New Compendium of a Myriad of Plants: A New Dataset Describing Ancient Chinese Plants

Xiaobin Shen[1] **Zhongqing Wang**[1*] **Shichen Li**[1] **Chu-Ren Huang**[2] **Guodong Zhou**[1]
[1]Natural Language Processing Lab, Soochow University, Suzhou, China
[2]The Hong Kong Polytechnic University, Hong Kong, China
xbshen@stu.suda.edu.cn wangzq@suda.edu.cn
scli06@stu.suda.edu.cn churen.huang@polyu.edu.hk
gdzhou@suda.edu.cn

Abstract

In ancient China, a variety of datasets depicted humanistic scenes, geographical features, and plants. However, these datasets, compiled long ago, often contain errors, lack comprehensiveness, and are inconsistent with modern realities. To meet current demands, we aim to expand and improve ancient datasets using large language model. Focusing on the *Great Compendium of Myriad Flowers*, an invaluable ancient plants dataset, we gather information on numerous previously excluded plants, carefully select and organize classical Chinese poetry and prose, and construct a comprehensive botanical encyclopedia knowledge system. Additionally, we collect ancient paintings and modern photographs of plants to enrich the dataset. Furthermore, we propose a novel multimodal plant classification model designed to integrate multimodal information from both classical and contemporary sources, enabling the extraction of plant-related information from classical Chinese poetry and prose. Extensive experiments demonstrate the importance of the proposed new ancient plants dataset, and also indicate the effectiveness of our proposed multimodal plant classification model.

1 Introduction

Classical Chinese texts are an essential carrier of Chinese civilization, documenting the rich knowledge accumulated over thousands of years in fields such as history, literature, medicine, agriculture, and astronomy. For example, the Compendium of Materia Medica (Zhou et al., 2024) (*Ben Cao Gang Mu*) recorded herbs, the Classic of Mountains and Seas (Liang et al., 2024) (*Shan Hai Jing*) depicted ancient humanistic scenes and geographical features, and the Great Compendium of Myriad Flowers (*Guang Qun Fang Pu*) documented ancient plants. However, these texts are ancient and mostly written in classical Chinese, which is highly concise and significantly different from modern Chinese, and are also inconsistent with modern realities in many aspects, thus posing considerable challenges for understanding and utilization.

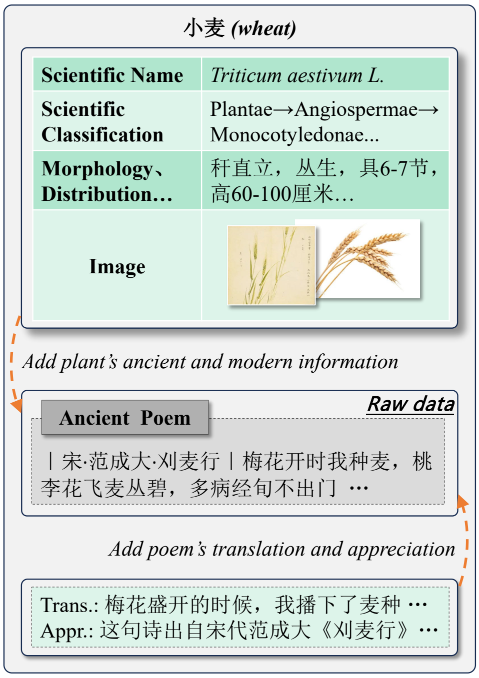


Figure 1: An example of our proposed dataset.

There have been some efforts to digitize these datasets and extract information from them (Yu et al., 2021; Feng et al., 2022; Tang et al., 2024a; Bao et al., 2025). Nevertheless, existing approaches often rely on shallow processing, such as basic digitization or tokenization (Liu et al., 2024a; Xu et al., 2019; Tian and Guo, 2022), and struggle to handle the linguistic characteristics of classical Chinese (Zhou et al., 2023; Congcong et al., 2023; Zhang and Li, 2023a), including condensed syntax, polysemy, and implicit references. They also fail to fully capture the rich historical, cultural, and scientific contexts embedded in these texts (Zheng et al., 2024; Song et al., 2020; Tang et al., 2024b), which can lead to superficial or even

*Zhongqing Wang is the corresponding author.

Findings of the Association for Computational Linguistics: ACL 2026, pages 1483–1498

inaccurate interpretations. Moreover, the scarcity of standardized, high-quality annotated datasets (Li et al., 2022; Cao et al., 2024b,a) further limits the ability of current methods to achieve deep semantic understanding and practical applicability.

Therefore, we plan to expand and improve upon ancient datasets and utilize LLM to construct a new dataset that describe ancient scenes and meet current demands. Specifically, in this study, we take the *Great Compendium of Myriad Flowers* (广群芳谱) dataset, which describes ancient plants, as our starting point.

The Great Compendium of Myriad Flowers is a highly valuable ancient botanical dataset. It covers various aspects of plants, including their growing environments, morphological characteristics, cultivation methods, offering extremely high reference value for ancient plant research. However, this dataset also has some notable shortcomings. The variety of plants included is not extensive enough. The number of classical literary works collected for each plant—which in our research specifically refers to classical Chinese poetry and prose—is limited and not systematically organized. Moreover, it lacks modern classification information and detailed encyclopedic knowledge about plants. Additionally, it does not contain ancient paintings depicting plants or modern plant photographs.

Therefore, based on the original dataset, we carry out extensive and in-depth data collection work. As shown in Figure 1, we gather information on a large number of plants that are not previously included. We also carefully select and organize a sufficient amount of classical literary works for each plant. At the same time, by leveraging LLM together with manual verification, we accurately classify the plants according to modern standards and construct a comprehensive botanical encyclopedia knowledge system. In addition, we translate and appreciate classical literary works, building a bridge between the ancient and the modern. Furthermore, we make every effort to collect ancient paintings depicting each plant as well as modern plant photographs to enrich the content of the dataset and enhance its application value.

After successfully constructing the new dataset, we carry out comprehensive and meticulous data collection work. Ultimately, we collect information on a total of 780 plant species and 44,298 classical literary works related to plants, and supplement these textual data with a curated collection of 1,911 images from both ancient and modern sources to provide comprehensive visual references for further analysis.

This dataset integrates knowledge from multiple disciplines, providing rich resources for interdisciplinary research in digital humanities, while facilitating the digital organization and dissemination of classical literary works, paintings, and other cultural materials.

In addition, we propose a novel multimodal plant classification model that integrates both classical and modern sources to extract plant-related information from classical texts.Extensive experiments demonstrate both the value of the dataset and the effectiveness of the proposed model.

2 Related Work

In this section, we discuss two related topics: digitization of classical literature and information extraction from classical literature.

2.1 Digitization of Classical Literature

The digitization of classical literature involves applying digital technologies to process the content of classical literature, with a key step being the use of OCR (Optical Character Recognition) to identify the text (Chen, 2022).

Early approaches used statistical analysis to conduct research at the content level (Li and Zheng, 2023; Qi et al., 2021; Rajeevan et al., 2020). Traditional machine learning and deep learning later used better representations to perform tasks including sentence segmentation, word segmentation, annotation, classification, and named entity recognition (Wicks and Post, 2021; Liu et al., 2023b; Yue et al., 2021; Jiang et al., 2022; Ihnaini et al., 2024). Recently, the development of LLMs has attracted considerable attention from researchers. They use parameter-efficient fine-tuning to adapt LLMs for digitization tasks (Tian et al., 2021; Zhang and Li, 2023b; Si et al., 2024; Qi et al., 2025).

2.2 Information Extraction from Classical Literature

In the task of entity extraction from classical literature, entities that generally receive attention include personal names, place names, official titles, and temporal expressions (Wang et al., 2023).

Previous studies extracted the entities from pre-Qin literature such as Gong Yang Zhuan (Liu et al., 2024b; Chang et al., 2022; Wu et al., 2021). Later, research further expanded to domain specific ancient literature (Zhang et al., 2020; Song et al.,

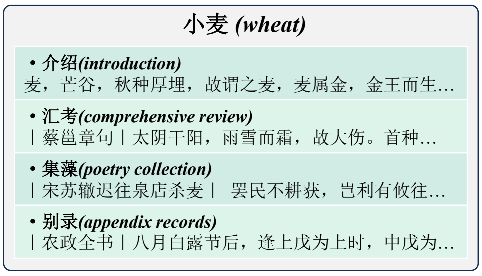


Figure 2: Content and organization of the original Great Compendium of Myriad Flowers dataset.

2020; Wu et al., 2023b), as well as the extraction of information related to war from specific historical periods (Liu and Liu, 2024; Congcong et al., 2023). In addition, studies have also focused on agronomic classical literature, addressing knowledge related to crops and agricultural technologies (Woodrum et al., 2025; Zheng et al., 2024). At the same time, numerous studies have focused on the construction of knowledge bases and knowledge graphs derived from classical literature through information extraction (Bolin, 2021; Zhao and Zhou, 2025; Xiang et al., 2025; Zheng et al., 2025).

Different from previous works, we consider plant terms in classical Chinese, which often undergo de-entityization. Rather than serving as stable entity references, they are commonly used to project abstract meanings. This semantic property fundamentally distinguishes plant information extraction from standard entity centric tasks.

Thus, we expand and annotate the dataset based on the Great Compendium of Myriad Flowers to explore the extraction of plant imagery from classical literary works. By organizing and digitally presenting classical literary works, paintings, and other cultural materials, this dataset helps better inherit and promote the excellent traditional Chinese culture.

3 Dataset Construction

In this section, we give the description of data collection, annotation, and also give some statistics and analysis of our proposed dataset.

3.1 Original Data Collection

The *Great Compendium of Myriad Flowers* (广群芳谱) is a highly valuable ancient botanical dataset. Compiled by Wang Hao and others during the Qing Dynasty. It covers various aspects of plants, including their growing environments, morphological characteristics, cultivation methods, uses, as well as related classical literary works, myths, and legends, offering extremely high reference value for ancient plant research. As shown in Figure 2, for each plant in the dataset, the content record various information from classical literary works and literary references about the plant in ancient times. We manually extract plant information along with the corresponding classical literary works and references, and organize them into a unified schema for each plant from classical Chinese texts. Finally we obtain a high quality, structured basic dataset, laying a solid foundation for subsequent in-depth research.

For this basic dataset, we adopt an organizational form centered around plants. Specifically, for each plant in the dataset, we treat it as an independent entry and attach in detail the corresponding classical literary works extracted precisely from original dataset after the plant entry.

3.2 Extended Data Collection

The original dataset has an obvious drawback: it contains too few plant species, with less than 500 different plants. To meet the needs of more extensive research and applications, we decide to expand the dataset.

First, we construct a list of commonly used plants based on the Catalogue of Life China[1]. The original list contains 47,474 plant species. Through the application of a scientific classification system and a category merging strategy, we consolidate this list to 1,923 commonly occurring plant species. The specific merging strategy is presented in the Appendix. Then, we manually collect possible candidate classical literary works corresponding to these plants from various classical Chinese anthologies[2].

Subsequently, as shown in Figure 3 (a), we design a three-stage framework to annotate the plant mentioned in classical literary work, using a dedicated model, LLM, and manual review. Specifically, we first train a dedicated model on historical data to generate a candidate plant list for each classical literary work. The LLM then, in a one shot manner without updating any model parameters (Zhang et al., 2024), selects the plant that best corresponds to each work, and the results are manually verified to produce the final mapping. During this process, we retain only entries that can

[1] http://www.sp2000.org.cn/

[2] https://www.gushiwen.cn/, http://gs.changrun.org/

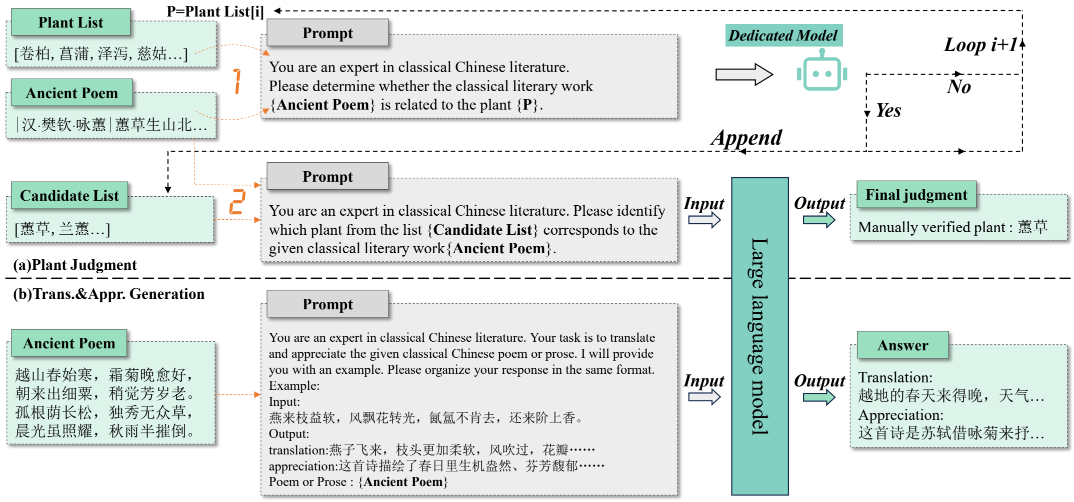


Figure 3: An example of (a) plant judgment, and (b) translation and appreciation generation.

be uniquely mapped to a single plant species.

3.3 Attributes Extraction

After completing the expansion and organization of the dataset, we extract relevant attributes for each plant and its corresponding classical literary works. These attributes include the encyclopedic knowledge of each plant, the translation and appreciation of each classical literary work, ancient paintings depicting each plant, and modern photographs of each plant.

Encyclopedic Information of Plants. Firstly, we retrieve the plant's entry from Wikipedia and other encyclopedic websites[3] and extract the required information, including the botanical name, scientific classification, morphological characteristics, planting situation, growth habits, and uses and organize these information into a fixed format. Then we leverage the classical literary works and related historical information recorded in original dataset to guide LLM to construct information of each plant's ancient morphological characteristics, planting situation, distribution.

Images of Plants. We collect ancient paintings depicting each plant, and modern photographs of each plant. For ancient paintings, we select works from historical atlases [4], these atlases contain a large number of ancient plant illustrations, which we manually collect and align with plant information. For modern photographs, during the process of collecting botanical information from the aforementioned encyclopedic websites, we also gather modern photographs from these sources.

Translation and Appreciation of classical literary works. As shown in Figure 3 (b), we employ a one-shot strategy to instruct the LLM to perform both translation and literary appreciation for each classical literary work in the dataset. The quality analysis of the translations and appreciations generated by the LLM is presented in the Appendix.

Figure 1 shows an example of an annotated data entry, which includes the original text and its corresponding labeled encyclopedic information, translation, appreciation, as well as modern and ancient images.

Dataset	Original	Proposed	Growth
Plant	474	780	64.6%
Image	0	1,911	∞
Poet	821	3,072	274.2%
C_T	61	174	185.2%
C_L	10,604	44,298	317.7%

Table 1: Comparison between the proposed dataset and the original dataset. C_T and C_L mean Classical Chinese Text and Classical Literary Work.

3.4 Statistics and Analysis

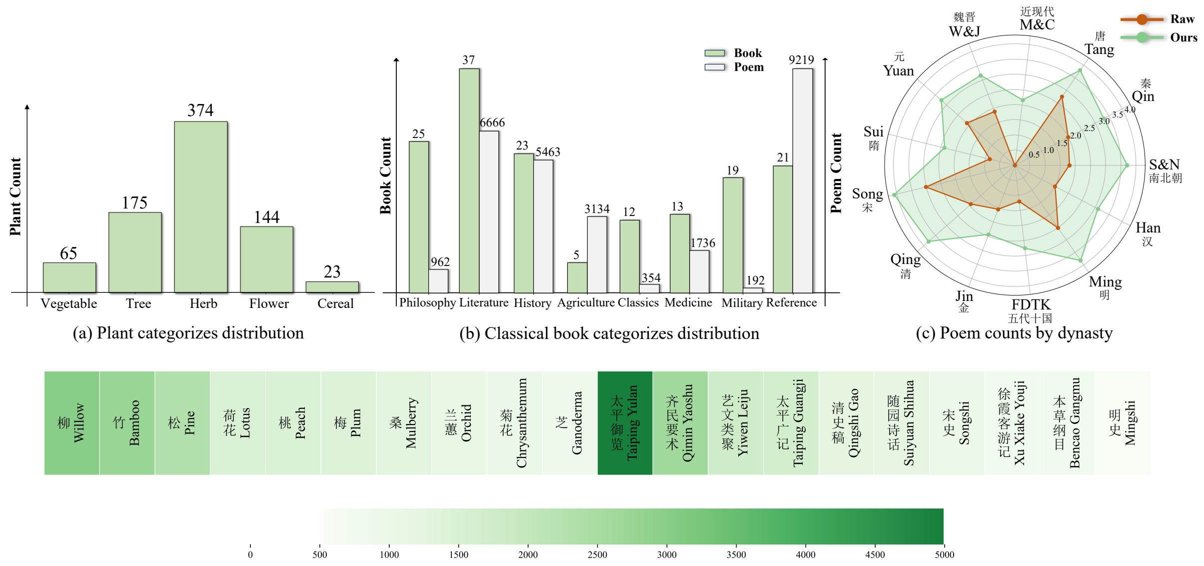
In this subsection, we give some statistics and analysis to evaluate the effective and importance of the proposed new dataset.

Overall Statistics

Table 1 demonstrates a basic comparison the proposed dataset and the original dataset. It is evident that the dataset has been significantly enriched in both diversity and volume. The data distribution

[3] https://zh.wikipedia.org/

[4] https://www.shuge.org/view/jin_shi_kun_chong_cao_mu_zhuang/



(d) Top-10 most frequent plants and classical texts in the dataset

Figure 4: Overview of the dataset: (a) and (b) depict the distribution of plant species among major categories and the number of classical books and literary works across classical book categories. (c) shows the number of classical literary works from different dynasties before and after dataset expansion, where the values in the figure are log transformed. Here, W&J, S&N, FDTK, and M&C refer to the Wei-Jin period, the Southern and Northern Dynasties, the Five Dynasties and Ten Kingdoms, and the Modern and Contemporary period, respectively. (d) shows the top 10 plants and classical texts with the largest number of associated classical literary works in the dataset.

has been further balanced: while the number of plant species increases by less than twofold, the total number of classical literary works grows by more than four times. In addition, to integrate multimodal information from both historical and contemporary perspectives, we associate each plant species with ancient paintings and modern photographs. As a result, a total of 1,911 images are collected for 780 plant species.

Statistics of Plants

Figure 4 (a) demonstrates that the taxonomic diversity of plant species in our dataset, along with a balanced distribution across five predefined categories, ensuring adequate representation for conducting fine-grained studies within each category. In addition, the top 10 plant species depicted in Figure 4 (d) have the highest frequency of occurrence in classical Chinese texts. These plants are commonly found in China, especially in ancient China, and are frequently celebrated in classical literary works, such as willows, chrysanthemums, and orchids.

Statistics by Dynasty

As shown in Figure 4 (c), Chinese history is divided into 13 distinct periods, and we analyze the number of classical literary works from each.

The dataset is predominantly composed of works from the Tang (A.D. 618-907), Song (A.D. 960-1279), Ming (A.D. 1368-1644), and Qing (A.D. 1636-1912) dynasties. These periods provide rich material and can serve as focal points for more in-depth research focused on specific dynasties when needed. Focusing on the trends of specific dynasties allows for more in-depth exploration of their unique cultural characteristics.

Statistics by Classical Chinese Text

As shown in Figure 4 (b) and (d), we conduct a statistical analysis of classical Chinese texts by category and report the ten most frequently occurring texts in the dataset.

From the Figure 4(b), our dataset covers a wide range of classical Chinese texts across diverse categories, ensuring rich textual diversity and capturing the cultural and historical contexts of plants. Figure 4(d) further shows that these texts span multiple historical periods, from the Song to the Qing Dynasty, and include various genres such as historical records, reference works, essay collections, and ancient novels. This diversity enables a comprehensive analysis of the dissemination and evolution of plant classification knowledge across different historical periods and cultural domains.

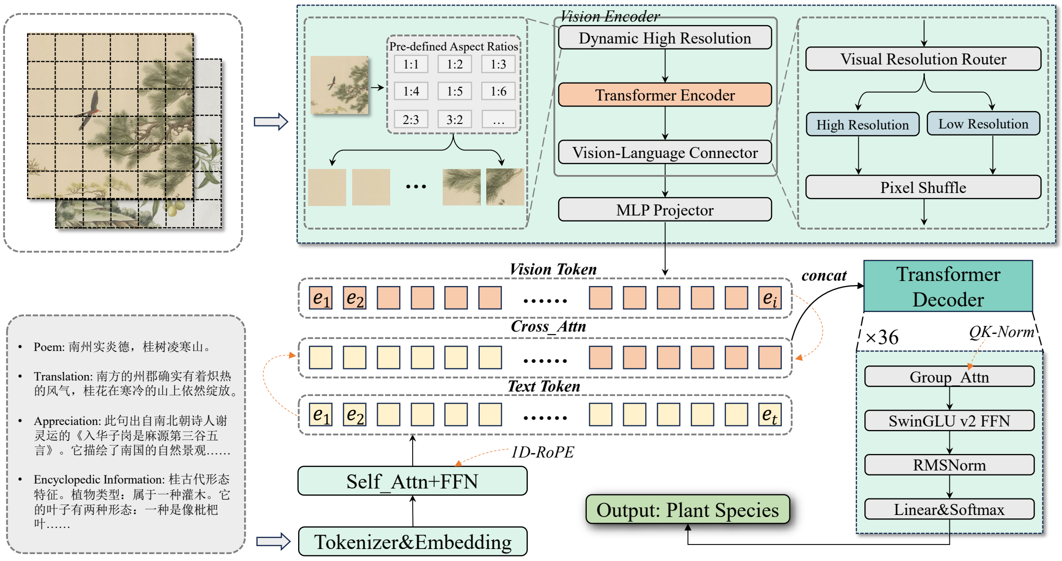


Figure 5: Overview of proposed multimodal plant classification model.

4 Multimodal Plant Classification Model

After collecting and constructing the new dataset, to verify the effectiveness of the dataset we have built, we propose a novel *plant multi-class classification* task based on classical literary works. The input of this task consists of classical literary work and images that depict the content of these classical literary work, while the output is the plant corresponding to the given classical literary work.

As shown in Figure 5, we propose a multimodal fusion model. This model can integrate textual information from classical literary work and image information to simultaneously predict the plant corresponding to the classical literary work.

4.1 Text Representation

Given an input text sequence $W = w_1, \ldots, w_t$ from classical literary work, we first tokenize the sequence and using an embedding matrix, thereby obtaining the hidden vector representations $H = h_1, \ldots, h_t$:

$$H = Embedding(w_1, \ldots, w_t) \quad (1)$$

Since the input consists of a sequential textual structure, we incorporate positional information via 1D-RoPE, which enables the model to capture relative token positions effectively. This sequence is then passed through a multi-layer self attention, combined with feed forward network at each layer, to generate context aware textual representations:

$$E_t = FFN(Self\_Attn(h_1, \ldots, h_t)) \quad (2)$$

4.2 Visual Representation

Given an input image, we first apply dynamic high resolution processing to divide it into multiple fixed size patches $P = p_1, \ldots, p_i$. For patches that may contain critical information, we preserve their high resolution to avoid the loss of fine details. These patches are then fed into the ViT (Dosovitskiy et al., 2021), whose core architecture is built upon the transformer encoder.

After obtaining the visual representations encoded by the ViT, we further adopt the visual resolution router (ViR) (Wang et al., 2025) technique to improve the inference efficiency of the model. The key idea is to evaluate the semantic complexity of each patch: regions containing rich or fine-grained information are preserved with more tokens to avoid semantic loss, while redundant or relatively simple regions are compressed into fewer tokens. This adaptive resolution routing strategy effectively reduces inference latency while maintaining semantic integrity, thereby striking a balance between efficiency and accuracy.

After all patches are processed by the above module, they are flattened and mapped into a vector space. These vectors are then fed into two MLPs to project them into the same dimensional space as the text embeddings.

$$E_v = MLP(VisionEncoder(p_1, \ldots, p_i)) \quad (3)$$

4.3 Multimodal Transformer

In the previous representations for both text and visual inputs, text tokens and visual patches share the same embedding dimension. The model then leverages cross attention to fuse information across modalities, after which the fused representations are concatenated and fed into a stack of transformer decoders for further processing.

$$\begin{aligned} E'_t, E'_v &= Cross\_Attn(E_t, E_v) \\ y &= Decoder(Concat(E'_t, E'_v)) \end{aligned} \tag{4}$$

It is worth noting that in the GQA component of our model, we adopt QK normalization (Henry et al., 2020). Before computing attention, both queries and keys are normalized to prevent large variations in magnitude from affecting the attention scores. This improves the stability of training and helps the model better align information across different modalities.

As a result, the model becomes more adept at capturing the relationships between space and time when processing images, text, or videos, thereby boosting its overall understanding ability.

4.4 Training

The goal of the proposed model is to maximize the output plant name P_n probability given the input classical literary work, its associated textual information W, and visual information P. Here, we use h_m to represent multimodal input. Therefore, we optimize the negative log-likelihood loss function:

$$\begin{aligned} \mathbf{h}_m &= \{W, P\} \\ \mathcal{L}_{\text{mpc}} &= -\frac{1}{N} \sum_{(P_n, h_m) \in \tau}^{N} \log P(P_n^{(i)} \mid \mathbf{h}_m^{(i)}; \theta) \end{aligned} \tag{5}$$

where θ is the model parameters, and (P_n, h_m) is a (plant, multimodal information) pair in training set τ, then:

$$\begin{aligned} &\log P(P_n \mid \mathbf{h}_{\text{m}}; \theta) \\ &= \frac{1}{N} \sum_{i=1}^{N} \log P(P_n^i | P_n^1, \ldots, P_n^{i-1}, h_m; \theta) \end{aligned} \tag{6}$$

where $P(P_n^i | P_n^1, \ldots, P_n^{i-1}, h_m; \theta)$ is calculated by the model.

5 Experiments

In this section, we introduce our experimental setup, showcase the experimental results, and analyze the impact of various factors on the experimental outcomes.

Method	P.	R.	F1.
Rule	0.4428	0.3855	0.4122
GPT-5	0.6880	0.6990	0.6934
MT5-base	0.3988	0.3404	0.3672
Llama2-7B	0.5711	0.5246	0.5468
Qwen2-7B	0.5497	0.5406	0.5451
Wu et al.	0.8040	0.7232	0.7651
Kang et al.	0.7670	0.7680	0.7680
Tang et al.	0.8035	0.7619	0.7824
Blouin et al.	0.5553	0.6127	0.5826
Ours	**0.8020**	**0.7893**	**0.7956**

Table 2: Comparison with baselines.

5.1 Setting

We randomly divide the proposed new dataset into a training set, a validation set, and a testing set, with a split ratio of 8:1:1. We employ InternVL3_5-8B[5] and fine-tune its parameters for our classical literary works plant classification model. We select the best model by early stopping using the eval loss on the validation dataset. The model parameters are optimized by AdamW(Loshchilov and Hutter, 2019) with a learning rate of 5e-5. The batch size is 8. Our experiments are carried out with an Nvidia RTX 4090 GPU. In evaluation, we calculate the Precision and Recall, and use F1 score as the final evaluation metric for classical literary works plant recognition.

5.2 Main Results

As show in Table 2, we conduct experimental comparisons on multiple baseline models, along with a rule-based experiment. The input consists solely of classical literary works. Among these baselines: *Rule* refers to rule-based matching, where semantic similarity analysis[6] is directly conducted between classical Chinese poetry and plant names; *Llama3-8B* (Touvron et al., 2023), *MT5-base* (Xue et al., 2021), *Qwen2-7B* (Yang et al., 2024) have all been pre-trained on large scale Chinese corpora and strike a practical balance between model capacity and computational efficiency, providing sufficient representational power, which enables them to be effectively adapted to our task.

Since our work is the first to address plant classification in Classical Chinese, there are no baselines

[5] https://huggingface.co/OpenGVLab/InternVL3_5-8B-HF

[6] https://huggingface.co/BAAI/bge-large-zh-v1.5

Method	P.	R.	F1.
Classical Literary Work	0.5497	0.5406	0.5451
+Translation	0.7134	0.6897	0.7014
+Appreciation	0.7117	0.7109	0.7113
+Encyclopedic	0.7244	0.7383	0.7313
+The Above Three	0.7742	0.7794	0.7768
+Image	0.7251	0.7160	0.7205
+*ALL*	**0.8020**	**0.7893**	**0.7956**

Table 3: Results of different factors.

specific to the domain for direct comparison. Therefore, we select state of the art information extraction models (Wu et al., 2023a; Kang et al., 2024; Feng et al., 2024; Blouin et al., 2024) developed for other Classical Chinese tasks as our baselines, we adopt them for our plant classification task.

From the experimental results, it is evident that simple rule-based methods struggle to achieve satisfactory performance, indicating their limited capability in handling complex linguistic patterns. In contrast, approaches based on large models, particularly fine-tuned models, can yield superior results.

Furthermore, our proposed multimodal fusion method achieves the highest performance, significantly outperforming all baseline models ($p < 0.05$). Additionally, compared with other works on information extraction from various types of ancient texts, our model consistently outperforms their results.

5.3 Impact of Different Factors

Table 3 shows the comprehensive results of the experiments by analyzing the impact of different factors. From the experimental results, it is clear that translation information, literary appreciation content, and encyclopedic knowledge all play significant roles in the classification task, offering substantial performance improvements compared to using literary work information alone. This demonstrates that providing translation and appreciation, helps the model better understand semantic content and capture subtler associations between text and plant entities, reducing ambiguity inherent in classical expressions. Moreover, by incorporating unified visual information, we effectively alleviate the difficulty of understanding classical Chinese. Ultimately, we find that integrating all types of information into a multimodal large model yields the best classification results. These results confirm that a multimodal approach is essential for maximizing performance in complex plant classification task.

Method	P.	R.	F1.
Llava-v1.6-7b	0.6524	0.6553	0.6538
Qwen2.5-VL-7b	0.8104	0.7694	0.7894
MiniCPM-V-2_6	0.7146	0.6871	0.7006
Gemma-3-4b-it	0.6996	0.7011	0.7003
Ours	0.8020	0.7893	0.7956

Table 4: Performance on different multimodal models.

5.4 Influence of Multimodal Models

To comprehensively evaluate the performance of multimodal models in plant classification tasks, we conducted a series of experiments using various well known multimodal large models, including *LLaVA* (Liu et al., 2023a), *Qwen2.5* (Bai et al., 2025), *MiniCPM* (Yao et al., 2024), and *Gemma3* (Team et al., 2025). The experimental results are presented in Table 4.

When comparing the results of these multimodal models with single modal baselines, we observe that all multimodal models achieve significant improvements. This not only validates the effectiveness of our proposed multimodal model but also indicates that incorporating additional information such as images can significantly assist in plant classification tasks. The multimodal information provides a more comprehensive view of the plants, enabling the model to make more accurate and reliable classifications.

6 Conclusion

In this study, we aim to expand and improve ancient datasets using LLM. Focusing on the Great Compendium of Myriad Flowers, we gather information on numerous previously excluded plants, carefully select and organize ancient poetic and literary materials, and construct a comprehensive botanical encyclopedia knowledge system. Additionally, we collect ancient paintings and modern plant photographs to enrich the dataset. Furthermore, we propose a novel multimodal plant classification model designed to integrate multimodal information from both classical and contemporary sources, enabling the extraction of plant-related information from classical Chinese poetry and prose. Extensive experiments demonstrate the importance and effectiveness of the proposed dataset and method. More detailed analysis can be found in Appendix.

Limitations

Although this study has achieved certain results, there are still limitations and multiple directions for expansion. In terms of data, the current ancient poetic dataset is limited in scale and the content related to plants it covers is not comprehensive enough, making it difficult for the model to generalize to classical Chinese texts beyond poetry and affecting the depth and breadth with which the model can extract plant information from literary works.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62376178),Jiangsu Key Laboratory of Language Computing(JSLCKeyLab 202500003) and Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, and 8 others. 2025. Qwen2.5-vl technical report. *Preprint*, arXiv:2502.13923.

Xiaoyi Bao, Zhongqing Wang, Jinghang Gu, and Chu-Ren Huang. 2025. Revisiting classical Chinese event extraction with ancient literature information. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8440–8451, Vienna, Austria. Association for Computational Linguistics.

Baptiste Blouin, Cécile Armand, and Christian Henriot. 2024. A dataset for named entity recognition and entity linking in chinese historical newspapers. In *LREC-COLING 2024*.

Chang Bolin. 2021. The construction and application for digital humanities knowledge base of ancient books based on word and entity annotation: A case study on <i>zhou qin han annals ofzizhitongjian</i>. *Library and Information Service*, 65(22):134.

Jiahuan Cao, Dezhi Peng, Peirong Zhang, Yongxin Shi, Yang Liu, Kai Ding, and Lianwen Jin. 2024a. Tonggu: Mastering classical chinese understanding with knowledge-grounded large language models. *arXiv preprint arXiv:2407.03937*.

Jiahuan Cao, Yongxin Shi, Dezhi Peng, Yang Liu, and Lianwen Jin. 2024b. CΘ{3} bench: A comprehensive classical chinese understanding benchmark for large language models. *arXiv preprint arXiv:2405.17732*.

Yu Chang, Peng Zhu, Chaoping Wang, and Chaofan Wang. 2022. Automatic word segmentation and part-of-speech tagging of ancient chinese based on bert model. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 141–145.

Li Chen. 2022. Digitalization of ancient books and construction of classical knowledge repository from the perspective of digital humanities. *Journal of library science in China*, 48(02):36–46.

Wei Congcong, Feng Zhenbing, Huang Shutan, Li Wei, and Shao Yanqiu. 2023. Ched: A cross-historical dataset with a logical event schema for classical chinese event detection. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 875–888.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. *Preprint*, arXiv:2010.11929.

Luan Feng, Du Yazhou, Luo Ying, Liu Zhipeng, and Batima Sairekjiang. 2024. Research on traditional chinese medicine named entity recognition based on shang han lun. *Information and Computer (Theoretical Edition)*, 36(20):19–22.

Ping Feng, Jingyu Kang, Biao Huang, Shulai Wang, and Jialun Li. 2022. A method of named entity recognition in classical chinese based on bert-ancient-chinese. In *2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pages 137–141.

Alex Henry, Prudhvi Raj Dachapally, Shubham Pawar, and Yuxuan Chen. 2020. Query-key normalization for transformers. *Preprint*, arXiv:2010.04245.

Baha Ihnaini, Weiyi Sun, Yingchao Cai, Zhijun Xu, and Rashid Sangi. 2024. Sentiment analysis of song dynasty classical poetry using fine-tuned large language models: A study with llms. In *2024 7th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 590–597.

Longjie Jiang, Qinyu C Chang, Huyin H Xie, and Zhuying Z Xia. 2022. Construction of segmentation and part of speech annotation model in ancient chinese. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 155–158.

Zongyuan Jiang, Jiapeng Wang, Jiahuan Cao, Xue Gao, and Lianwen Jin. 2023. Towards better translations from classical to modern chinese: A new dataset and a new method. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 387–399. Springer.

Wenjun Kang, Jiali Zuo, Yiyu Hu, and Mingwen Wang. 2024. Historical event detection in classical chinese based on incremental pre-training and external knowledge. In *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 191–200, Taiyuan, China. Chinese Information Processing Society of China.

Mengfei Li and Xiaohuan Zheng. 2023. Identification of ancient chinese medical prescriptions and case data analysis under artificial intelligence gpt algorithm: a case study of song dynasty medical literature. *Ieee Access*, 11:131453–131464.

Yuqing Li, Yuxin Zhang, Bin Wu, Ji-Rong Wen, Ruihua Song, and Ting Bai. 2022. A multi-modal knowledge graph for classical chinese poetry. In *Findings of the association for computational linguistics: EMNLP 2022*, pages 2318–2326.

Ke Liang, Chu-Ren Huang, and Xin-Lan Jiang. 2024. From text to historical ecological knowledge: The construction and application of the Shan jing knowledge base. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7521–7530, Torino, Italia. ELRA and ICCL.

Chao-Lin Liu, Wei-Ting Chang, Chang-Ting Chu, and Ti-Yong Zheng. 2024a. Machine learning and data analysis for word segmentation of classical chinese poems: illustrations with tang and song examples. *Digital Scholarship in the Humanities*, 39(1):228–241.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023a. Visual instruction tuning. *Preprint*, arXiv:2304.08485.

Liu Liu, Chufei Liu, Wenqi Li, Dongbo Wang, and Shuiqing Huang. 2024b. Perspective of digital humanities on person names in chinese pre-qin classic. *ACM Journal on Computing and Cultural Heritage*, 17(2):1–19.

Yunxiang Liu and Mengtian Liu. 2024. Research on named entity recognition of traditional chinese medicine text based on roberta-bilstm-crf. In *2024 9th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, volume 9, pages 130–135. IEEE.

Zhou Liu, Hongsu Wang, and Peter K Bol. 2023b. Automatic biographical information extraction from local gazetteers with bi-lstm-crf model and bert. *International Journal of Digital Humanities*, 4(1):195–212.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *Preprint*, arXiv:1711.05101.

Hengnian Qi, Hao Yang, Zhaojiang Wang, Jiabin Ye, Qiuyi Xin, Chu Zhang, and Qing Lang. 2025. Ancientglyphnet: an advanced deep learning framework for detecting ancient chinese characters in complex scene. *Artificial Intelligence Review*, 58(3):88.

Lin Qi, Yuwei Wang, Jindong Chen, Mengjie Liao, and Jian Zhang. 2021. Culture under complex perspective: a classification for traditional chinese cultural elements based on nlp and complex networks. *Complexity*, 2021(1):6693753.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. *Preprint*, arXiv:2103.00020.

MS Rajeevan, Chakala Mallikarjuna, Anoop VS, and 1 others. 2020. Nlp-driven knowledge extraction and thematic classification of translated ancient indian medical texts. *Rajeevan, MS, Mini devi, B., Anoop, VS, & Mallikarjuna, C.(2025). NLP-driven Knowledge Extraction and Thematic Classification of Translated Ancient Indian Medical Text. Reimagining LIS Education: Collaborative Integration of Indian Knowledge System with NEP*, 1:351.

Shijing Si, Siqing Zhou, and Yugui Zhang. 2024. Exploring the capabilities of chatgpt in ancient chinese translation and person name recognition. *Corpus-based Studies across Humanities*, 2(2):221–234.

Bingyan Song, Zhenshan Bao, YueZhang Wang, Wenbo Zhang, and Chao Sun. 2020. Incorporating lexicon for named entity recognition of traditional chinese medicine books. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 481–489. Springer.

Xuemei Tang, Zekun Deng, Qi Su, Hao Yang, and Jun Wang. 2024a. Chisiec: an information extraction corpus for ancient chinese history. *arXiv preprint arXiv:2403.15088*.

Xuemei Tang, Zekun Deng, Qi Su, Hao Yang, and Jun Wang. 2024b. CHisIEC: An information extraction corpus for Ancient Chinese history. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3192–3202, Torino, Italia. ELRA and ICCL.

Xuemei Tang, Zekun Deng, Qi Su, Hao Yang, and Jun Wang. 2024c. Chisiec: An information extraction corpus for ancient chinese history. *Preprint*, arXiv:2403.15088.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 197 others. 2025. Gemma 3 technical report. *Preprint*, arXiv:2503.19786.

Huishuang Tian, Kexin Yang, Dayiheng Liu, and Jiancheng Lv. 2021. Anchibert: A pre-trained model for ancient chinese language understanding and generation. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Yanzhi Tian and Yuhang Guo. 2022. Ancient Chinese word segmentation and part-of-speech tagging using data augmentation. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 146–149, Marseille, France. European Language Resources Association.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Nick J Turland, John Harry Wiersema, Fred R Barrie, Werner Greuter, David L Hawksworth, Patrick Stephen Herendeen, Sandra Knapp, Wolf-Henning Kusber, De-Zhu Li, Karol Marhold, and 1 others. 2018. *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Koeltz botanical books.

Dongbo Wang, Chang Liu, Zhixiao Zhao, Si Shen, Liu Liu, Bin Li, Haotian Hu, Mengcheng Wu, Litao Lin, Xue Zhao, and 1 others. 2023. Gujibert and gujigpt: Construction of intelligent information processing foundation language models for ancient texts. *arXiv preprint arXiv:2307.05354*.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, Zhaokai Wang, Zhe Chen, Hongjie Zhang, Ganlin Yang, Haomin Wang, Qi Wei, Jinhui Yin, Wenhao Li, Erfei Cui, and 56 others. 2025. Internvl3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *Preprint*, arXiv:2508.18265.

Rachel Wicks and Matt Post. 2021. A unified approach to sentence segmentation of punctuated text in many languages. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3995–4007.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.

Charles Woodrum, Brian Joseph, Antoine Haaker, and Christopher Brown. 2025. Named entity recognition on ancient languages using large language models: An exploratory study. In *International Conference on Human-Computer Interaction*, pages 430–439. Springer.

Lingjing Wu, Chuang Zhang, Mengqiu Xu, and Ming Wu. 2021. Ancient chinese recognition method based on attention mechanism. In *2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, pages 309–313. IEEE.

MC Wu, LT Lin, Y Qi, SQ Huang, DB Wang, and L Liu. 2023a. Plant knowledge mining and organization construction in pre-qin classics from the perspective of digital humanities. *Library and Information Service*, 67(12):103–113.

Ziwei Wu, Yuxuan Liu, Qingao Huo, and Wendong Zhang. 2023b. Named entity recognition of tcm classics based on sikubert and multi-feature embedding. In *2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 73–77. IEEE.

Shate Xiang, Huanxiang Lin, Fen Cai, and Zhehan Jiang. 2025. Integrating knowledge graphs with ancient chinese medicine classics: challenges and future prospects of multi-agent system convergence. *Chinese Medicine*, 20(1):168.

Han Xu, Wang Hongsu, Zhang Sanqian, Fu Qunchao, and Liu Jun. 2019. Sentence segmentation for classical chinese based on lstm with radical embedding. *The Journal of China Universities of Posts and Telecommunications*, 26(02):1–8.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, and 43 others. 2024. Qwen2 technical report. *Preprint*, arXiv:2407.10671.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, Qianyu Chen, Huarong Zhou, Zhensheng Zou, Haoye Zhang, Shengding Hu, Zhi Zheng, Jie Zhou, Jie Cai, Xu Han, and 4 others. 2024. Minicpm-v: A gpt-4v level mllm on your phone. *Preprint*, arXiv:2408.01800.

Xuehan Yu, Lin He, and Jian Xu. 2021. Extracting events from ancient books based on roberta-crf. *Data Analysis and Knowledge Discovery*, 5(7):26–35.

Du Yue, Wang Dongbo, Jiang Chuan, Xu Runhua, Li Bin, Xu Chao, and Xu Chenfei. 2021. Construction and application of an automatic entity recognition model for classics based on deep learning in digital humanities. *Library and Information Service*, 65(3):100–108.

Dezheng Zhang, Chao Xia, Cong Xu, Qi Jia, Shibing Yang, Xiong Luo, and Yonghong Xie. 2020. Improving distantly-supervised named entity recognition for traditional chinese medicine text via a novel back-labeling approach. *IEEE Access*, 8:145413–145421.

Yixuan Zhang and Haonan Li. 2023a. Can large language model comprehend Ancient Chinese? a preliminary test on ACLUE. In *Proceedings of the Ancient Language Processing Workshop*, pages 80–87, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Yixuan Zhang and Haonan Li. 2023b. Can large language model comprehend ancient chinese? a preliminary test on aclue. *arXiv preprint arXiv:2310.09550*.

Zhen Zhang, Yuhua Zhao, Hang Gao, and Mengting Hu. 2024. Linkner: Linking local named entity recognition models to large language models using uncertainty. *Preprint*, arXiv:2402.10573.

YueYan Zhao and WenJie Zhou. 2025. Unlocking ancient wisdom with modern tools: A new approach to the revitalization of ancient texts based on generative artificial intelligence. *The Journal of Academic Librarianship*, 51(3):103055.

Xiang Zheng, Mingjie Li, and Jingpeng Deng. 2025. Simplifying knowledge acquisition of ancient chinese agricultural books: a method for structure function recognition of bibliographic summaries. *Aslib Journal of Information Management*, pages 1–21.

Xiang Zheng, Mingjie Li, Ze Wan, and Yan Zhang. 2024. Knowledge mining and graph visualization of ancient chinese scientific and technological documents bibliographic summaries based on digital humanities. *Library Hi Tech*, 42(6):1693–1721.

Bo Zhou, Qianglong Chen, Tianyu Wang, Xiaomi Zhong, and Yin Zhang. 2023. Wyweb: A nlp evaluation benchmark for classical chinese. *arXiv preprint arXiv:2305.14150*.

Ying Zhou, Yang Zhou, and Yao Liu. 2024. Research on ontology construction based on the knowledge system of ‘compendium of materia medica’. In *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 642–645.

A Similarity Analysis among Plant, Text and Image

Table 5 presents the similarity metrics within our dataset, including three perspectives: text–image similarity, plant–image similarity, and ancient–modern (i.e., anc. and mod.) image similarity. These metrics are designed to evaluate how well the textual descriptions align with the corresponding visual representations, how closely plant-related information is reflected in images, and how consistent ancient illustrations are when compared with their modern counterparts.

To compute these similarities, we adapt CLIP (Radford et al., 2021), a large scale pre-trained neural network model that jointly encodes images and text into a shared embedding space. The semantic similarity between an image and its associated text or other image is then computed as the cosine similarity between their respective embeddings. We calculate the similarity for all pairs in the dataset and then compute the mean of these values to obtain a global measure of the dataset. By leveraging its powerful cross modal alignment capabilities, CLIP allows us to quantitatively assess the degree of semantic consistency across different modalities in our dataset, thus providing an important reference for evaluating data quality and guiding downstream tasks.

Source	Target	Similarity
Image	Plant	0.3604
Image	Text	0.3716
Anc.Image	Mod.Image	0.4182

Table 5: Results of similarity analysis.

Latin name	Chinese name
Astragalus aksuensis	阿克苏黄芪
Astragalus albicans	革果黄芪
Astragalus alopecias	长尾黄芪
Lagerstroemia stenophylla	狭叶紫薇
Lagerstroemia subcostata	南紫薇
Lagerstroemia venusta	西双紫薇

Table 6: An example of original data.

B Strategy of Plant List Merging

In the data augmentation stage, we find that the raw plant list obtained from the Catalogue of Life China contains too many entries, and we need to reduce the list size to meet our requirements. Table 6 shows a portion of the original data. In the full dataset, many entries exhibit similar patterns, where a single plant species appears under multiple fine-grained taxonomic categories.

Therefore, we adopt a rule-driven merging strategy based on botanical naming conventions. In botanical taxonomy, the Latin scientific name of a plant follows a binomial nomenclature, where the first term denotes the genus, and the subsequent term specify the species (Turland et al., 2018).

Leveraging this convention, we perform plant list merging at the genus level. Specifically, for each plant entry, the Latin name is tokenized using whitespace delimiters, and the first token is extracted as the genus identifier. All entries sharing the same genus are then grouped and merged into a single category. For example, the data shown in the table are merged into Astragalus and Lagerstroemia.

This deterministic rule enables the consolidation of multiple fine-grained species level records into a unified genus level representation.

The proposed rule-based strategy is fully unsupervised, does not rely on external taxonomic databases, and ensures high interpretability and reproducibility. By reducing fine-grained taxonomic redundancy while preserving core taxonomic structure, this approach effectively decreases the size of the plant list and improves its suitability for related data augmentation tasks. Based on the collected raw classical poems from multiple sources, we perform deduplication, and the final dataset retains only plants that are associated with more than 10 classical poems.

C Influence of Dynasties and Categories

Table 7 presents the classification performance across different plant categories. Overall, the model performs relatively well on trees and cereals, indicating strong classification ability for these categories. In contrast, the performance on flowers is comparatively lower, reflecting the fact that flowers are often described metaphorically in classical texts, which increases the difficulty of identification. Moreover, for vegetables and grasses, the results show a certain degree of imbalance: while the model achieves high precision in some cases, it is accompanied by relatively low recall, suggesting missed detections. In general, the model demonstrates strong performance on more straightforward categories, whereas categories with more implicit

Attribute	**Herb**	**Tree**	**Flower**	**Vegetable**	**Cereal**	**ALL**
Pre_Qin	1.0000	-	1.0000	-	1.0000	1.0000
Han	-	-	1.0000	-	0.8000	0.8333
W&J	1.0000	1.0000	0.7500	-	1.0000	0.8947
S&N	-	-	1.0000	-	0.6000	0.7778
Sui	-	1.0000	1.0000	-	1.0000	0.9286
Tang	1.0000	0.6667	0.8043	-	0.9333	0.8267
FDTK	1.0000	1.0000	0.6667	-	-	0.7500
Song	0.7500	0.8000	0.7983	1.0000	1.0000	0.8205
Jin	-	-	1.0000	-	-	1.0000
Yuan	-	1.0000	0.7273	-	1.0000	0.7895
Ming	0.6000	0.8000	0.6905	1.0000	1.0000	0.7193
Qing	-	-	0.9091	-	1.0000	0.9286
M&C	-	-	0.6667	-	-	0.6667
ALL	0.7500	0.8276	0.7868	1.0000	0.9385	0.8207

Table 7: Results of different dynasties and categories.

Translation					
Score	Expert_1	Expert_2	Expert_3	Expert_4	Consistency
Faithfulness	1.200	1.685	1.805	1.815	0.603
Accuracy	1.310	1.595	1.760	1.780	0.638
Fluency	1.445	1.690	1.960	1.970	0.645
Appreciation					
Score	Expert_1	Expert_2	Expert_3	Expert_4	Consistency
Faithfulness	1.655	1.64	1.97	1.965	0.694
Accuracy	1.315	1.775	1.855	1.74	0.589
Fluency	1.835	1.920	2.000	1.985	0.878

Table 8: Human evaluation of LLM generated translations and appreciations

or metaphorical expressions remain challenging.

Finally, we analyze the performance of our proposed model on different plant categories across various historical periods, with the results presented in Table 7. From these results, we can observe that classical literary works from periods such as the Yuan and Ming dynasties often employ more subtle and nuanced metaphors to describe plants, making information extraction more challenging. At the same time, the results reveal the aesthetic preferences of poets in different eras, showing which plant species are more frequently celebrated or symbolically emphasized in their works. This analysis not only highlights the temporal variations in literary representation but also provides insights into the cultural and ecological significance of specific plants throughout Chinese history.

D Evaluation of LLM Generated Translations and Appreciations

To evaluate the quality of the LLM generated translations and literary appreciations, we invite four domain experts—one PhD student and three Master's students—to conduct an assessment. Following prior work on Classical Chinese processing (Jiang et al., 2023), we adopt three evaluation dimensions—faithfulness, accuracy, and fluency—to systematically examine the generated results. We use the AvgAgr metric (Wiebe et al., 2005) to measure inter annotator agreement.

The evaluation scheme is based on classical Chinese poems and their corresponding translations or appreciations, using the following scoring rubric:

- 0: completely not conforming (relevance below 30%)
- 1: basically conforming (relevance between 30% and 60%)

Case	Label	Literary Work	+ALL
渊明岂但隐逸人，渊明素怀诸葛志， 清香不独占秋天，菊潭一滴三千岁。 Tao Yuanming, more than a recluse, bore Zhuge Liang's ambition. Chrysanthemums bloom beyond autumn, their scent endures millennia.	菊花 Chrys.	菊花 ✓ Chrys.	菊花 ✓ Chrys.
春风满面喜津津，纵有痴拳不忍嗔， 窃恐意观安注脚，笑他何事与何人。 Spring breeze brings delight, foolish blows stir no wrath. I fear his thoughts chase notes, laughing at nothing, with no one.	含笑 Michelia	菊花 ✗ Chrys.	含笑 ✓ Michelia
六出玉盘金屈卮，青瑶丛里出花枝， 清香自信高群品，故与江梅相并时。 Six jade plates, golden cup, a flower blooms from green clusters. Fragrance rivals the finest, blooming with the river plum.	水仙 Narcissus	山矾花✗ Ixora	水仙✓ Narcissus

Table 9: Case analysis.

- 2: mostly conforming (relevance above 60%)

Based on the results of human evaluation presented in Table 8, the translations and appreciations generated by the LLM exhibit several notable advantages.

In terms of translation, firstly, regarding faithfulness, although there are some variations in the scores given by different experts, the overall relatively high scores suggest that the LLM generated translations can capture the core essence of the original classical Chinese poems. Secondly, for accuracy, the scores imply that the translations are able to precisely transfer the semantic and cultural information from the original poems. Finally, in terms of fluency, the high scores across the board show that the generated translations read smoothly and naturally.

As for the appreciations, the high faithfulness scores demonstrate that the LLM can generate analyses that are closely related to the original poems. Moreover, the extremely high fluency scores for appreciations imply that the generated content is well organized and logically coherent, with a high level of language proficiency, which enhances the readability and persuasiveness of the appreciations.

E Case Study

Table 9 presents three examples of classical Chinese texts from our dataset, along with the results of plant classification under two settings: using the raw input, and combining all annotation information. Through these examples, we can clearly observe the varying levels of difficulty involved in extracting plant-related information from classical Chinese texts.

Firstly, when the text explicitly states the plant being described, even the simplest baseline methods are capable of correctly identifying the target plant.

However, this is not always the case. For instance, in the second example, classical texts often introduce other plants as contrasts or foils, which can easily mislead the model and result in incorrect classification.

Furthermore, in third example, the description of plants frequently relies on metaphors, allusions, and deep cultural background knowledge. Such implicit expressions pose an even greater challenge for accurate classification, as they require a certain level of cultural understanding.

F Prompt Template

We design the prompt by jointly incorporating role specification, explicit capability guidance, and output space constraints, enabling the model to perform more robust, interpretable, and controllable reasoning in the task of plant classification from Classical Chinese texts.

The prompt template used for plant classification is as follows:

Role

- You are an expert in plant information extraction for Classical Chinese.
- Your task is to extract the plant described in Classical Chinese texts.

Skills

- Auxiliary information includes translations of Classical Chinese, literary appreciation, related encyclopedic knowledge, and images.
- You should combine the above information to carefully analyze the connotations of the Classical Chinese and make accurate judgments.

Constraints

- Your answer must select only from the candidate plant list:{plant_list}
- Only provide the plant name, without any other characters.

Input information: {text + image}