

Article

Not peer-reviewed version

A Data-Driven Strategy for Profiling of TCM Herb Properties by Network Pharmacology and Deep Learning

[Jingyi Jian](#) , [Wenping Huang](#) , Tingting Zhang , [Erwin Adams](#) , [Ann Van Schepdael](#) *

Posted Date: 27 February 2026

doi: 10.20944/preprints202602.1543.v1

Keywords: traditional Chinese medicine property; *KEGG* enrichment analysis; deep learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Data-Driven Strategy for Profiling of TCM Herb Properties by Network Pharmacology and Deep Learning

Jingyi Jian ¹, Wenping Huang ¹, Tingting Zhang ², Erwin Adams ¹ and Ann Van Schepdael ^{1,*}

¹ KU Leuven - University of Leuven, Pharmaceutical Analysis, Department of Pharmaceutical and Pharmacological Sciences, O&N2, PB 923, Herestraat 49 3000 Leuven, Belgium

² Institute of Pharmaceutical Analysis/International Cooperative Laboratory of Traditional Chinese Medicine Modernization and Innovative Drug Development of Ministry of Education (MOE) of China, College of Pharmacy, Jinan University, Guangzhou 510632, PR China

* Correspondence: ann.vanschepdael@kuleuven.be

Abstract

Extensive experimental data on Traditional Chinese Medicine are available in literature and databases. However, many studies focus on specific diseases or pathways with small sample sizes. As a result, the fundamental pharmacological basis underlying TCM herb properties remains insufficiently elucidated. Based on the concept of the multi-component, multi-target, multi-pathway network of TCM, a data-driven strategy was developed for the profiling of TCM herb properties through network pharmacology and deep learning, facilitating the exploration of the scientific evidence underlying TCM herb properties. Large-scale ingredient and target data of TCM herbs were curated from the *HERB2.0* database. *KEGG* pathway enrichment was conducted for each herb with relative frequency profiling of distinct property groups. Deep learning models were developed and optimized for classification with visual explanation. As a result, high-relative frequency pathways were highly concentrated in five systems (endocrine, immune, nervous, signal transduction, cell growth and death) of *KEGG*. Herbs with distinct properties exhibited a V-shaped trend (Hot>Warm>Neutral<Cool<Cold) in terms of the abundance of ingredients, targets and high-frequency pathways. The HeteroGAT model improved classification accuracy and provided visual explanations at the ingredient–target–pathway level. We demonstrated a viable strategy to profile TCM property classification from a holistic perspective on ingredients, targets, and pathways, which could help elucidate the scientific basis of TCM properties. However, further advances in model refinement and data matrices are required to enhance the effectiveness of this strategy.

Keywords: traditional Chinese medicine property; *KEGG* enrichment analysis; deep learning

1. Introduction

Traditional Chinese Medicine (TCM) theory is rooted in the principles of “holism” and “syndrome differentiation”, emphasizing the dynamic balance in the human body. According to TCM theory, diseases are conceptualized through hot syndromes and cold syndromes [1]. In order to achieve syndrome-specific treatment, the property (also known as “Qi” or “Nature”) of TCM herbs is classified into: cold/cool for hot syndromes and hot/warm for cold syndromes [2,3]. However, the mechanisms underlying TCM herbs’ property classification remain insufficiently elucidated [4]. Such limitations constrain the advancement of quality evaluation for TCM herbs, which are currently based on quantifying selected or abundant components rather than herbs’ pharmacological effects in the context of TCM theory.

Based on the synergistic interaction of the “multi-component, multi-target, multi-pathway”-paradigm of TCM [5], emerging evidence from multiple studies provides mechanistic insights

through system biology approaches. Studies on the therapeutic mechanisms of herbs have amassed extensive data on bioactive ingredients and their molecular targets. Network pharmacology [6] for these ingredients and targets facilitates a clarification of how herbs achieve therapeutic effects at the pathway level. To bridge TCM classifications with modern biological frameworks, from the perspective of hot/cold syndromes, studies [1,7] have investigated the omics characteristics such as gene expression and metabolic profiles of hot or cold syndromes. Regarding herbal medicines, the common characteristics and distinctions among herbs with different properties are being actively investigated, focusing on the chemical structural features [8] and pathway-level regulatory effects of TCM herbs [9,10]. Moreover, machine/deep learning models have been developed for TCM herbs property classification tasks [11,12] to support standardized quality control.

However, many current researches through network pharmacology for TCM, with usually small or medium sample sizes, target a specific disease in biomedicine to illustrate the mechanisms of TCM prescriptions or herbs, offering limited insight into the fundamental principles and conceptual foundations of TCM. In addition, classification models of TCM herbs mainly use only ingredient features, and currently lack a comprehensive application integrating ingredient, target and pathway of TCM herbs. Furthermore, many classification studies reduce the label system to a binary task by merging Hot-Warm and Cold-Cool, although in TCM theory these categories are not fully equivalent, which may oversimplify the representation of herbal properties.

In this study, based on the principle “multi-component, multi-target, multi-pathway”, we propose a strategy for the profiling of TCM herbs property combining network pharmacology and deep learning. After *KEGG* pathway enrichment for each TCM herb from a large sample size, results were analyzed across groups with different property labels to reveal the biological insights of TCM classification systems. Afterwards, classification models were developed by deep learning, with herbs categorized into five property classes (hot, warm, neutral, cool, and cold), applying ingredient, target and pathway data for pattern recognition and interpretability analysis.

2. Results and Discussions

2.1. Overview of Data Characteristics

HERB2.0 is a comprehensive database that integrates information on the ingredients and targets of TCM herbs, derived from literature, high-throughput experiments and cross-referenced with other major TCM databases such as *SymMap*, *TCM-ID* and *TCMSP*. This integration of diverse sources helps reduce potential source-specific bias. Herbs with TCM property labels were categorized into five groups: hot, warm, neutral, cool, and cold, following TCM theory where hot/warm herbs exhibit therapeutic efficacy in treating cold syndromes, and cold/cool herbs exhibit efficacy in treating hot syndromes, while neutral herbs exhibit minimal bias. Through the *HERB2.0* ingredient set (see S1 in the supplementary material), the number of property-specific herbs with successfully enriched pathways employing target sets of *HERB2.0* is presented in Table 1. Relative frequencies of all targets and pathways in distinct property groups were listed in S2 and S3.

Table 1. Number of herbs with distinct property labels employing *HERB2.0* target sets.

Properties	Number of herbs
Hot	36
Warm	530
Neutral	256
Cool	255
Cold	504

2.2. Property Profiling Through Network Pharmacology

When applying the *HERB2.0* ingredient and target sets, the box plots (Figure 1) show the mean and median of the number of targets of each herb group and the distribution of herbs in each group. Both figure panels exhibit a V-shaped trend: the abundances of ingredients and targets were observed to progressively increase, forming a “hot > warm > neutral < cool < cold” hierarchy.

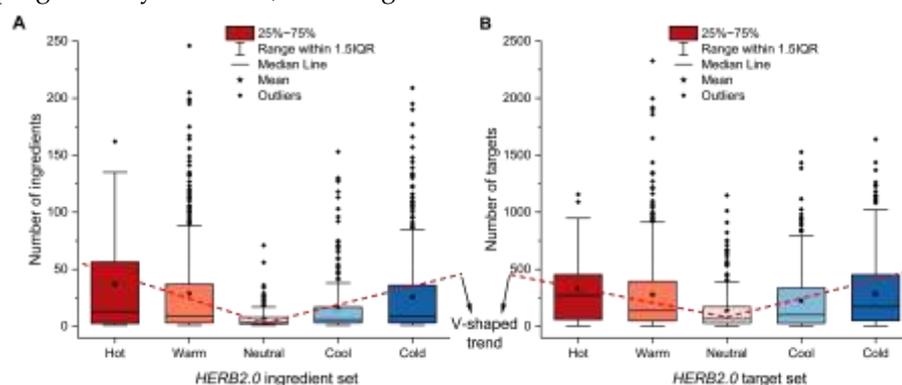


Figure 1. Target abundances of herbs with distinct properties employing (A) *HERB2.0* ingredient set, (B) *HERB2.0* target set.

The heatmap in Figure 2 was constructed to organize enriched pathways by primary and secondary systems of the *KEGG*, when using the *HERB2.0* target set, with the primary systems listed on top of the figure. Pathways within each secondary system were sorted by their relative frequency across all herbs. The various color-coded heatmaps represent the relative frequency of pathways within the five groups of TCM properties.

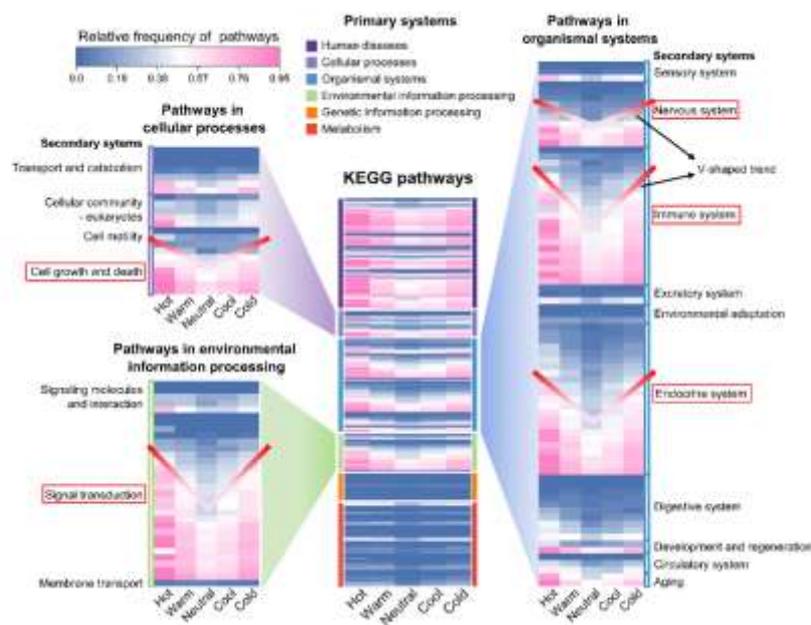


Figure 2. Relative frequency profiling of pathways for property groups employing *HERB2.0* target set.

At the top of the central heatmap, most pathways belonging to the primary system “Human diseases” show high relative frequency, which might be attributed to the extensive focus on disease-related research resulting in abundance of targets identified and pathway annotations. The other high-relative frequency pathways clustered predominantly within five secondary systems: Endocrine system, Immune system, Nervous system, Signal transduction, and Cell growth and death, revealing that TCM herbs exert their therapeutic effects primarily through the multi-pathway regulation within these five secondary systems. These findings align with previous studies [13,14] demonstrating TCM's effects on similar biological activities within nervous, endocrine and immune systems. Of note, both relative frequencies of pathways and total counts of high-relative frequency pathways in each group in the heatmap recapitulate the V-shaped trend observed in Figure 1, adhering to the “hot > warm > neutral < cool < cold” hierarchy.

To evaluate the consistency of these findings and mitigate potential database-specific bias, the *SwissTargetPrediction* (STP) target set was applied through the same procedure. A similar V-shaped hierarchy was observed in both the box plot and the heatmap (see Figure A1-2), although the trend was not as strong as in the *HERB2.0* results. This is likely because the overall abundance of predicted targets and enriched pathways is much lower, reflecting the more conservative nature of predictive computational algorithms (STP) compared with the literature-curated database (*HERB2.0*) that aggregates a broader range of reported associations.

This pattern might imply that hot/cold herbs, which tend to contain more ingredients to influence more targets and pathways hold the option of a synergistic “multi-component, multi-target, multi-pathway” interaction, resulting in different therapeutic efficacy than milder (warm/cool and neutral) herbs. However, given the limitations of current data for quantitative pathway enrichment analysis for TCM herbs, it remains unclear whether hot/warm and cold/cool herbs exert differential effects (e.g., up-regulation and down-regulation) on the same set of pathways.

2.3. Establishment of Classification Models of TCM Herb Properties by Deep Learning

Although the network pharmacology analyses with large sample size suggested certain patterns and potential associations between ingredients, targets, pathways, and herb property labels, the resulting patterns are highly complex and cannot be readily extracted into common features of herbs with different property categories by simple models. Deep learning provides a powerful solution to

address such complexity. In this study, we first applied a conventional approach by constructing MLP-based classification models using ingredient, target, and pathway data of herbs, and subsequently optimized the classification task using the HeteroGAT framework. On the other hand, the classification performance achieved based on these patterns can reflect whether the information of the ingredient, target, and pathway data is genuinely informative and contributes meaningfully to the model.

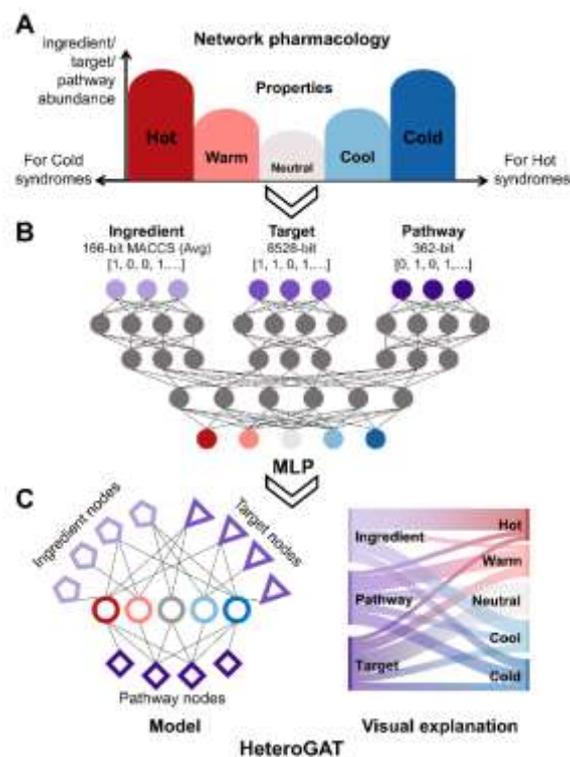


Figure 3. Strategy for property profiling of TCM herbs through (A) network pharmacology, (B) MLP classification models and (C) HeteroGAT classification models with visual explanation.

In Figure 3, the overall strategy for profiling TCM herb properties is illustrated. After obtaining ingredient, target, and pathway data of herbs through network pharmacology (Figure 3A), the multi-level features were applied to construct MLP-based classification models (Figure 3B). In order to avoid imbalance in feature contribution, where high-dimensional features (8528-bit features of target) dominate while low-dimensional ones (166-bit features of ingredient and 362-bit features of pathway) can be underrepresented, each feature type was first processed by a separate sub-MLP to extract compact latent representations, and then concatenated and fed into the final MLP classifier for training.

To address the imbalance among the five property classes (only 36 herbs were labeled as hot), and considering that each class has equal importance in clinical applications, BalAcc (mean of per-class recalls) was selected as the main evaluation metric by averaging the value of recall across all five classes. In addition, the macro-averaged area under the receiver operating characteristic curve (Macro AUC) was used to provide a complementary assessment of overall discriminative ability. Ingredient, target, and pathway features were used independently and in various combinations to train the models. The results of five-fold cross-validation (see S4 in the supplementary material) for MLP models are summarized in Table 2, reporting the mean performance metrics. It can be observed that the BalAcc and Macro AUC values achieved by the MLP models with different feature inputs remain at a low level. This can be attributed to (1) the intrinsic difficulty of the five-class classification task, which increases classification complexity compared to binary tasks; and (2) the sparsity of target and pathway features, which limits their representational power in the model.

Table 2. Balanced accuracy and Macro AUC of MLP and HeteroGAT models with different feature combinations.

Feature combinations	MLP		HeteroGAT	
	BalAcc	Macro AUC	BalAcc	Macro AUC
I	0.3632 ± 0.0310	0.6616 ± 0.0272	0.4638 ± 0.0300	0.6983 ± 0.0203
T	0.3658 ± 0.0285	0.6483 ± 0.0219	0.3575 ± 0.0375	0.6264 ± 0.0219
P	0.3257 ± 0.0300	0.6294 ± 0.0240	0.3286 ± 0.0347	0.5849 ± 0.0241
I-T	0.3646 ± 0.0303	0.6540 ± 0.0245	0.4594 ± 0.0323	0.6954 ± 0.0214
I-P	0.3497 ± 0.0321	0.6522 ± 0.0221	0.4617 ± 0.0343	0.7035 ± 0.0233
I-T-P	0.3652 ± 0.0294	0.6527 ± 0.0254	0.4586 ± 0.0305	0.7016 ± 0.0209

I = Ingredient features; B = Target features; P = Pathway features

To further improve model performance, we subsequently developed a HeteroGAT framework (Figure 3C), in which ingredients, targets, and pathways were represented as individual nodes with their own features and connected to herb nodes in the heterogeneous graph, thereby enabling the model to leverage structural relations and flexible feature inputting. Table 2 also reports the mean BalAcc and Macro AUC values of HeteroGAT models. When ingredient features were employed, both BalAcc and Macro AUC showed a marked improvement compared to the MLP models (paired t-test, two-tailed, $p < 0.01$). This improvement can be attributed to the fact that, unlike the MLP models which averaged MACCS fingerprints across all ingredients of a herb and thereby diluted individual information, the HeteroGAT model retained the fingerprint of each ingredient as an independent node and leveraged the attention mechanism to learn and weigh their relative contributions, thus distinguishing important ingredients from noise and improving classification accuracy. However, when trained on target or pathway features, the HeteroGAT models did not exhibit better performance over the MLP models. Meanwhile, when trained on ingredient combined with target (I-T) and/or pathway features, the performance of HeteroGAT models was not improved, compared to the HeteroGAT model with only ingredient features, indicating that current target and pathway data derived from network pharmacology remain too sparse to independently support or improve herb property classification tasks.

Table 3. Statistics of recall per class of MLP and HeteroGAT models with ingredient feature.

Models	Statistics of recall per class				
	Hot	Warm	Neutral	Cool	Cold
MLP	0.31 ± 0.14	0.52 ± 0.06	0.27 ± 0.09	0.17 ± 0.08	0.54 ± 0.10
HeteroGAT	0.66 ± 0.14	0.49 ± 0.08	0.35 ± 0.09	0.36 ± 0.12	0.45 ± 0.13

Nevertheless, with a more detailed inspection of class-wise recall (Table 3), the MLP model achieved much higher recall of two majority classes (Warm and Cold), while the minority classes (hot, neutral, cool) were poorly classified. This indicates a bias toward majority classes, where the model learned imbalanced patterns that inflated overall accuracy, but failed to capture minority classes. In contrast, the HeteroGAT model with ingredient features improved classification for minority classes (paired t-test, two-tailed, $p < 0.05$), especially the Hot herbs, and improved the BalAcc (Table 2), with an acceptable decrease in recall for the majority classes. This trade-off is valuable in the five-class task where all classes are of equal importance.

Table 4. Comparison of the performance of the HeteroGAT model with previous works.

Resources	Tasks	Features	Models	Performance
TCMID (583 herbs)	Meridian binary classification	Ingredient (4922)	RF, SVM, kNN [15]	BalAcc: 0.67

<i>TCMSP</i> , <i>ETCM</i> (393 herbs)	Property binary classification	Ingredient (12793)	RF, SVM, GNB [11]	AUC: 0.82
<i>TCMID</i> (459 herbs)	Property binary classification	Ingredient (8075)	GCN [12]	Acc: 0.84 F1: 0.85
<i>HERB2.0</i> (1581 herbs)	Property five-class classification	Ingredient (18410) Target (8528) Pathway (362)	HeteroGAT	BalAcc: 0.46 Macro AUC: 0.70

RF = Random Forest; SVM = Support Vector Machine; kNN = k-Nearest Neighbors;
GNB = Gaussian Naive Bayes; GCN = Graph Convolutional Network

Compared with some reported classification models for property or meridian classification tasks of TCM herbs (Table 4), the proposed models yielded relatively lower accuracy and AUC. This can be attributed to differences in data sources, sample sizes, and task types. Most prior studies addressed properties as a binary classification task while Hot and Warm as well as Cool and Cold cannot be fully merged in clinical practice. To the best of our knowledge, our study represents the first attempt to train deep learning models for herb property classification through the more difficult five-class classification which has less distinct class boundaries and lower baseline accuracy than the binary classification. Nonetheless, the HeteroGAT framework has demonstrated its potential, as it is more powerful in capturing structural information and allows the incorporation of diverse node types with rich feature representations, making it particularly suitable for investigation of multi-component, multi-target, multi-pathway network in TCM. However, the current model performance still requires further improvement. At present, ingredient features remain the most informative contributors to model performance. Target and pathway features, while showing preliminary associations in network pharmacology analyses, do not yet provide sufficient and effective information for classification due to limitations in current data. The improvement of target and pathway information could be expected to provide more directional and quantitative analyses of TCM herbs and will play an increasingly valuable role to enhance classification models for TCM herb properties.

2.4. Property Profiling Through HeteroGAT

On the other hand, by leveraging the trained HeteroGAT models, it is possible to derive the contribution (importance weights) of each node to five properties when models make decisions for classification, thereby enabling the profiling of ingredient, target, and pathway nodes in relation to herb property classification. Specifically, the attention coefficients from the trained HeteroGAT models were extracted, and aggregated across models.

The contribution weights of nodes to properties were first analyzed for each model of the five-fold cross-validation, and then averaged across all models to obtain an integrated profile. Detailed outputs were provided in Supplementary Material S5.

To further visualize the contribution of each node to the properties, some top nodes with the highest weights to properties from ingredients, targets, and pathways were selected, respectively, and then merged and illustrated in a Sankey diagram (Figure 4). The widths of the flows in the diagram reveal that each ingredient, target, or pathway node contributes with different weights to distinct herb properties. Notably, the high-weight nodes suggest potential biological significance underlying TCM properties, they currently represent mathematical associations that require experimental validation. While the current results are limited by data quality constraints and the necessity for further model refinement, nevertheless, this strategy provides a prioritized list of candidates, offering a data-driven framework for exploring the complex mechanistic basis of TCM properties.

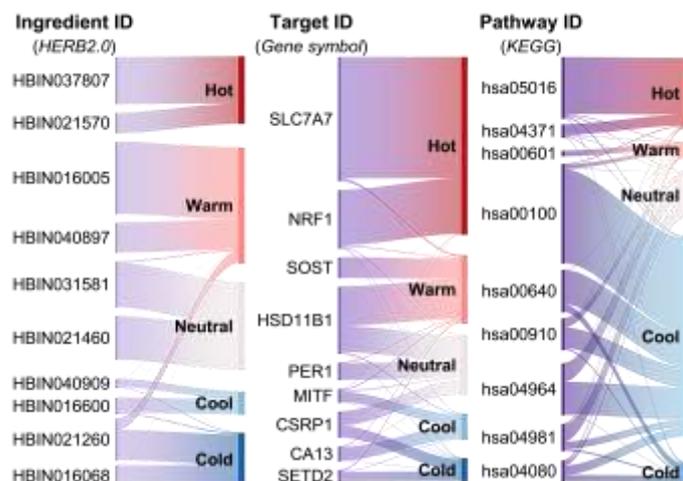


Figure 4. Visual explanation of the importance to properties of selected ingredient, target and pathway nodes.

3. Materials and Methods

3.1. Data Acquisition

Information of property classifications, SMILES (Simplified Molecular Input Line Entry System) structures and related targets of ingredients for 2,415 TCM herbs were extracted from the *HERB2.0* database [16]. All SMILES structures were subsequently submitted to *SwissADME* [17], and ingredients with low Gastrointestinal (GI) absorption were filtered out to finally collect the *HERB2.0* ingredient set. The *HERB2.0* target set was obtained by collecting related targets of the *HERB2.0* ingredient set. Afterwards, for comparison, the same ingredient set was imported into *SwissTargetPrediction (STP)* [18] and an additional *STP* target set was derived through collecting predicted targets with high probability (>0.5) by *STP*.

3.2. KEGG Pathway Enrichment Analysis

For each TCM herb, target gene symbols were extracted from the *HERB2.0* target set and the *STP* target set, separately. *KEGG* pathway enrichment analysis was conducted using the *clusterProfiler* package (v4.14.4) within the R environment (v4.4.1), with *Homo sapiens (hsa)* designated as the reference organism. Pathways meeting a significance threshold of $q\text{-value} < 0.01$ were retained for subsequent analysis.

3.3. Statistical Analysis of Network Pharmacology

Herbs without any significantly enriched *KEGG* pathways were excluded. The remaining herbs were grouped according to their property labels. For each group, three metrics were analyzed: (1) the number of ingredients/targets of each herb, (2) relative frequency of individual targets within the group, and (3) relative frequency of individual *KEGG* pathways within the group. The relative frequency p_j was calculated by $p_j = n_j/N$, in which N represents the number of herbs in a group and n_j represents the number of herbs in a group that is with a specific target j or enriched a specific *KEGG* pathway j .

3.4. Classification Models of TCM Herbs Properties by MLP

For the Multilayer Perceptron (MLP)-based model, herb property labels were defined as hot, warm, neutral, cool, and cold. Based on the data of *HERB2.0* and subsequent results of network pharmacology, ingredient features were encoded using the 166-bit Molecular ACCess System keys (MACCS) fingerprint, and the overall herb representation was obtained by averaging the fingerprint vectors of all its constituent ingredients. Target features of a herb were expressed as Boolean values,

indicating whether a given target was influenced by the herb, while pathway features were encoded in the same way, representing the activation status of each pathway. To address class imbalance in the five-class task, the loss function was the focal loss L [19].

$$L = -\alpha_y * (1 - p_y)^\gamma * \log(p_y)$$

Where p_y is the softmax probability of the true class y , γ is the focusing parameter, and α_y is the (optional) class weight. All models were trained using a five-fold cross-validation strategy with random seeds set to 1995, 218, 1994, 1014, and 2019. The best-performing models were selected based on the balanced accuracy (BalAcc) metric.

3.5. Classification Models of TCM Herbs Properties by HeteroGAT

For the heterogeneous graph attention network (HeteroGAT)-based model [20], a heterogeneous graph was constructed with four node types: herb, ingredient, target, and pathway. Ingredient nodes were represented by their MACCS fingerprints; target nodes were initialized with unique IDs; and pathway nodes were annotated using KEGG pathway classification codes. Edges were created between herb nodes and ingredient/target/pathway nodes if a herb contained an ingredient, influenced a target, or linked to an enriched pathway. The five-fold cross-validation training was likewise formulated as the five-class classification of MLP with focal loss as the loss function.

To evaluate the contribution of nodes to herb classification, the gradient-weighted attention attribution method, adapted from Grad-CAM [21], was employed. Specifically, attention coefficients learned by the HeteroGAT model were combined with their corresponding gradients to derive node-level importance scores, which quantify the relative contribution of non-herb nodes to herb classification decisions, producing interpretable importance values for downstream analysis.

4. Conclusions

This study offers a strategy to explore the scientific basis of the properties of TCM herbs through large-scale analysis of herb-target-pathway relationships by network pharmacology and deep learning. Extending prior investigations that focused on a subset of highest significant pathways, our approach linked the property labels with holistic profiling of pathway patterns across a large sample size of TCM herbs. Results revealed the correlation between the property labels and the abundances of ingredients, targets and pathways. Furthermore, pathways that are influenced by TCM herbs are predominantly concentrated in five systems (endocrine, immune, nervous, signal transduction, cell growth and death). Subsequently, deep learning models were developed and optimized for property classification to further profile the multi-ingredient, multi-target, multi-pathway pattern of TCM herbs, with the HeteroGAT framework enhancing performance through its structural advantages and showing potential for further improvement. The findings derived from the strategy reveal that the potential correlation between analyzed data and TCM properties could offer clues to facilitate the future studies on TCM mechanisms.

To further improve the profiling strategy, it is essential to address current data limitations. For instance, some herbs share similar ingredient/target compositions, but with distinct property labels, suggesting there may be unidentified active ingredients or targets underlying the properties of these herbs. Moreover, quantitative data on ingredient abundances, target activities, and herb-level effects remain incomplete. These limitations preclude definitive conclusions on whether herbs of distinct properties differentially regulate shared targets or pathways, and also restrict the performance of classification models.

Nonetheless, this strategy demonstrated the viability of profiling of herbs integrating ingredient, target and pathway information as a useful exploration under the present conditions, offering not only a valuable perspective to understand and decipher TCM theory, but also a prioritized list of key ingredients, targets, and pathways to guide subsequent experimental validation. In order to enable deeper study on TCM's "multi-component, multi-target, multi-pathway" network, future progress

requires: (1) continuous model refinement such as upgrading feature encoding through incorporating molecular descriptors and 3D structural data to reduce information loss, and constructing dynamic interaction networks to achieve deeper feature fusion across ingredients, targets, and pathways. (2) the integration of more comprehensive and diverse TCM-specific information, such as ingredient profiles, target interaction networks, multi-omics datasets, and pathway annotations. Such systematic improvements will further strengthen the reliability of this strategy and provide a more robust tool for exploring the biological mechanisms underlying TCM herb properties.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org. Table S1: Results of SwissADME; Table S2: Number and relative frequency of targets; Table S3: Relative frequency of pathways; Table S4: Results of five-fold cross-validation; Table S5: Importance for visual explanation.

Author Contributions: Conceptualization, J.J. and A.V.S.; methodology, J.J. and W.H.; software, J.J.; formal analysis, J.J.; resources, W.H., T.Z., E.A. and A.V.S.; data curation, J.J.; writing—original draft preparation, J.J.; writing—review and editing, W.H., T.Z., E.A. and A.V.S.; visualization, J.J.; supervision, A.V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the China Scholarship Council (CSC) (No. 202206787001).

Data Availability Statement: The original contributions presented in this study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Averaged area under the receiver operating characteristic curve
HeteroGAT	Heterogeneous Graph Attention Network
KEGG	Kyoto Encyclopedia of Genes and Genomes
MLP	Multilayer Perceptron
SMILES	Simplified Molecular Input Line Entry System
STP	SwissTargetPrediction
TCM	Traditional Chinese Medicine

Appendix A

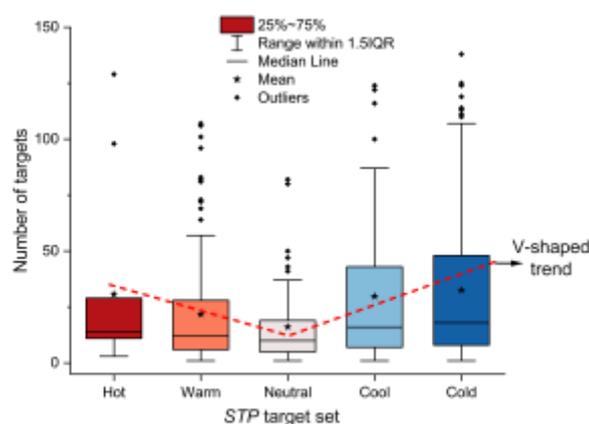


Figure A1. Target abundances of herbs with distinct properties employing the STP target set. Similar to the results from the HERB2.0 target set, the V-shaped trend is observed across the five properties, although the overall abundance of targets is much lower.

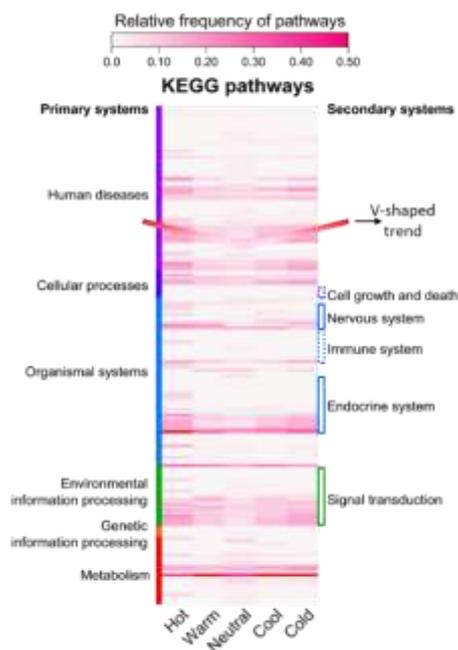


Figure A2. Relative frequency profiling of pathways for property groups employing the STP target set. The pattern is less pronounced, likely due to the lower abundance of STP-derived targets resulting in a reduced number of enriched pathways. Nevertheless, the hierarchy remains visible, particularly within systems such as the nervous system, endocrine system, and signal transduction.

References

1. Chen, P.; Wang, B.-Y.; Zhang, P.; Li, S. Cold and Hot Syndromes in Traditional Chinese Medicine: Insights from the Perspective of Immunometabolic Homeostasis. *World Journal of Traditional Chinese Medicine* **2024**, *10*, 434–442, doi:10.4103/wjtc.wjtc_53_23.
2. Wang, B.-Y.; Chen, P.; Zhang, P.; Li, S. Biological Mechanism of Traditional Chinese Medicine Formula and Herbs in Treating Diseases from the Perspective of Cold and Hot. *World Journal of Traditional Chinese Medicine* **2024**, *10*, 274–283, doi:10.4103/wjtc.wjtc_41_23.
3. Liu, J.; Feng, W.; Peng, C. A Song of Ice and Fire: Cold and Hot Properties of Traditional Chinese Medicines. *Front Pharmacol* **2020**, *11*, 598744, doi:10.3389/fphar.2020.598744.
4. Eigenschink, M.; Dearing, L.; Dablander, T.E.; Maier, J.; Sitte, H.H. A critical examination of the main premises of Traditional Chinese Medicine. *Wien Klin Wochenschr* **2020**, *132*, 260–273, doi:10.1007/s00508-020-01625-w.
5. Li, X.; Liu, Z.; Liao, J.; Chen, Q.; Lu, X.; Fan, X. Network pharmacology approaches for research of Traditional Chinese Medicines. *Chin J Nat Med* **2023**, *21*, 323–332, doi:10.1016/S1875-5364(23)60429-7.
6. Li, S.; Zhang, B. Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chinese Journal of Natural Medicines* **2014**, *11*, 110–120, doi:10.3724/sp.J.1009.2013.00110.
7. Li, R.; Ma, T.; Gu, J.; Liang, X.; Li, S. Imbalanced network biomarkers for traditional Chinese medicine Syndrome in gastritis patients. *Sci Rep* **2013**, *3*, 1543, doi:10.1038/srep01543.
8. Wei, G.; Fu, X.; Wang, Z. Multisolvant Similarity Measure of Chinese Herbal Medicine Ingredients for Cold-Hot Nature Identification. *J Chem Inf Model* **2019**, *59*, 5065–5073, doi:10.1021/acs.jcim.9b00682.
9. Fu, X.; Mervin, L.H.; Li, X.; Yu, H.; Li, J.; Mohamad Zobir, S.Z.; Zoufir, A.; Zhou, Y.; Song, Y.; Wang, Z.; et al. Toward Understanding the Cold, Hot, and Neutral Nature of Chinese Medicines Using in Silico Mode-of-Action Analysis. *J Chem Inf Model* **2017**, *57*, 468–483, doi:10.1021/acs.jcim.6b00725.
10. Zhou, Y.; Xu, B. New insights into molecular mechanisms of "Cold or Hot" nature of food: When East meets West. *Food Res Int* **2021**, *144*, 110361, doi:10.1016/j.foodres.2021.110361.

11. Huang, Z.; Li, Y.; Cheng, H.; Li, G.; Liang, Z. Definition of the molecular bases of cold and hot properties of traditional Chinese medicine through machine learning. *Pharmacological Research - Modern Chinese Medicine* **2022**, *4*, doi:10.1016/j.prmcm.2022.100124.
12. Yang, M.; Liu, W. Classification of cold and hot medicinal properties of Chinese herbal medicines based on graph convolutional network. *Digital Chinese Medicine* **2024**, *7*, 356–364, doi:10.1016/j.dcm.2025.01.008.
13. Li, Y.; Li, R.; Ouyang, Z.; Li, S. Herb Network Analysis for a Famous TCM Doctor's Prescriptions on Treatment of Rheumatoid Arthritis. *Evid Based Complement Alternat Med* **2015**, *2015*, 451319, doi:10.1155/2015/451319.
14. Luo, C.; Yu, H.; Yang, T.; Bai, C.; He, B.; Yan, Y.; Liu, T.; Wang, J.; Gu, X. Data Mining and Systematic Pharmacology to Reveal the Mechanisms of Traditional Chinese Medicine in Recurrent Respiratory Tract Infections' Treatment. *Evid Based Complement Alternat Med* **2020**, *2020*, 8979713, doi:10.1155/2020/8979713.
15. Wang, Y.; Jafari, M.; Tang, Y.; Tang, J. Predicting Meridian in Chinese traditional medicine using machine learning approaches. *PLoS Comput Biol* **2019**, *15*, e1007249, doi:10.1371/journal.pcbi.1007249.
16. Gao, K.; Liu, L.; Lei, S.; Li, Z.; Huo, P.; Wang, Z.; Dong, L.; Deng, W.; Bu, D.; Zeng, X.; et al. HERB 2.0: an updated database integrating clinical and experimental evidence for traditional Chinese medicine. *Nucleic Acids Research* **2024**, *53*, D1404-D1414, doi:10.1093/nar/gkae1037.
17. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7*, 42717, doi:10.1038/srep42717.
18. Daina, A.; Michielin, O.; Zoete, V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Research* **2019**, *47*, W357-W364, doi:10.1093/nar/gkz382.
19. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017; pp. 2980–2988.
20. Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; Yu, P.S. Heterogeneous Graph Attention Network. In Proceedings of the The World Wide Web Conference, San Francisco, CA, USA, 2019; pp. 2022–2032.
21. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017; pp. 618–626.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.