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

Anonymous ACL submission

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 (LLM). 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 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 multi-modal plant classification model designed to integrate multi-modal information from both classical and contemporary sources, enabling the extraction of plant-related information from classical Chinese texts. Extensive experiments demonstrate the importance of the proposed dataset, and also indicates the effectiveness of our proposed multi-modal plant classification model.

1 Introduction

Chinese classical 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.

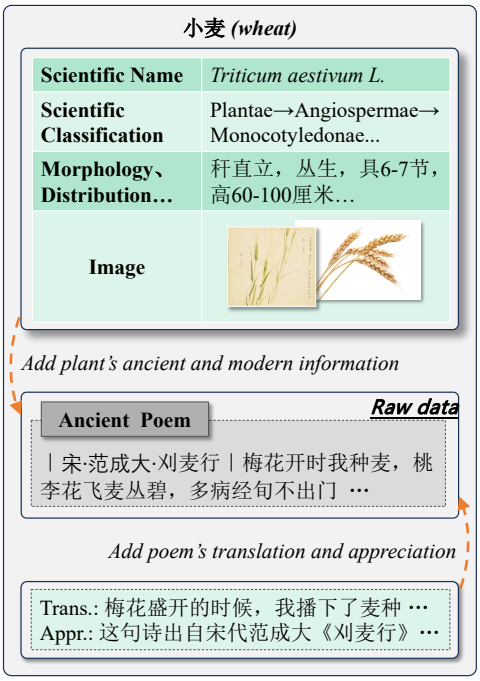

Figure 1: An example of our proposed dataset.

There have been some efforts to digitize these datasets and extract information from them (Yu et al., 2021; Xie et al., 2022; 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., 2022; Yu et al., 2020; Tian and Guo, 2022), and struggle to handle the linguistic characteristics of classical Chinese (Zhou et al., 2023; Congcong et al., 2023), including condensed syntax, polysemy, and implicit references. They also fail to fully capture the rich historical, cultural, and scientific contexts embedded in these texts (Mengcheng et al., 2023; Feng et al., 2024; Tang et al., 2024b), which can lead to superficial or even 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 new datasets that describe ancient scenes and meet current demands. Specifically, in this study, we take the *Great Compendium of Myriad Flowers* 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 ancient poems and literary works collected for each plant is limited and not systematic. 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 were not previously included. We also carefully select and organize a sufficient amount of ancient poetic and literary materials for each plant. At the same time, by leveraging LLM, we accurately classify the plants according to modern standards and construct a comprehensive botanical encyclopedia knowledge system. In addition, we translate and appreciate classical poems, building a bridge between the ancient and the modern. Furthermore, we made 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 Chinese poetry and prose works related to plants, and supplement this 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 materials and strong support for interdisciplinary research in the field of digital humanities. In addition, by organizing and digitally presenting ancient poetic works, paintings, and other cultural materials, this dataset helps better inherit and promote the excellent traditional Chinese culture.

Furthermore, we propose a novel multi-modal plant classification model designed to integrate multi-modal information from both classical and contemporary sources, enabling the extraction of plant-related information from classical Chinese texts. Extensive experiments demonstrate the importance of the proposed dataset, and also indicates the effectiveness of our proposed multi-modal plant classification model.

2 Related Works

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 ancient texts, 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 (Lingxiang and Xiting, 2006; Wei et al., 2014; Liu 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 (Yue et al., 2019; Na, 2018; Yue et al., 2021; Qi et al., 2021). 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, 2023; Si et al., 2024; Qi et al., 2025).

2.2 Information Extraction from Classical Literature

In the task of entity extraction from ancient texts, entities that generally receive attention include personal names, place names, official titles, and temporal expressions (Yang and Dongbo, 2024).

Previous studies extracted the entities from pre-Qin texts such as Gong Yang Zhuan (Yafen, 2013; Dongbo et al., 2018; Mengcheng et al., 2023). Later, research further expanded to domain-specific

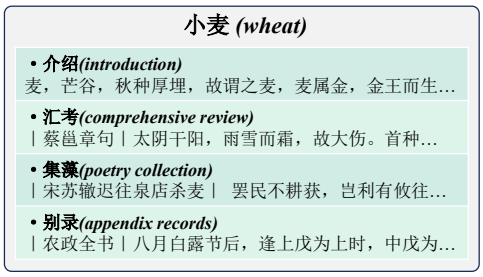

Figure 2: Content and organization of the original Great Compendium of Myriad Flowers dataset.

ancient texts (Chao and Wenbo, 2021; Feng et al., 2024; Jing et al., 2022), as well as the extraction of war-related information from specific historical periods (Xuehan et al., 2021; Zhangchao et al., 2020). In addition, studies have also focused on agronomic classical texts, addressing knowledge related to crops and agricultural technologies (Na, 2021; Zheng et al., 2024). At the same time, numerous studies have focused on the construction of knowledge bases and knowledge graphs derived from classical texts through information extraction (Bolin et al., 2021; Liu et al., 2023b; Wang et al., 2023; Wei et al., 2024).

Different from previous works, we expand and annotate the dataset based on the Great Compendium of Myriad Flowers to explore the extraction of plant imagery from classical Chinese poetry and prose. By organizing and digitally presenting ancient poetic 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 (*Guang Qun Fang Pu*) 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 poems, 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 ancient poetic and literary references about the plant in ancient times.

We extract plant information and the corresponding ancient poetic and literary references for each plant from the original dataset. Additionally, we remove data with a very low frequency of occurrence and also eliminated duplicate entries. Through this series of rigorous processing steps, we finally 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 ancient text descriptions 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 five hundred different plants. To meet the needs of more extensive research and applications, we decided to expand the dataset.

First, we constructed a list of commonly-used plants based on the Catalogue of Life China[1]. This list includes 47,474 plant species. By applying a scientific classification and category-merging strategy, we reduced the total number to 1,923. Then, we collected possible candidate ancient poems and literary works corresponding to these plants from various sources, such as ancient poetry collections and historical documents[2].

Subsequently, as shown in Figure 3 (a), we train a dedicated model solely using past data. The function of this model is to determine whether a given candidate ancient poem or literary work is truly related to a given plant. Then, we combine each plant from the plant list with an ancient poem and input them into the model for evaluation, resulting in a candidate list of possible plants for each poem. Finally, by carefully designing a reference example for the LLM, we use the LLM to determine the plant that best corresponds to each poem. Since this process does not require updating any model parameters, it can be regarded as a one-shot approach. (Zhang et al., 2024).

3.3 Attributes Extraction

After completing the expansion and organization of the data, we extracted relevant attributes for each plant and its corresponding ancient poems and essays. These attributes include the encyclopedic

[1] http://www.sp2000.org.cn/
[2] https://www.gushiwen.cn/

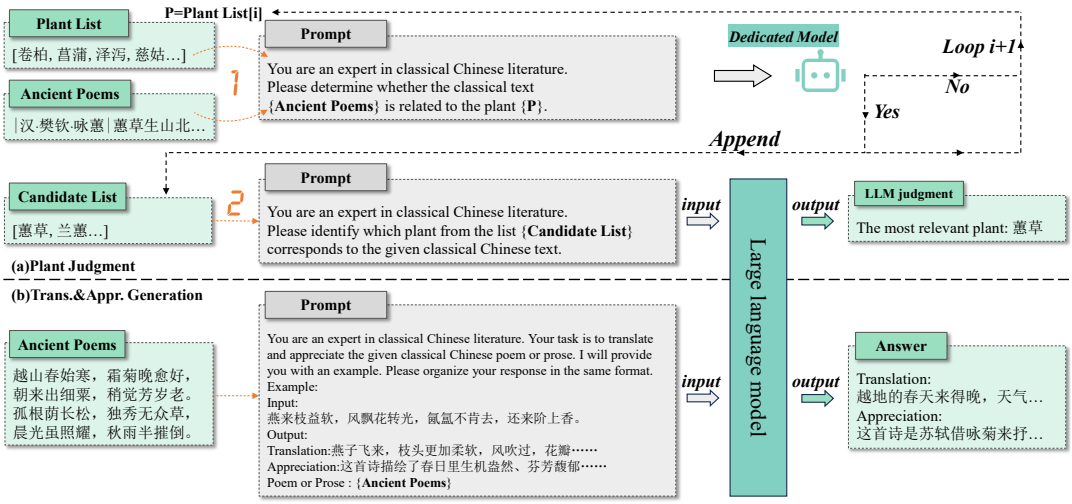

Figure 3: An example of (a) plant judgment, and (b) translation and appreciation generation.

knowledge of each plant, the translations and appreciations of each ancient poem or essay, 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. Then we leverage the classical poems 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 Poems. We employ a one-shot prompt strategy to instruct the LLM to perform both translation and literary appreciation for each entry in the dataset. As shown in Figure 3 (b), we guide the model to produce accurate translations that preserve the semantic and stylistic nuances of the original text.

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

Table 1: Comparison between the proposed dataset and the original dataset.

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.

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 has been further balanced: while the number of plant species increases by less than twofold, the total number of classical Chinese poetry grows by more than four times.

In addition, to integrate multi-modal information from both historical and contemporary perspectives, we associate each plant species with ancient

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

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

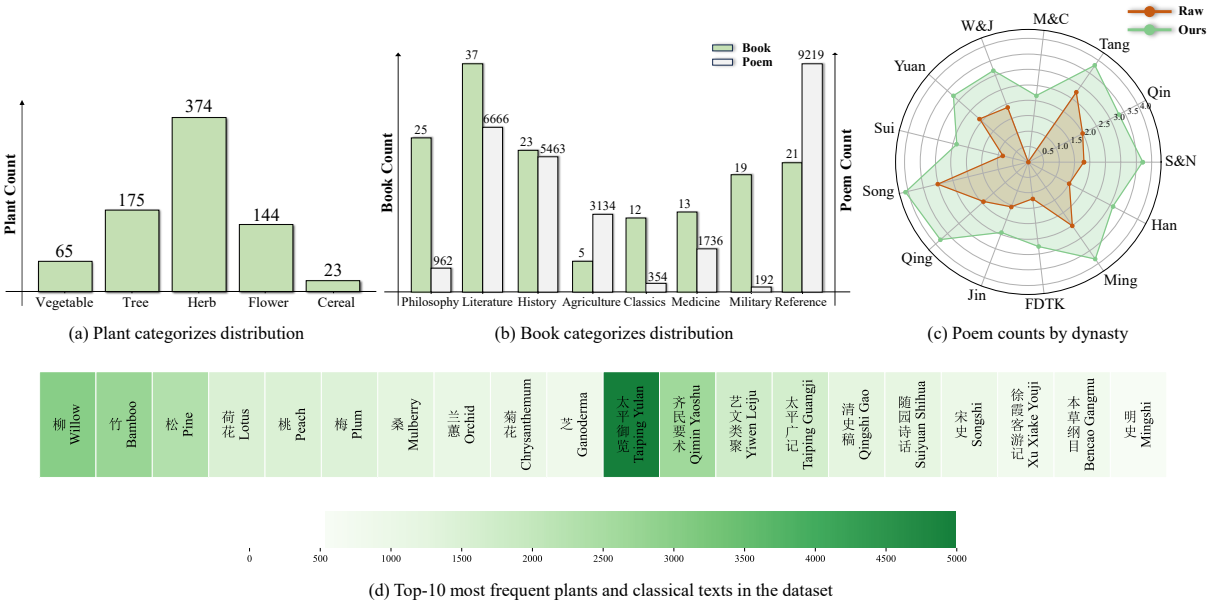

Figure 4: Overview of the dataset: (a) and (b) depict the distribution of plant species among major categories and the number of books and poems across book categories. (c) shows the number of classical poetry and prose works from different dynasties before and after dataset expansion, where the values 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 poems in the dataset.

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) demonstrate that the taxonomic diversity of plant species in our dataset, along with a well-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 were frequently celebrated in ancient Chinese poetry and literature, 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 poetry and prose works from each.

The dataset is predominantly composed of works from the Tang, Song, Ming, and Qing dynasties. These periods provide rich material and can serve as focal points for more in-depth, dynasty-specific research when needed. Focusing on the trends of specific dynasties allows for more in-depth exploration of their unique cultural characteristics.

Statistics by Classical Book

As shown in Figure 4 (b) and (d), we conduct a statistical analysis of the number of books across different categories and the ten most frequently appearing ancient books in the dataset.

It can be clearly seen from the figure that these ancient books span a wide range of historical periods, covering various eras from the Song Dynasty to the Qing Dynasty. At the same time, they also exhibit great diversity in terms of types, belonging to different categories including historical books, reference books, collections of essays, and collections of ancient novels, among others. Therefore, by analyzing the historical periods and types to which these ancient books belong, we can gain a more comprehensive understanding of the dissemination, development, and evolution of ancient plant classification knowledge across different historical periods and cultural fields.

4 Multi-modal 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 classification* task based on ancient poems and essays. The input of

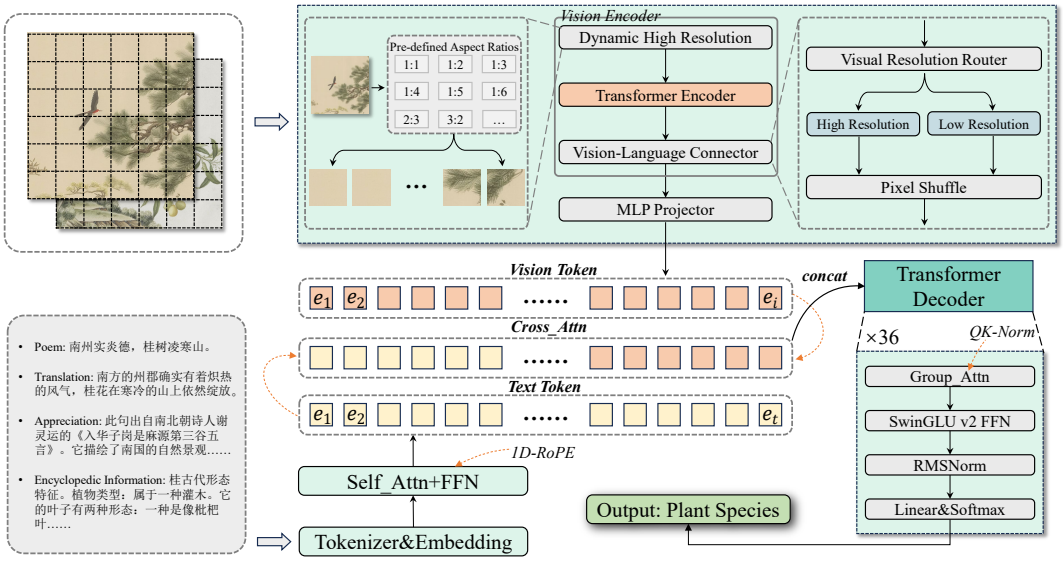

Figure 5: Overview of proposed multi-modal plant classification model.

this task consists of ancient poems or essays and images that depict the content of these poems or essays, while the output is the plant corresponding to the given ancient poems or essays.

As shown in Figure 5, we propose a multi-modal fusion model. This model can integrate textual information from ancient poems and image information to simultaneously predict the plant corresponding to the ancient poem or prose.

4.1 Text Representation

Given an input text sequence $W = w_1, \ldots, w_t$ from ancient poems and essays, 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 networks 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 Multi-modal 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.

$$E'_t, E'_v = Cross_Attn(E_t, E_v)$$
$$y = Decoder(Concat(E'_t, E'_v) \quad (4)$$

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 poem, its associated textual information W, and visual image data P. Here, we use h_m to represent multi-modal input. Therefore, we optimize the negative log-likelihood loss function:

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

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

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

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.

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.

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
Mengcheng 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.

We employ InternVL-3.5[5] and fine-tune its parameters for our classical Chinese poetry plant recognition Model. We select the best models by early stopping using the eval loss on the validation dataset. The model parameters are optimized by Adam(Kingma and Ba, 2017) 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 Chinese poetry 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 Chinese poetry. 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) are the popular LLMs, we fine-tune them based on our proposed dataset. We also employ varaious state-of-the-art classical text information extraction models (Mengcheng et al., 2023; Kang et al., 2024; Feng et al., 2024; Blouin et al., 2024) as baselines, we adopt them for our proposed 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

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

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

Method	P.	R.	F1.
Poem	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.

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 multi-modal models.

contrast, approaches based on large models, particularly fine-tuned models, can yield superior results.

Furthermore, our proposed multi-modal 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 text information alone. This demonstrates that providing translations 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, incorporating image information can also effectively enhance the performance of plant classification. Ultimately, we found that integrating all types of information into a multi-modal large model yields the best classification results. This demonstrate that multi-modal learning effectively leverages complementary information from textual, knowledge-based, and visual sources. These results confirm that a multi-modal approach is essential for maximizing performance in complex plant classification tasks.

5.4 Influence of Multimodal Models

In addition, as shown in Table 4, we conduct experiments with various multi-modal 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), which respectively represent vision-language alignment models, advanced instruction-following multi-modal models, lightweight efficient architectures, and open-source models with strong multi-modal reasoning capabilities.

Compared with the single-modal baselines, all models achieved significant improvements, including those without extensive pre-training on large-scale Chinese corpora, which further validates the value of our dataset in supporting multi-modal plant information extraction tasks. Thess not only validates the effectiveness of our proposed multi-modal model but also indicates that incorporating additional information such as images can significantly assist in plant classification tasks.

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 multi-modal plant classification model designed to integrate multi-modal information from both classical and contemporary sources, enabling the extraction of plant-related information from classical Chinese texts. Extensive experiments demonstrate the importance and effectiveness of the proposed dataset and method. More detail analysis of consistency among text and image, influence of dynasties and categories, and examples of case study, 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 poetry and prose dataset is limited in scale, and the plant-related content it covers is not comprehensive enough, affecting the depth and breadth of the model's mining of plant information in ancient poetry and prose.

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, Wan Chen, Li Bin, Chen Xinyu, Feng Minxuan, and Wang Dongbo. 2021. Construction and application of a digital humanities knowledge base of ancient texts based on word and entity annotation: A case study of zizhi tongjian·zhou, qin, and han records. *Library and Information Service*, 65(22):134–142.

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. CŒ{3} bench: A comprehensive classical chinese understanding benchmark for large language models. *arXiv preprint arXiv:2405.17732*.

Sun Chao and Zhang Wenbo. 2021. Research progress and challenges in named entity recognition of terminology in traditional chinese medicine ancient texts. *China Journal of Chinese Materia Medica*, 36(11):6843–6845.

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.

Wang Dongbo, Gao Ruiqing, Shen Si, and Li Bin. 2018. Research on automatic recognition of basic entity components of historical events in pre-qin classics. *Journal of the National Library of China*, 27(1):65–77.

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.

Xie Jing, Liu Jiangfeng, and Wang Dongbo. 2022. Research on named entity recognition in ancient chinese medical literature: A case study of the flat-lattice enhanced sikubert pre-trained model. *Library Tribune*, 42(10):51–60.

Wenjun Kang, Jiali Zuo, Yiyu Hu, and Mingwen Wang. 2024. . 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.

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization. *Preprint*, arXiv:1412.6980.

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.

Xia Lingxiang and Huang Xiting. 2006. A conceptual analysis of the meaning of self-reliance in ancient books. *Acta Psychologica Sinica*, (6):916–923.

Chang Liu, Dongbo Wang, Haotian Hu, Yiqin Zhang, and Bin Li. 2022. Research on automatic word segmentation of ancient chinese classics with external feature fusion for digital humanities: A case study of the sikubert pre-trained model. *Library Tribune*, 42(06):44–54.

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

Liu Liu, Huang Shuiqing, Meng Kai, Li Bin, Wang Dongbo, and Su Xinning. 2020. A humanistic computing study on female figures in the three commentaries of the spring and autumn annals. *Library and Information Service*, 64(23):109–123.

Liu Liu, Qi Yue, Liu Chufei, Li Wenqi, and Wang Dongbo. 2023b. Research on ancient text citation and construction of full-text knowledge base under computational humanities. *Journal of Intelligence*, 42(12):1498–1512.

Wu Mengcheng, Lin Litao, Qi Yue, Huang Shuiqing, Wang Dongbo, and Liu Liu. 2023. Research on mining and organizing plant knowledge from pre-qin classics from the perspective of digital humanities. *Library and Information Service*, 67(12):103–113.

Li Na. 2018. Construction of an automatic alias extraction model for local gazetteer classics based on conditional random fields. *Journal of Chinese Information Processing*, 32(11):41–48, 61.

Li Na. 2021. Construction of a joint automatic recognition model for multiple types of named entities in gazetteer-class ancient texts. *Library Forum*, 41(12):113–123.

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.

Zhang Qi, Jiang Chuan, Ji Youshu, Feng Minxuan, Li Bin, Xu Chao, and Liu Liu. 2021. Construction of an integrated word segmentation and part-of-speech tagging model for multi-domain pre-qin classics. *Data Analysis and Knowledge Discovery*, 5(3):2–11.

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.

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.

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.

Linxu Wang, Jun Wang, and Wei Tong. 2023. Using ontology to organize chinese ancient books in the digital age. *Proceedings of the Association for Information Science and Technology*, 60(1):712–716.

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.

Meng Wei, Wang Xifa, Ma Sulin, and Ma Lihua. 2014. Analysis of medication for heart failure based on ancient medical case literature data. *China Journal of Chinese Materia Medica*, 29(3):898–900.

Zhang Wei, Wang Hao, Wang Dongbo, Qiu Jingwen, Li Xiaomin, and Deng Sanhong. 2024. Promoting literary cognition through data association: A semantic organization method for metaphorical cultural schemas in classical poetry. *Library and Information Service*, 68(04):109–123.

Zhiqiang Xie, Jinzhu Liu, and Genhui Liu. 2022. (construction and application of classical Chinese nested named entity recognition data set). In *Proceedings of the 21st Chinese National Conference on Computational Linguistics*, pages 406–416, Nanchang, China. Chinese Information Processing Society of China.

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.

Yu Xuehan, He Lin, and Xu Jian. 2021. Research on ancient chinese historical event extraction method based on roberta-crf. *Data Analysis and Knowledge Discovery*, 5(07):26–35.

Tang Yafen. 2013. Research on automatic recognition of personal names in pre-qin classical chinese texts. *New Technology of Library and Information Service*, (Z1):63–68.

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.

Liu Yang and Wang Dongbo. 2024. Current research on intelligent information processing of classical books. *Library and Information Service*, 68(23):120–138.

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.

Jingsong Yu, Yi Wei, Yongwei Zhang, and Hao Yang. 2020. Research on classical chinese word segmentation based on non-parametric bayesian model and deep learning. *Journal of Chinese Information Processing*, 34(06):1–8.

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.

Yuan Yue, Wang Dongbo, Huang Shuiqing, and Li Bin. 2019. Exploring the differences of different part-of-speech tag sets in entity extraction from classics. *Data Analysis and Knowledge Discovery*, 3(3):57–65.

Yixuan Zhang and Haonan Li. 2023. 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.

Li Zhangchao, Li Zhongkai, and He Lin. 2020. Research on war event extraction technology in zuo zhuan. *Library and Information Service*, 64(07):20–29.

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.

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

Table 5: Results of similarity analysis.

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,

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 6: Results of different dynasties and categories.

Case	Label	Poem	+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 7: Case analysis.

and how consistent ancient illustrations are when compared with their modern counterparts.

To compute these similarities, we adopt 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.

B Influence of Dynasties and Categories

Table 6 presents the classification performance across different plant categories. Overall, the model performs relatively well on trees and cereals, indicating strong recognition 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 or metaphorical expressions remain challenging.

Finally, we analyzed the performance of our proposed model on different plant categories across various historical periods, with the results presented in Table 6. From these results, we can observe that classical poems 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 were 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.

C Case Study

Table 7 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 recognition, as they require a certain level of cultural understanding.