of informed consent, and all participants provided consent for data collection. The privacy and right to self-determination of participants and subjects were strictly protected, with all personal identifying information removed from the dataset. In addition, this study complied with all other applicable ICLR ethical principles.

C Supplementary Ablation Study Results

Supplementary ablation study results, as shown in Table 6, demonstrate the outcomes of individual ablation experiments, thereby validating the effectiveness of our model.

D Influence of Hyperparameters

The main hyperparameters of our model include λ_s, λ_a, δ, and κ, with $\lambda_s = 0.3$, $\lambda_a = 0.3$, $\delta = 0.01$, and $\kappa = 0.5$. Here, λ_s and λ_a primarily control the weights of the graph relations derived from zang–fu organ and meridian theories, δ determines the degree of graph pruning, and κ regulates the separation among different samples. The experimental results under varying hyperparameter settings are shown in the Figure 2. Performance is optimal when λ_s and λ_a are set to 0.3; values that are too high or too low slightly degrade performance, indicating that symptom weights significantly affect both model precision and recall. Excessively high values suppress the contribution of acupoint information, whereas excessively low values lead to insufficient theoretical guidance. The optimal δ is 0.01, suggesting that moderate pruning effectively removes redundant information and improves model performance.

E Supplementary Case Studies

In this section, we present representative case studies in Table 7 to demonstrate the clinical utility of KnowAcu. In the "Acupoint Set" column, dark green highlights denote matches with the clinical ground truth, while blue and cyan signify principal and adjunct acupoints, respectively. By synergizing the Zang-fu and Meridian graphs, KnowAcu identifies latent syndrome patterns even when clinical manifestations are highly diverse.

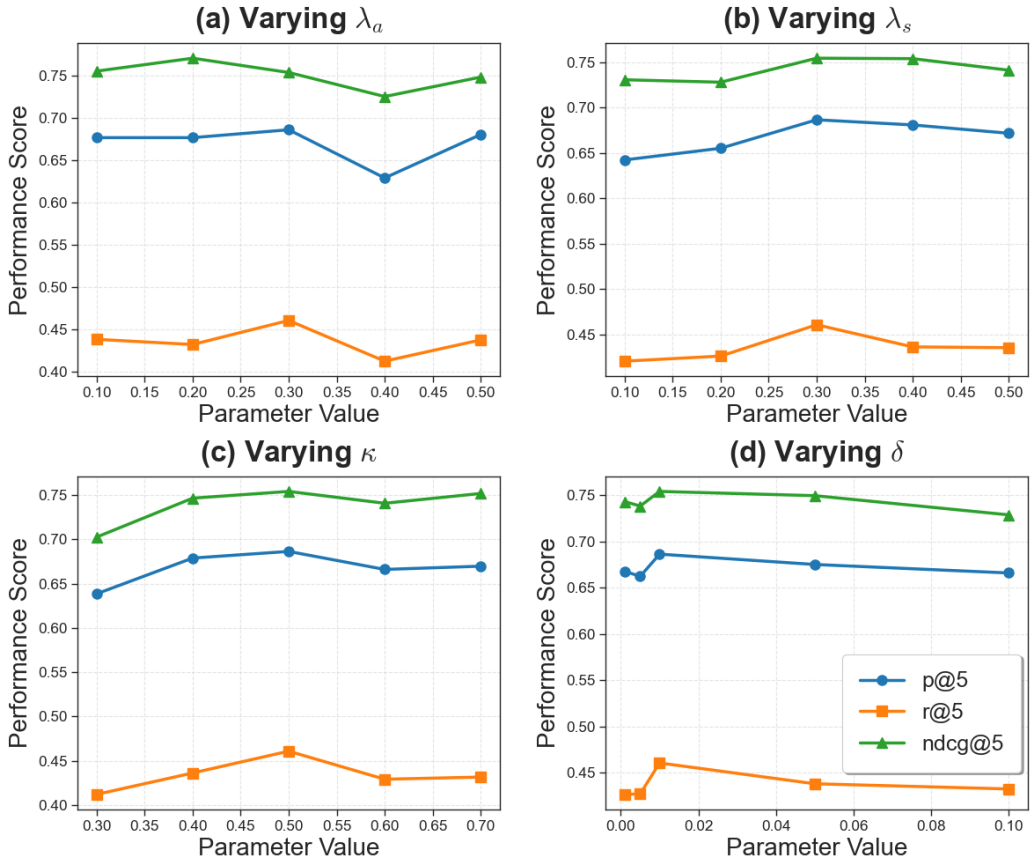


Figure 2: Influence of Hyperparameters

F Theory of Acupuncture

We consulted experts in the field of acupuncture and developed the Zang-fu relationship graph SS^z and the meridian relationship graph AA^m. A brief overview of the relevant theories is provided below.

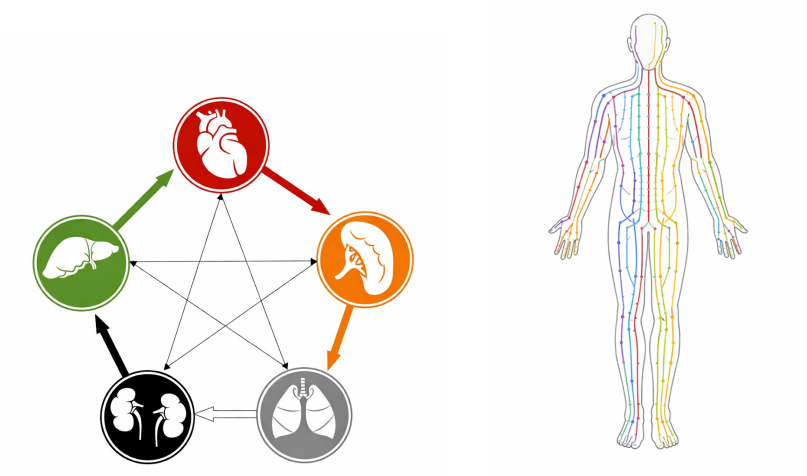
F.1 Zang-Fu Theory

Traditional Chinese Medicine (TCM) is founded on the theory of Zang-fu Leung et al. (2017); Jiang & Zou (2013), which provides a systematic framework for understanding human physiology and pathology, shows in Figure 3a. The five zang organs—Heart, Liver, Spleen, Lung, and Kidney—are responsible for storing vital substances and maintaining fundamental life activities, whereas the six fu organs—Gallbladder, Stomach, Small Intestine, Large Intestine, Bladder, and Sanjiao—primarily function in the reception, transformation, and transportation of food, fluids, and waste.

Within this theoretical system, symptoms are viewed as external manifestations of internal functional disturbances of specific zang–fu organs or their interconnections. Each organ system is associated with characteristic physiological roles and pathological patterns, dysfunction in these roles leads to corresponding symptom clusters. Therefore, symptom analysis in TCM emphasizes holistic pattern recognition rather than isolated signs, enabling the inference of underlying organ imbalances and guiding diagnosis and treatment.

F.2 Meridian Theory

The meridian theory Wang et al. (2010); Paladini et al. (2017) shows in Figure 3b is a core component of Traditional Chinese Medicine (TCM), describing an internal network that connects or-



(a) Zang-Fu Theory (b) Meridian Theory

Figure 3: Acupuncture Theory

Table 6: Supplementary ablation study results

Method	p@5	p@10	p@15	p@20	r@5	r@10	r@15	r@20	n@5	n@10	n@15	n@20
KnowAcu	**0.69**	**0.48**	**0.38**	**0.31**	**0.46**	**0.61**	**0.70**	**0.75**	**0.75**	**0.78**	**0.83**	**0.86**
w/o SS^z	0.66	0.47	0.38	0.31	0.44	0.57	0.65	0.70	0.73	0.75	0.80	0.84
w/o AA^m	0.66	0.47	0.38	0.31	0.43	0.57	0.65	0.71	0.72	0.75	0.80	0.83
w/o loss_pn	0.65	0.45	0.36	0.30	0.42	0.57	0.65	0.71	0.72	0.75	0.80	0.83
w/o loss_pp	0.63	0.43	0.36	0.29	0.41	0.56	0.65	0.71	0.69	0.73	0.78	0.82
w/o loss_cp	0.65	0.46	0.37	0.30	0.43	0.57	0.65	0.70	0.72	0.75	0.80	0.84

gans, limbs, and sensory structures into an integrated functional system. Meridians are conceptual pathways through which qi and blood circulate, ensuring physiological coordination and systemic balance across the body.

According to this theory, the meridian system consists primarily of twelve regular meridians, each corresponding to a specific zang or fu organ, along with extraordinary vessels and collateral branches. Pathological changes in organs are believed to be transmitted along meridians and manifested as localized or systemic symptoms, while therapeutic stimulation at specific acupoints can modulate meridian activity to restore functional homeostasis. This framework provides the theoretical basis for acupuncture and supports symptom-oriented diagnosis and intervention in TCM.

F.3 Primary–Auxiliary Acupoint Theory

In Traditional Chinese Acupuncture, the primary–auxiliary acupoint theory refers to a structured principle for acupoint selection in clinical treatment.

Specifically, primary acupoints (main acupoints) are selected according to the core pathogenesis and dominant symptoms of a disease and play a leading therapeutic role, whereas auxiliary acupoints are combined to reinforce, modulate, or target secondary symptoms through synergistic regulation of meridians, zang–fu organs, and qi–blood dynamics. This combinatorial strategy emphasizes coordinated action rather than isolated acupoint effects and reflects the holistic and hierarchical nature of acupuncture prescriptions.

From a theoretical perspective, the primary–auxiliary acupoint framework embodies key Traditional Chinese Medicine principles, including syndrome differentiation, meridian correspondence, and functional complementarity, and serves as a foundational guideline for constructing effective and interpretable acupuncture treatment protocols.

G BASELINE DETAILS

- NFM He & Chua (2017).NFM performs the prediction task by integrating the linear second-order feature interaction modeling of factorization machines with the non-linear capability of neural networks to capture higher-order feature interactions.
- BPRMF Rendle et al. (2012). BPRMF for personalized ranking is the maximum a posteriori estimator obtained through a Bayesian analysis of the problem.
- NGCF Wang et al. (2019). NGCF is a novel recommendation framework inspired by the neural factorization machine (NFM) He & Chua (2017), which explicitly injects high-order user–item connectivity into embedding learning through recursive graph-based message propagation to enhance collaborative filtering performance.
- NCF He et al. (2017). NCF aimes to devise neural network-based methods to address the central challenge in recommendation — collaborative filtering — using implicit feedback.
- GTC Sun et al. (2025). GTC is a self-supervised heterogeneous graph representation framework that collaboratively integrates GNN-based local schema encoding and Transformer-based multi-hop semantic modeling through cross-view contrastive learning to effectively mitigate oversmoothing and enhance long-range information capture.
- DGNN Wang et al. (2024). DGNN learns comprehensive node representations by decoupling the exploration of attribute and graph spaces into distinctive embeddings while enforcing structural consistency between their reconstructed adjacency matrices using a shared latent factor.
- PEGCN Li et al. (2024). PEGCN introduces an enhanced graph convolutional network with property fusion, which integrates acupoint attribute embeddings learned by a property feature graph encoder and employs an attention-enhanced GCN to capture complex symptom–acupoint relationships.
- 4SDrug Tan et al. (2022). 4SDrug enables set-to-set comparison via symptom/drug set representations and similarity metrics. It aggregates symptom sets by importance for accurate representation, and refines drug sets via intersection-based augmentation—applying knowledge- and data-driven penalties to ensure safety.
- RippleNet Wang et al. (2018). RippleNet is an end-to-end framework that treats a user's historically clicked items as seeds in a knowledge graph and iteratively propagates preferences along links to discover hierarchical potential interests for recommendation.
- CPGNN Lin et al. (2026). CPGNN introduced an acupoint compatibility prediction framework with graph neural networks (GNN) and fine-tuned bidirectional encoder representations from transformers (termed GNN-BERT-Attention).
- HARNN Yang et al. (2026). HARNN preprocessed textual AMT data to build an acupoint prescription database designed for machine learning applications.

H PROMPT TEMPLATE OF LLMS

We utilized the Qwen3 Yang et al. (2025b) LLM in three distinct aspects: (1) Table 8 illustrates the prompt template used to generate a counter-prediction mapping matrix based on general TCM knowledge, enabling the inference of possible symptom sets from a given acupoint set. (2) Table 9 illustrates the prompt template used to score each acupoint in a prescription based on the principal–auxiliary acupoint theory, thereby distinguishing the relative importance of acupoints; and (3) Table 10 illustrates the prompt template used to generate a synthetic dataset based on general TCM knowledge, producing acupoint treatment plans corresponding to real symptom sets.

Table 7: Example recommended acupoint set for a given symptom set on a real dataset

Symptoms	Recommended Acupoint Set	Ground Truth
Difficulty in raising the arm, limited movement, right shoulder pain, difficulty extending and abducting, heat-relief pain, aversion to wind and cold, difficulty dressing, combing hair, touching back, cold pain in the lower back.	Jianzhen (SI9), Jianqian (EX-UE12), Quchi (LI11), Jianyu (LI15), Waiguan (SJ5), Hegu (LI4), Jianliao (SJ14), Ashi point, Fengchi (GB20), Tianzong (SI11), Quyuan (SI13), Shousanli (LI10), Lieque (LU7), Tianzhu (BL10), Jianwaishu (SI14).	Jianzhen (SI9), Jianqian (EX-UE12), Jianyu (LI15), Ashi point, Quchi (LI11), Waiguan (SJ5), Hegu (LI4), Fengchi (GB20), Tianzhu (BL10), Jianliao (SJ14), Tianzong (SI11), Shousanli (LI10), Jingjiaji (Extra), Jianwaishu (SI14), Jianjing (GB21), Jianzhongshu (SI15), Quyuan (SI13).
Full-body convulsions, acute onset, rapid breathing, high fever, head shaking, tongue rolling, agitation, phlegm-heat seizure, unconscious, wiry-slippery pulse, floating-rapid pulse, yellow tongue coating.	Shuigou (GV26), Hegu (LI4), Yintang (EX-HN3), Taichong (LR3), Shijing points (12 wells), Chize (LU5), Shixuan (EX-UE), Fenglong (ST40), Dazhui (GV14), Neiguan (PC6).	Hegu (LI4), Shuigou (GV26), Yintang (EX-HN3), Taichong (LR3), Fenglong (ST40), Zhongwan (CV12).
Sweating, cyanotic lips and nails, cold pathogen stagnation, palpitations, stabbing or squeezing chest pain, shortness of breath, anxiety, chest tightness, choppy pulse, purple tongue, triggered by cold, pale complexion.	Shanzhong (CV17), Yinxi (HT6), Ximen (PC4), Neiguan (PC6), Fenglong (ST40), Qimen (LR14), Zhiyang (GV9), Zhongwan (CV12), Taichong (LV3).	Shanzhong (CV17), Ximen (PC4), Yinxi (HT6), Neiguan (PC6), Shenque (CV8), Zhiyang (GV9)
Vomiting, dizziness, palpitations, alternating cold and heat, phlegm-fluid obstruction, slippery pulse, epigastric fullness, poor appetite, white greasy coating.	Neiguan (PC6), Weishu (BL21), Zusanli (ST36), Zhongwan (CV12), Ganshu (BL18), Yinlingquan (SP9), Qihai (CV6), Gongsun (SP4), Yanglingquan (GB34), Fenglong (ST40), Sanyinjiao (SP6).	Weishu (BL21), Neiguan (PC6), Zusanli (ST36), Zhongwan (CV12), Shanzhong (CV17), Fenglong (ST40)
Dizziness, headache, head fullness/pressure, heaviness of the head, vertigo, blurred vision, tinnitus.	Fengchi (GB20), Taichong (LR3), Hegu (LI4), Quchi (LI11), Sanyinjiao (SP6), Baihui (GV20), Xingjian (LR2), Zusanli (ST36), Ashi point, Ganshu (BL18).	Fengchi (GB20), Hegu (LI4), Taiyang (EX-HN5), Baihui (GV20), Ashi point, Taichong (LR3), Quchi (LI11), Sanyinjiao (SP6), and Touwei (ST8).
Posterior occipital pain, neck and back pain, limb dysfunction, neck stiffness, neck rigidity, cervical pain.	Ashi point, Tianzhu (BL10), Houxi (SI3), Shenmai (BL62), Lieque (LU7), Jingjiaji (EX), Hegu (LI4), Xiaohai (SI8).	Ashi point, Tianzhu (BL10), Houxi (SI3), Shenmai (BL62), Jingjiaji (EX), Fengfu (GV16), Kunlun (BL60)
head and face erythema and blisters with electric-like pain.	Feishu (BL13), Fengmen (BL12), Jiajizhi (BL), Ashi point, Back three needles, Hegu (LI4), Temporal three needles, Ximen (PC4), Quchi (LI11), Dazhu (BL11), Shenshu (BL23), Zhongfu (LU1), Waiguan (SJ5), Jiexi (ST41).	Fengmen (BL12), Feishu (BL13), Jiajizhi (BL), Ashi point, Hegu (LI4), Ximen (PC4), Waiguan (SJ5), Qimen (LR14), Dazhu (BL11), Riyue (LR13)

Table 8: Prompt Template for Counter-prediction Mapping Matrix

Prompt Template for Counter-prediction Mapping Matrix
[User] You are a Traditional Chinese Medicine (TCM) expert. Now you will take a closed-book exam. Based on general TCM knowledge, identify the main symptoms that the acupoint "{acup}" is used to treat and assign a confidence score (0.0–1.0). **Reference symptom list**: {symptom_list} **Requirements**: 1. Only include symptoms you consider effective. 2. Provide scores with two decimal places. **Mandatory output**: Output only a single valid JSON string. Example format: {"headache": 0.95, "stomachache": 0.60}

Table 9: Prompt Template for Acupoint Score

Prompt Template for Acupoint Score
[User] You are a TCM expert well-versed in the "Principal–Auxiliary Acupoint Theory". **Current Task**: For the given symptoms {symptom_list}, score the acupoints in the prescription {acu_str} according to their importance in treating these symptoms. Use space to separate the scores and do not output any other content. **Reference symptom list**: {symptom list} **Requirements**: 1. Only include symptoms you consider effective. 2. Provide scores with two decimal places. **Mandatory output**: Output only a single valid JSON string. Example format: {"headache": 0.95, "stomachache": 0.60}

Table 10: Prompt Template for Synthetic Dataset

Prompt Template for Synthetic Dataset
[User] You are a TCM expert proficient in integrating acupuncture and herbal medicine. I will provide a set of {clinical symptoms} and the corresponding {herbal prescription}. Please select the most appropriate acupoints from the provided {valid acupoint list} based on the principle of "using acupoints to replace herbs", and rank them according to their importance. {Valid Acupoint List} (only select from here): {valid_acups_str} {Symptoms}: symptoms {Reference Herbs}: herbs **Current Task**: 1. Analyze the effects of the herbal prescription. 2. Generate an acupuncture prescription using only names from the {valid acupoint list}. 3. Include 5–12 acupoints, covering both principal and auxiliary points as much as possible. 4. Output only the acupoint names, separated by spaces, with no explanations. **Output Example**: Zusanli Hegu Taichong Fengchi Baihui **Your Answer**: