

Journal of Plant Research

Herbarium specimens in the age of artificial intelligence: from herbarium image identification to Integrative Taxonomic AI --Manuscript Draft--

Manuscript Number:	JPRE-D-26-00177R3									
Full Title:	Herbarium specimens in the age of artificial intelligence: from herbarium image identification to Integrative Taxonomic AI									
Article Type:	Current Topics in Plant Research									
Funding Information:	<table><tr><td>Japan Society for the Promotion of Science (No. 120258806)</td><td>Dr. Atsuko Takano</td></tr><tr><td>National Research Foundation of Korea (RS-2025-00508625)</td><td>Prof. Sangtae Kim</td></tr><tr><td>Japan Society for the Promotion of Science (JSPS24K09576)</td><td>Dr. Atsuko Takano</td></tr><tr><td>National Research Foundation of Korea (RS-2021-NR063105)</td><td>Prof. Sangtae Kim</td></tr></table>		Japan Society for the Promotion of Science (No. 120258806)	Dr. Atsuko Takano	National Research Foundation of Korea (RS-2025-00508625)	Prof. Sangtae Kim	Japan Society for the Promotion of Science (JSPS24K09576)	Dr. Atsuko Takano	National Research Foundation of Korea (RS-2021-NR063105)	Prof. Sangtae Kim
Japan Society for the Promotion of Science (No. 120258806)	Dr. Atsuko Takano									
National Research Foundation of Korea (RS-2025-00508625)	Prof. Sangtae Kim									
Japan Society for the Promotion of Science (JSPS24K09576)	Dr. Atsuko Takano									
National Research Foundation of Korea (RS-2021-NR063105)	Prof. Sangtae Kim									
Abstract:	Herbarium specimens are physical, verifiable records that form the basis of taxonomic knowledge and biodiversity research. Their large-scale digitization has produced extensive collections of high-resolution images and associated specimen metadata, creating conditions in which artificial intelligence (AI) can play an important role in plant taxonomy, collection management, and ecological research. Early AI applications have primarily focused on automated species identification based on individual specimen images. Although increasingly accurate, such approaches remain limited by their emphasis on single-specimen label prediction and by treating identification outputs as final analytical decisions. Recent methodological advances—including segmentation-based preprocessing, automated trait extraction, structured extraction of label data, detection of potentially misidentified specimens, and multimodal integration of visual, textual, and genetic information—extend AI applications beyond species identification toward broader analytical frameworks, encompassing taxonomic interpretation as well as ecological and biodiversity research. In these approaches, specimens are placed within a shared analytical space, and identification results are used to support comparisons across multiple specimens rather than being treated as final decisions for single individuals. This multi-specimen perspective enables quantitative examination of species boundaries, morphological variation, data inconsistencies, and taxonomic stability within curated collections. In this context, AI serves not as an ultimate decision-maker but as a decision-support tool embedded in expert-guided workflows and biodiversity knowledge infrastructures. These developments can be summarized as Integrative Taxonomic AI, an approach that employs learned morphospaces to interpret and refine taxonomic categories by integrating multimodal evidence and curated specimen data under expert guidance.									
Corresponding Author:	Sangtae Kim Sungshin Women's University KOREA, REPUBLIC OF									
Corresponding Author Secondary Information:										
Corresponding Author's Institution:	Sungshin Women's University									
Corresponding Author's Secondary Institution:										
First Author:	Atsuko Takano									
First Author Secondary Information:										
Order of Authors:	Atsuko Takano Hyeonji Moon									

	Masato Shirai
	Shuichiro Tagane
	Diego Tavares Vasques
	Chung-Kun Lee
	Amjad Khan
	Jeongwoo Seo
	Sujeong Han
	Chanho Park
	Jaesung Lee
	Takashi Akihiro
	Sangtae Kim
Order of Authors Secondary Information:	
Author Comments:	Dear Dr. Noriko Inada, an Editor-in-Chief of the Journal of Plant Research, Please find attached our manuscript entitled “Herbarium specimens in the age of artificial intelligence: from herbarium image identification to Taxonomic AI” for consideration for publication in the Journal of Plant Research. We have edited the manuscript based on the editor's comments. This review was developed from discussions and presentations at the Japan–Korea Collaboration Symposium and Workshop held at the University of Tokyo last year, entitled “Innovating Plant Taxonomy with AI: Shaping the Future through Japan–Korea Collaboration.” The event was jointly organized by Sangtae Kim and Atsuko Takano and supported by a bilateral collaboration program funded by the Japan Society for the Promotion of Science and the National Research Foundation of Korea. The symposium and workshop brought together researchers working at the interface of plant systematics and artificial intelligence, and this manuscript synthesizes and further develops the key scientific themes that emerged from that collaborative exchange. In this review, we examine recent advances in AI applications to digitized herbarium specimens, including automated species identification, segmentation-based preprocessing, automated trait and label extraction, multimodal integration, and their broader implications for taxonomy and biodiversity knowledge infrastructures. Rather than focusing solely on automated species identification, the manuscript emphasizes a shift toward multi-specimen, expert-guided taxonomic interpretation. We use the term Taxonomic AI to describe this integrative framework, in which representation learning, curated specimen data, multimodal evidence, and professional taxonomic judgment are brought together within specimen-based workflows. Given the broad, interdisciplinary scope of the Journal of Plant Research, we believe this review will be of strong interest to its audience—especially to those working in plant systematics and biodiversity research, as well as to researchers interested in connecting traditional taxonomic practice with emerging advances in artificial intelligence. This manuscript is original, has not been published previously, and is not under consideration elsewhere. All authors have approved the submission and agree with its content. Thank you very much for your consideration. Sangtae Kim (on behalf of all authors) Department of Biology Sungshin Women's University

	Seoul, 01133 Republic of Korea
Response to Reviewers:	We have deleted the numbers from the titles of the sections, as well as the section numbers referred to in the text. Thanks. Sangtae Kim

Herbarium specimens in the age of artificial intelligence: from herbarium image identification to Integrative Taxonomic AI

Atsuko Takano^{1,2}, Hyeonji Moon³, Masato Shirai⁴, Shuichiro Tagane⁵, Diego Tavares Vasques⁶, Chung-Kun Lee^{2,1}, Amjad Khan⁷, Jeongwoo Seo³, Sujeong Han⁸, Chanhoo Park⁹, Jaesung Lee⁵, Takashi Akihiro¹⁰, and Sangtae Kim^{3*}

¹ Institute of Natural and Environmental Sciences, University of Hyogo, Kobe, Japan

² Museum of Nature and Human Activities, Hyogo, Sanda, Japan

³ Department of Biology, Sungshin Women's University, Seoul, Republic of Korea

⁴ Interdisciplinary Faculty of Science and Engineering, Shimane University, Matsue, Japan

⁵ The Kagoshima University Museum, Kagoshima University, Korimoto, Japan

⁶ Graduate School of Science, The University of Tokyo, Tokyo, Japan

⁷ Graduate School of Human and Environmental Sciences, University of Hyogo, Kobe, Japan

⁸ Department of Artificial Intelligence, Chung-Ang University, Seoul, Republic of Korea

⁹ Fisheries Science Institute, Chonnam National University, Gwangju, Republic of Korea

¹⁰ Faculty of Life and Environmental Sciences, Shimane University, Matsue, Japan

ORCID

Atsuko Takano: 0000-0002-8345-5080

Hyeonji Moon: 0000-0001-8561-755X

Masato Shirai: 0009-0002-6047-8121

Shuichiro Tagane: 0000-0002-1974-7329

Diego Tavares Vasques: 0000-0001-7978-2516

Chung-Kun Lee: 0000-0003-3080-6905

Amjad Khan: 0000-0002-5337-7237

Jeongwoo Seo: 0009-0006-2294-6747

Sujeong Han: 0009-0002-9444-6240

Chanhoo Park: 0000-0003-4047-1918

Jaesung Lee: 0000-0002-3757-3510

Takashi Akihiro: 0000-0001-8349-3290

Sangtae Kim: 0000-0002-1821-4707

*Corresponding author:

Sangtae Kim (amborella@sungshin.ac.kr)

Abstract

Herbarium specimens are physical, verifiable records that form the basis of taxonomic knowledge and biodiversity research. Their large-scale digitization has produced extensive collections of high-resolution images and associated specimen metadata, creating conditions in which artificial intelligence (AI) can play an important role in plant taxonomy, collection management, and ecological research. Early AI applications have primarily focused on automated species identification based on individual specimen images. Although increasingly accurate, such approaches remain limited by their emphasis on single-specimen label prediction and by treating identification outputs as final analytical decisions. Recent methodological advances—including segmentation-based preprocessing, automated trait extraction, structured extraction of label data, detection of potentially misidentified specimens, and multimodal integration of visual, textual, and genetic information—extend AI applications beyond species identification toward broader analytical frameworks, encompassing taxonomic interpretation as well as ecological and biodiversity research. In these approaches, specimens are placed within a shared analytical space, and identification results are used to support comparisons across multiple specimens rather than being treated as final decisions for single individuals. This multi-specimen perspective enables quantitative examination of species boundaries, morphological variation, data inconsistencies, and taxonomic stability within curated collections. In this context, AI serves not as an ultimate decision-maker but as a decision-support tool embedded in expert-guided workflows and biodiversity knowledge infrastructures. These developments can be summarized as *Integrative Taxonomic AI*, an approach that employs learned morphospaces to interpret and refine taxonomic categories by integrating multimodal evidence and curated specimen data under expert guidance.

Keywords:

Artificial intelligence, Automated species identification, Herbarium specimens, Integrative Taxonomic AI, Morphospace

Introduction

Since Luca Ghini, a botanist and physician at the University of Bologna, devised the format of botanical specimen sheets in the 16th century, herbarium specimens have served as essential research materials for studying plants and their diversity (Heberling et al. 2019), providing physical and verifiable references that form the basis of taxonomic knowledge. Herbaria enabled the systematic assembly of dispersed observations and records of individual species at a single location, functioning as centralized repositories of biological information, analogous to modern digital data centers that integrate distributed data streams.

For centuries, the collectors, curators, and users of these specimens were primarily plant taxonomists. Today, approximately 3,800 herbaria worldwide collectively house nearly 400 million specimens (Thiers 2025). Herbarium specimens are thus inseparably linked to plant taxonomy, which aims at a comprehensive and systematic understanding of plant diversity. In recent decades, this traditional taxonomic role has expanded substantially. Since the establishment of the Global Biodiversity Information Facility (GBIF) in 1999, large-scale digitization and archiving efforts have accelerated the use of herbarium specimens and other natural history collections for biodiversity research by making specimen images and associated metadata widely accessible in standardized digital formats. This transformation is exemplified by the emergence of major virtual herbaria worldwide. Examples include the Integrated Digitized Biocollections (iDigBio; <https://www.idigbio.org/>) in the United States, the Distributed System of Scientific Collections (DiSSCo) in the European Union, the Australasian Virtual Herbarium (AVH; <https://avh.chah.org.au/>), the Re flora Virtual Herbarium (<https://reflora.jbrj.gov.br/>) in Brazil, the Virtual Herbarium Germany (VH/de; <https://herbarium.gbif.de/>), and the Chinese Virtual Herbarium (CVH; <https://www.cvh.ac.cn/>) (reviewed in Hussein et al. 2022; Nelson and Ellis 2019; Takano et al. 2019). There are also institution-level portals such as that of the Royal Botanical Gardens, Kew (<https://data.kew.org/>).

Together, these initiatives have made millions of specimen images and associated metadata openly accessible, enabling large-scale, image-based analyses that were previously impractical. In iDigBio alone, 45.6% of the 150 million specimen records and 81% of the 66 million media files are plant-derived (accessed November 8, 2025), illustrating the scale to which plant specimen data are now available digitally. The availability of such vast image resources has broadened the use of herbarium data beyond traditional taxonomic research to fields such as ecology, biodiversity science, and evolutionary biology. At the same time, this unprecedented data volume has created favorable conditions for data-intensive computational approaches, including machine-learning methods that rely on large, well-labeled datasets. Although taxonomic and geographic coverage remains uneven, publicly accessible herbarium images now exist in quantities sufficient to support large-scale image-based modeling efforts.

This review synthesizes recent advances in applying artificial intelligence (AI) to digitized herbarium specimens, with particular emphasis on automated species classification, image segmentation,

trait extraction, label data extraction, and emerging multimodal approaches that integrate image, textual, and genetic data. While several recent reviews have summarized methodological developments in herbarium image analysis (e.g., Guo et al. 2025; Hussein et al. 2022), the present review emphasizes large-scale, real-world herbarium applications, including identification challenges related to dataset structure and morphology, misidentification detection, regional implementation in Asia, and broader implications for taxonomy, collection management, and biodiversity research. In this context, recent developments point beyond AI-assisted species identification toward broader forms of taxonomic interpretation within learned morphological representation spaces (morphospace *sensu lato*; Budd 2021; Cavanagh et al. 2021; Raup 1966), marking a shift in how AI-derived outputs are interpreted within taxonomic workflows. This conceptual transition is summarized in Fig. 1, which outlines the overall analytical framework linking AI-based processing, data integration, and taxonomic interpretation.

Advances in automated plant image classification

Efforts to automate plant identification date back more than five decades (Morse 1971), and automated identification systems based on plant images began to emerge in the 1990s (reviewed by Edwards and Morse 1995; Pankhurst 1991). By the early 2000s, large-scale digitization initiatives at major herbaria in Europe and North America significantly expanded the availability of digitized herbarium specimen images (Nelson and Ellis 2019; Thiers 2024). This progress facilitated advancements in automated species identification systems (e.g., Belhumeur et al. 2008; White et al. 2006). Rapid improvements in image classification techniques have driven recent progress in plant image analysis, including applications to herbarium specimens (see also Soltis et al. 2020 for an overview of machine learning applications in plant biology). International competitions aimed at improving image recognition have played a significant role, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC, 2010–2017) and ImageCLEF, part of the Cross Language Evaluation Forum (CLEF). Within ImageCLEF, LifeCLEF focuses on species identification and is subdivided into PlantCLEF (plants), BirdCLEF (birds), and AnimalCLEF (animals). The specific themes and configurations of these tasks vary slightly from year to year, and PlantCLEF has been held annually since 2011. In plantCLEF, the input images have evolved from controlled leaf scans to multi-organ photographs and, more recently, to complex images containing multiple plant species within a plot. Thus, input images have shifted from controlled leaf scans to diverse, real-world, multi-organ and multi-species scenes. In addition, cross-domain datasets, such as images of herbarium specimens and living plants, have been used (Carranza-Rojas et al. 2017; Chulif et al. 2023; Ong et al. 2025). These advances soon led to the development of automated plant identification applications, such as LeafSnap (Kumar et al. 2012) and Pl@ntNet (Joly et al. 2016). LeafSnap demonstrated species identification from isolated leaves photographed under controlled conditions, while Pl@ntNet extended automated identification to images of living plants captured in natural environments. Beyond angiosperms, identification systems have also been developed

for other plant groups (e.g., freshwater algae: Geronimo et al. 2023; bryophytes: Luo et al. 2025), as well as for non-plant organisms such as cyanobacteria (Topaloglu et al. 2025).

Agricultural applications also emerged early, including non-destructive estimation of fruit maturity (oil palm: Fadilah et al. 2012; strawberry: Liu et al. 2014), detection of leaf diseases (Sladojevic et al. 2016), and identification of abnormalities in crop fields (Christiansen et al. 2016). This research area continues to grow rapidly. In 2025 alone, numerous image-based systems for detecting pests and diseases in crops such as grape (Al-Shamayleh and Ibrahim 2025), tea (Wen et al. 2025), rice (Raman and Jayaraman 2025), soybean (Zhang et al. 2025), and date palm (Mir et al. 2025) were developed. In parallel, high-resolution hyperspectral imaging has been used to identify crop provenance (e.g., wolfberry: Li et al. 2025), detect early disease symptoms (Goyal et al. 2025), and classify cereal seeds (Hu et al. 2025).

Within this broader context, automated plant image classification has been developed across a wide range of image types, including scans of pressed material, images of living plants captured in natural or agricultural settings, and images captured under controlled conditions. Digitized herbarium specimen images represent a distinct visual domain: specimens are two-dimensionally flattened and may exhibit color fading or physical damage. Specimens are typically imaged under standardized conditions, with consistent backgrounds, fixed viewpoints, and accompanying scale and color references. These characteristics facilitate large-scale comparative analysis and model training but also introduce challenges, including two-dimensional compression and non-morphological artifacts such as labels and mounting materials. Recognition of both the strengths and constraints of specimen images has led to dedicated research on automated species identification in herbarium collections, discussed in the following section.

Development of automated species identification systems using herbarium images

Automated identification of herbarium specimens using AI-based models has gained attention as herbaria confront growing backlogs and limited taxonomic capacity. The ongoing digitization of specimens has generated a continuous influx of new material requiring timely identification, while large numbers of specimens already housed in collections have not been systematically re-examined or updated. In addition, many herbaria retain substantial backlogs of unidentified specimens from historical collections (sometimes unmounted), reflecting the cumulative effects of limited taxonomic capacity over time. This situation is not confined to small or under-resourced institutions; even large, internationally recognized herbaria face similar challenges. For example, historical collections from the Korean Peninsula collected by Emile Joseph Taquet at the Royal Botanic Garden Edinburgh (E) and by Takenoshin Nakai at the University of Tokyo (TI) continue to contain numerous unidentified specimens, thereby complicating collection management and taxonomic synthesis. These challenges have stimulated the development of deep learning-based species identification systems using herbarium specimen images

since the late 2010s.

To date, most studies have emphasized model development and benchmarking for specific taxonomic groups or regional floras rather than sustained deployment in herbarium operations. Table 1 summarizes representative examples of such studies, illustrating that existing research has largely emphasized model construction and performance evaluation across large and taxonomically diverse groups (Carranza-Rojas et al. 2017; de Lutio et al. 2022; Younis et al. 2018), region-specific floras (Park et al. 2024; Shirai et al. 2022; Xie et al. 2025), and taxonomically challenging taxa characterized by high morphological similarity (Arno et al. 2025; Kajihara et al. 2025).

Performance evaluation has played a central role in herbarium-based image classification research, with studies commonly relying on two evaluation metrics: Top-1 accuracy and macro-averaged F1 scores (Russakovsky et al. 2015; Sokolova and Lapalme 2009). Top-1 accuracy provides an intuitive measure of overall classification success and remains a standard metric in large-scale image recognition benchmarks (Russakovsky et al. 2015). In contrast, macro-averaged F1 scores weight performance equally across all taxa, regardless of their representation in the dataset, and are therefore particularly informative for herbarium-derived datasets that exhibit severe class imbalance. Accordingly, many studies report both metrics, as overall accuracy alone can obscure poor performance on underrepresented taxa common in herbarium datasets (de Lutio et al. 2022; Shirai et al. 2022).

Regarding model design, a clear evolution in architectural choices can be observed across the studies summarized in Table 1. Early and mid-scale investigations predominantly employed convolutional neural networks (CNNs), including Inception (Szegedy et al. 2015), ResNet (He et al. 2016), EfficientNet (Tan and Le 2019), and their variants, achieving moderate to high classification accuracy depending on dataset composition and curation level (Carranza-Rojas et al. 2017, 2018; Little et al. 2020; Shirai et al. 2022; Tan et al. 2019; Xie et al. 2025). More recent work has increasingly incorporated vision transformers (ViTs) (Dosovitskiy et al. 2021) and hybrid CNN–ViT architectures (Chen et al. 2021; d'Ascoli et al. 2021; Wu et al. 2021), motivated by their ability to model long-range dependencies and global morphological structure through self-attention mechanisms (Arno et al. 2025; Kajihara et al. 2025; Park et al. 2024). Beyond single-model architectures, several studies have therefore adopted ensemble strategies that combine predictions from independently trained CNN- and transformer-based models. Such approaches aim to exploit complementary error patterns rather than introduce new representational mechanisms and frequently outperform single-model baselines in large-scale, highly imbalanced datasets, particularly in terms of macro-averaged F1 scores (de Lutio et al. 2022; Little et al. 2020; Park et al. 2024).

Despite these advances in model architecture and evaluation strategies, practical constraints related to image resolution remain a significant bottleneck in herbarium-based image classification. A key methodological issue is the gap between the high resolution of digitized herbarium specimens and the much smaller image sizes used for deep learning models. Recently digitized herbarium specimens often

exceed 20 megapixels per image, enabling observation of fine details such as surface hairs. However, in most AI-based studies, these images are resized to approximately 60~260 kilopixels (196 X 292–512 X 512) (Table 1). This reduction reflects the substantial computational demands associated with high input resolutions, including increased GPU memory usage and longer training times. Consequently, a considerable amount of diagnostically useful fine morphological information is not being fully utilized in the current classification pipeline. Importantly, evidence from studies employing higher-resolution inputs or multi-scale strategies suggests that retaining finer morphological detail can improve classification performance, particularly for taxonomically challenging groups (Kajihara et al. 2025; Park et al. 2024; Shirai et al. 2022; Vaishnav et al. 2022). However, increasing input image size inevitably raises computational demands, making the selection of an appropriate resolution a practical trade-off between morphological resolution and feasibility. Systematic, simulation-based evaluations that jointly assess image resolution, performance gains, and computational cost would therefore be valuable for identifying optimal input sizes for each taxon.

Collectively, the studies summarized in Table 1 demonstrate that herbarium-based automated identification has matured methodologically but remains constrained by persistent biological and data-related constraints. These constraints can be broadly categorized as intrinsic morphological complexity and extrinsic dataset limitations. We first consider the intrinsic factor: morphological complexity, which imposes fundamental constraints on achievable performance. Plant identification from images represents a fine-grained classification problem, as closely related species often differ only in subtle morphological traits (Guo et al. 2025). In addition, herbarium specimens frequently exhibit significant morphological variation within species across different life-cycle stages (e.g., flowering versus fruiting). At the same time, some taxa possess extremely small or obscured reproductive organs, which require microscopic examination and dissection, thereby making species-level discrimination more difficult. Examples of such intrinsic difficulty include taxonomically challenging genera such as *Bulbostylis*, *Salix*, and *Iris*, in which reliable species delimitation often depends on reproductive characters that are frequently absent in herbarium specimen images (Arno et al. 2025; Park et al. 2024; Shirai et al. 2022). Additionally, low-resolution training images can obscure diagnostic features, such as hair structures, thereby having a particularly significant impact on small-bodied taxa. Taken together, these intrinsic limitations reflect the biological properties of the specimens themselves and cannot be fully resolved through improvements in model architecture or expanded training datasets. In this context, future studies may also benefit from focusing on taxon-specific models targeting morphologically complex or species-rich groups (e.g., Poaceae, Cyperaceae, Salicaceae, Fabaceae, and Asteraceae), rather than broad, multi-family classification.

We next consider the extrinsic factor: dataset-related constraints that arise from data quality and long-tailed class distributions before model training. At the dataset level, data quality is the most critical extrinsic factor, encompassing both the reliability of specimen identification through expert validation

and the intrinsic quality of the specimens. Accordingly, expert validation remains crucial, and non-specimen images (e.g., field photographs, botanical illustrations) or specimen records containing insufficient morphological information (e.g., specimens consisting only of twigs) should be removed during the filtering process (Arno et al. 2025; de Lutio et al. 2022; Kajihara et al. 2025; Shirai et al. 2022). Another major extrinsic constraint is the extremely long-tailed distribution of herbarium datasets, characterized by highly skewed class frequencies: widespread species are represented by many images, whereas only a few images represent rare or narrowly distributed species. This pattern is further reinforced by uneven geographic sampling, as easily accessible lowland regions tend to be overrepresented, while alpine or remote habitats remain comparatively undersampled. Such an imbalance inherently leads to reduced classification performance, necessitating consideration of both minimum and maximum image counts per species when constructing training datasets (Arno et al. 2025; Kajihara et al. 2025; Park et al. 2024; Shirai et al. 2022; Xie et al. 2025). In practice, restricting analyses to taxa with sufficient image counts improves accuracy but reduces the number of taxa included, creating a clear trade-off between coverage and performance. Several technical strategies have been proposed to mitigate data imbalance, including ensemble models, increased input image resolution, and specialized loss functions such as ArcFace (Deng et al. 2020) to address class imbalance (de Lutio et al. 2022; Little et al. 2020; Park et al. 2024).

In response to these methodological and dataset-level challenges, a major milestone was the series of "Herbarium Challenge" competitions hosted by the New York Botanical Garden (NYBG) on Kaggle from 2019 to 2022 as part of the Fine-Grained Visual Categorization (FGVC) workshop series (FGVC6–FGVC9). The Herbarium 2019 dataset (FGVC6; 46,469 images, 683 taxa) focused on Melastomataceae and was carefully curated by taxonomic experts to reflect realistic long-tailed distributions of specimens (Little et al. 2020; Tan et al. 2019). The subsequent Herbarium 2020 dataset (FGVC7; 1,169,039 images, 32,094 taxa) expanded coverage to vascular plants from the Americas, although competition results have not been systematically reported in the literature. This effort was further extended by the Herbarium 2021 "Half-Earth" dataset (FGVC8; 2,500,779 images, 64,500 taxa), which increased geographic and taxonomic scope by integrating specimens from five major herbaria worldwide (de Lutio et al. 2022). However, despite its scale, it covered only about 14% of vascular plant species in the sampled regions (Park et al. 2024). To overcome this limitation, the Herbarium 2022 dataset (FGVC9), NAFlora-1M (1,050,182 images, 15,501 species), was developed using specimens from 54 North American herbaria and achieved 90.06% coverage of known vascular plant species in the region (Park et al. 2024).

Notably, recent studies demonstrate that automated identification models can also contribute to herbarium curation by assisting in the detection of potentially misidentified specimens. By highlighting specimens that are repeatedly or inconsistently classified across multiple training runs, such approaches enable curators to prioritize expert review and efficiently improve data quality (Hollister et al. 2025;

Shirai et al. 2022). Together, these developments underscore a shift from automated identification as a stand-alone classification task toward AI-assisted decision-support frameworks, in which models complement rather than replace taxonomic expertise, supporting identification, quality control, and iterative refinement within herbarium-based workflows.

The feasibility of deploying these approaches at scale is further shaped by regional differences in digitization capacity and data availability, an issue that is particularly pronounced in Asia.

Current status of herbarium specimen digitization in Asia

Asia harbors exceptionally high plant diversity, reflecting its wide climatic range, complex topography, and heterogeneous vegetation patterns, and includes multiple biodiversity hotspots of global conservation importance (Myers et al. 2000; Tietje et al. 2023). Despite this biological richness, Asia has long been regarded as a data-poor region in terms of digitally accessible plant specimen data compared with Europe and North America (Wang et al. 2020). According to GBIF records, only 5.2 million of the 59.4 million digitized plant specimen images worldwide (8.78%) originate from Asia. Moreover, only 34.31% of the 15.2 million plant specimens recorded from Asia have been digitized (accessed December 25, 2025). This substantial discrepancy underscores the persistent gap between Asia's biodiversity richness and its representation in global digital repositories.

To address this imbalance, several coordinated initiatives have been launched to improve the accessibility and integration of Asian plant biodiversity data. Among these, the Mapping Asia Plants (MAP) project under the Asian Biodiversity Conservation and Database Network (ABCDNet) aims to establish foundational infrastructure through standardized vascular plant checklists and distribution databases (Li et al. 2020; Sun et al. 2024; Wang et al. 2020; Xiao et al. 2023; Xu et al. 2020; Xue et al. 2020). Complementary digitization efforts also target Asian plant specimens housed in overseas collections, particularly in the United States (e.g., All Asia TCN; <https://allasiatcn.org/index.php>).

In parallel with these initiatives, the online availability of digitized herbarium specimen images within Asia has expanded rapidly. Major virtual herbaria and institutional portals are summarized in Table 2. However, this expansion remains remarkably uneven. While the Chinese Virtual Herbarium (CVH) hosts nearly seven million images, many other national and university herbaria face challenges with incomplete digitization or restricted web access. This digital heterogeneity poses a fundamental challenge: the lack of a uniform, continent-wide dataset makes the development of a universal AI identification system impractical.

In response to these structural constraints, AI research in Asia has increasingly adopted modular approaches. Regional or taxon-specific identification systems have therefore emerged as more feasible alternatives (Shirai et al. 2022; Xie et al. 2025). By targeting restricted geographic regions or morphologically coherent taxonomic groups, researchers can achieve high identification accuracy even with manageable dataset sizes and limited computational resources. As digitization continues to fill

existing data gaps, these localized efforts are paving the way for more integrated and expansive systems. The following sections review specific case studies that demonstrate how these targeted AI approaches are being implemented to address the structural and data-related constraints of Asian herbarium digitization.

Segmentation-based preprocessing for removing non-plant information

Herbarium specimen images contain substantial amounts of non-plant information that can confound AI-based analyses if not adequately addressed. Representative examples of these preprocessing strategies and their effects on model attention are shown in Fig. 2. In addition to biologically meaningful plant structures, specimen images routinely include non-plant components such as labels, barcodes, color palettes, rulers, background areas outside the specimen, and digitization artifacts introduced during curatorial and imaging practices (Fig. 2a). These sources of non-biological variation—arising from differences in specimen mounting, paper aging, and digitization methods—can divert model attention away from diagnostically informative plant structures and bias predictions toward herbarium-specific visual cues (Ariouat et al. 2025b; Arno et al. 2025; Han et al. 2026; Sklab et al. 2025; Tessler and Little 2025). As a result, deep learning models trained on unprocessed herbarium images can exhibit reduced generalization across taxa and herbaria, particularly for taxa represented by few specimen images (Han et al. 2026). This limitation was also demonstrated by Xie et al. (2025), who reported a species identification accuracy of only 24.5% when applying a Chinese vascular plant identification system to herbarium images from the Royal Botanic Gardens, Kew, highlighting the impact of inter-herbarium image heterogeneity on model generalization.

To mitigate these effects, several preprocessing strategies have been developed to reduce or neutralize non-plant information in herbarium specimen images. A common approach is to blur textual regions, which effectively suppresses explicit textual information such as labels and barcodes and has demonstrated robust generalization performance (Fig. 2b). This approach has been widely adopted in major benchmark datasets, including Herbarium 2019, 2021, and 2022, in which text and barcode regions were detected using models such as PhotoOCR (Bissacco et al. 2013), EAST (Zhou et al. 2017), and CRAFT (Baek et al. 2019), followed by Gaussian blurring (de Lutio et al. 2022; Park et al. 2024; Tan et al. 2019). Similarly, Arno et al. (2025) applied CRAFT-based text detection and blurring to specimens from Cyperaceae and Rhamnaceae and further removed non-plant components and background regions outside specimen margins through cropping. Despite their practicality, text blurring approaches have inherent limitations. Because non-textual elements—including background paper texture, color palettes, photographs, and digitization artifacts—remain intact, these methods do not fully standardize image inputs and may still allow models to rely on herbarium-specific visual biases. In addition, incomplete detection of textual regions and incidental loss of morphological information have been reported in large-scale applications (de Lutio et al. 2022).

From both a conceptual and biological perspective, segmentation-based preprocessing provides the most comprehensive and logically coherent framework for removing non-plant information. By isolating plant regions at the pixel level in specimen images, segmentation ensures that classification models learn plant-derived morphological features (Fig. 2c). Unlike text blurring approaches, this strategy addresses non-plant variation at its source and offers a unified framework for image standardization across herbaria. Segmentation-based preprocessing has been applied to a wide range of analytical tasks using herbarium images, including trait identification, morphological measurement, and species classification. In trait analyses, segmentation has been used to isolate specific organs, such as leaves (Hussein et al. 2021; Triki et al. 2022b; Weaver et al. 2020) and reproductive structures (Goëau et al. 2020). Beyond organ-level approaches, whole-plant segmentation has been applied to extract trait information (Ariouat et al. 2025a; Ariouat et al. 2025b; Sklab et al. 2025). It has also been extended to species identification in particular taxonomic groups, including ferns (White et al. 2020), Piperaceae (Kajihara et al. 2025), and *Viola* (Han et al. 2026; Lee et al. 2025).

Moreover, beyond improvements in classification accuracy, segmentation substantially enhances the biological interpretability of deep learning models (Ariouat et al. 2025a). Gradient-weighted Class Activation Mapping (Grad-CAM) highlights image regions that most strongly influence model predictions (Selvaraju et al. 2017). These analyses consistently show that models trained on segmented images predominantly attend to plant structures (Fig. 2e), whereas models trained on unprocessed or text-blurring images often focus on non-plant components (Ariouat et al. 2025a; Ariouat et al. 2025b; Shirai et al. 2022; Xie et al. 2025) (Fig. 2d).

Despite these advantages, segmentation-based preprocessing remains technically challenging. Accurate segmentation is challenging when plant organs are thin (e.g., peduncles, petioles, stems, and roots), densely overlapping, or similar in color to the background in dried specimens (Ariouat et al. 2025a; Ariouat et al. 2025b; Han et al. 2026; Kajihara et al. 2025; Sklab et al. 2025). Such challenges can result in partial loss of morphological information and, in some cases, marginal or inconsistent improvements in classification performance (Ariouat et al. 2025a; Sklab et al. 2025). In Ariouat et al. (2025a, 2025b), images processed with segmentation showed marginal improvements in classification performance for relatively large organs, such as leaves and fruits, whereas performance decreased slightly for thin or small structures, such as thorns and flowers. Similarly, in a Piperaceae species identification study by Kajihara et al. (2025), no substantial difference in classification accuracy was observed between segmented and unprocessed images. These observations indicate that, although segmentation provides the most conceptually robust preprocessing framework, its practical effectiveness depends on segmentation quality, target traits, and taxonomic context.

Shirai et al. (2022) experimentally demonstrated that, under low-resolution input settings (e.g., 299×299 pixels), non-plant elements—including textual information—did not significantly reduce classification accuracy. They therefore suggested that explicit removal of non-plant components is not

necessary to improve performance. However, Grad-CAM analyses revealed that models sometimes still attend to such non-plant elements during prediction (Fig. 2d), suggesting that models may still use herbarium-specific visual features, which can reduce their performance on specimens from other herbaria.

Taken together, segmentation appears to offer a particularly comprehensive approach for removing non-plant information in AI-based analyses of herbarium specimen images. While alternative methods such as text blurring offer practical advantages for large-scale datasets, segmentation more directly enforces biological relevance, facilitates cross-herbarium standardization, and enhances model interpretability. Beyond preprocessing for species identification, segmentation has also been widely applied to trait-based analyses, as discussed in the next section. As segmentation algorithms continue to improve (e.g., Han et al. 2026; Sklab et al. 2025), this approach is likely to play an increasingly central role in the development of reliable, transparent, and taxonomically meaningful AI workflows for herbarium-based research (Ariouat et al. 2025a).

Automated extraction of trait information

Building on advances in AI-based image analysis techniques, including classification, object detection, and segmentation, automated extraction of trait information from herbarium specimen images has become possible, enabling large-scale phenological studies that were previously limited by labor-intensive manual approaches (Pearson et al. 2020; Soltis et al. 2020). With more than 300 years of accumulated collections, European herbaria provide unique opportunities to reconstruct long-term ecological and evolutionary dynamics (Williamson et al. 2025). Herbarium specimens thus represent valuable resources for investigating long-term biological responses to environmental change, including climate-driven shifts in phenology (Katal et al. 2022).

Early large-scale phenological studies using herbarium specimens relied primarily on manual scoring of reproductive status. For example, Hart et al. (2014) analyzed approximately 10,000 *Rhododendron* specimen images from across the Himalayas and demonstrated that flowering times have advanced in recent decades. While their study relied on manual scoring, Williamson et al. (2025) subsequently developed an AI-based classifier capable of distinguishing flowering from non-flowering specimens using more than eight million images representing approximately 200,000 species, along with GBIF metadata. By comparing pre-1970s and post-1970s data across World Wildlife Fund (WWF) terrestrial ecoregions, they revealed globally advanced flowering phenology, with particularly strong shifts in tropical regions. These findings addressed a major gap in large-scale phenological research, which had previously focused mainly on temperate regions. Similarly, Bateux et al. (2025) used an AI-based classification model to automatically extract phenological states, including budding, flowering, fruiting, and non-reproductive stages, from approximately 600,000 herbarium specimen images representing around 700 species, and documented widespread shifts in flowering time.

Beyond phenology, automated trait extraction has been increasingly applied to studies of

taxonomy, evolution, ecology, plant adaptation to climate change, and the construction of large-scale trait datasets (Weaver et al. 2023). Whereas early phenological studies focused mainly on reproductive organs such as flower buds, flowers, and fruits (Lorieul et al. 2019), automated trait extraction has since expanded to non-reproductive organs, including leaves, roots, and spines (Sklab et al. 2025; Weaver et al. 2023). These approaches include the classification of organs or phenological states (Ariouat et al. 2025b; Lorieul et al. 2019; Sklab et al. 2025; Younis et al. 2020), organ detection (Ott et al. 2020; Triki et al. 2022a), organ counting (Davis et al. 2020; Goëau et al. 2020), and organ measurement (Pearson et al. 2020; Weaver et al. 2020; Weaver et al. 2023).

Leaves have received particular attention because they constitute a large proportion of herbarium specimens, are consistently preserved, and exhibit relatively stable and quantifiable morphological characteristics. Moreover, many leaf traits are closely associated with climate and environmental gradients (Li et al. 2020; Wright et al. 2017). For example, Wilde et al. (2023) demonstrated systematic variation in leaf size along latitudinal and climatic gradients in *Syzygium* and *Ficus*. LeafMachine2, developed by Weaver et al. (2023), provides an automated framework for extracting quantitative leaf traits, including leaf area, perimeter, length, and width, as well as shape-related traits such as aspect ratio and lobedness. It can also detect plant organs beyond leaves. In addition, several models have been developed to identify detailed leaf characters, including arrangement, structure, form, margin, and venation (Ott et al. 2020; Younis et al. 2018).

Automated approaches have also been applied to the analysis of biotic interactions. Meineke et al. (2020) developed a model capable of automatically detecting two major types of insect herbivory (interior feeding and leaf margin feeding), building on earlier evidence that insect damage on herbarium specimens has increased over the past century (Meineke et al. 2019). The automation of herbivory detection enables analyses at unprecedented spatial and temporal scales, thereby providing new insights into long-term plant–insect interactions.

Phenological and morphological traits are fundamental to species delimitation, the interpretation of morphological variation, and the analysis of plant responses to environmental change. By automating trait extraction from specimen images, recent approaches extend these analyses from small, manually curated datasets to continent- or global-scale biodiversity studies. However, the interpretability and reliability of such large-scale inferences depend not only on accurate species identification but also on temporally and geographically structured label metadata. The automated extraction and structuring of such metadata, which complements image-based trait analysis, are discussed in the next section.

Automated extraction of label information

Although hundreds of millions of herbarium specimens exist worldwide, fewer than half are currently digitized and accessible online. Because large-scale ecological and evolutionary inferences depend critically on collection dates and locality information, the analytical value of specimen images is

realized only when label metadata are reliably structured and interoperable. Manual transcription, however, remains a major bottleneck in digitization workflows, particularly for historical collections containing handwritten or multilingual labels. Consequently, automated approaches to label extraction have become essential for accelerating metadata generation and improving data quality at scale.

In response to this need, contemporary systems have converged on modular, multi-stage architectures. Reliable label extraction typically follows a pipeline in which (1) label regions are localized within specimen images, (2) textual content is transcribed using optical character recognition (OCR) or handwritten text recognition (HTR), and (3) extracted text is transformed into structured metadata through downstream semantic processing. Because specimen images contain multiple visual elements—including plant material, barcodes, color charts, and mounting artifacts—accurate label localization is a prerequisite for robust transcription and subsequent structuring. OCR-based extraction has been applied to herbarium labels since the early 2010s (Drinkwater et al. 2014). Takano et al. (2024) employed named entity recognition (NER) to structure OCR-derived text into discrete entities, including collection dates, localities, and scientific names, in both Japanese and English. While effective for clearly formatted fields, rule-based or entity-based systems often struggle with narrative descriptions, such as habitat notes or remarks. More recently, large language model (LLM)–based approaches have introduced greater flexibility in semantic interpretation. Guralnick et al. (2025) integrated natural language processing with ChatGPT-4 to extract Darwin Core–compliant metadata (Darwin Core: Wiczorek et al. 2012), and Stenhouse and Thrall (2025) developed SpeciMate, an open-source system that similarly leverages LLMs to generate structured data. This progression reflects a broader shift from rule-based entity extraction toward context-aware semantic modeling. However, as many LLMs are frequently updated or controlled by external providers, their outputs may vary over time, making reproducibility and long-term consistency difficult to ensure. This highlights the need for careful validation and documentation.

A major limitation of conventional OCR is its difficulty in deciphering handwritten text, which is prevalent in historical collections. Advances in HTR have begun to mitigate this challenge. Sadek et al. (2024) demonstrated automated detection of printed and handwritten annotations in historical botanical works, and Turnbull et al. (2025) introduced Hespi, a pipeline combining Tesseract OCR for printed text with TrOCR (Li et al. 2023) for handwritten elements, followed by multimodal LLM-based validation. While effective for English-language materials, reliance on TrOCR currently limits applicability across diverse linguistic contexts. Multilingual platforms such as Transkribus (Colutto et al. 2019), a deep-learning–based handwriting recognition system trainable on custom ground-truth data, offer broader language coverage, though performance depends on the availability of high-quality training corpora. In Japan, Chagi and Kawai (2025) further demonstrated that vision–language models (VLMs), including ChatGPT and Qwen2.5-VL-7B, achieved over 90% accuracy in recognizing handwritten labels at the herbarium of the University of Tokyo (TI). Collectively, these developments illustrate expanding technical capacity for automated transcription across historical and multilingual collections, albeit with

ongoing challenges in generalization, language coverage, and model stability.

Importantly, automated label extraction complements advances in automated trait analysis. When semantically structured label metadata are integrated with image-derived morphological traits, herbarium specimens become multi-layered, machine-readable data entities that combine temporal, geographic, and morphological dimensions. Rather than functioning solely as inputs for species-level classification, specimens increasingly serve as integrated repositories of biological evidence within shared analytical frameworks. This convergence also supports downstream multimodal analyses, while providing a foundation for later discussion of more integrative, decision-support uses of AI in taxonomy.

Multimodal approaches for herbarium-based biodiversity studies

Multimodal learning refers to the integration of heterogeneous data types—such as images, text, genetic sequences, ecological metadata, or behavioral information—within a unified analytical framework. In contrast, most previous AI-based species identification studies have relied primarily on a single modality, typically specimen or field images. Although image-based models have achieved substantial success, their reliance on visual information alone reduces robustness in taxonomically complex groups, where closely related species differ only in subtle morphological traits (Javed et al. 2025; Lhayani et al. 2025; Manolache et al. 2025). Incorporating complementary sources of evidence can therefore improve species-level inference. Herbarium specimens inherently provide such multimodal information, combining images with structured label metadata (Guo et al. 2025). Together with recent advances in extracting traits and label data, this enables integration within shared analytical frameworks.

Building on this technical feasibility, empirical studies demonstrate that multimodal AI achieves higher classification performance than unimodal approaches. In plants, models integrating specimen or field images with DNA barcode sequences have been applied to taxonomically large and morphologically challenging families such as Asteraceae, Poaceae, and Fabaceae (Kösters et al. 2025; Lhayani et al. 2025). Other approaches combine images with ecological and geographic metadata—including taxonomic rank, longitude, and locality—yielding measurable improvements over image-only models (Javed et al. 2025). Similar strategies have been reported across other organismal groups, including the integration of images with ecological and taxonomic metadata for animals and cryptic species complexes (Li et al. 2025; Manolache et al. 2025; Sastry et al. 2025), fusion of image and genetic data in animals (Kösters et al. 2025; Yang et al. 2022), combination of images with meristic characters in fishes (Masuda et al. 2020), and integration of microscopic images with expert-validated morphological descriptions in cyanobacteria (Blanco et al. 2025).

Beyond research settings, multimodal AI is increasingly embedded in widely used identification and vision–language systems. General-purpose models such as ChatGPT and Google Lens integrate visual inputs with contextual or textual information to generate structured interpretations. Plant identification platforms such as Pl@ntNet and Flora Incognita (Mäder et al. 2021) similarly combine

image-based recognition with user-provided contextual cues—such as organ type, habitat, or taxonomic hierarchy—to refine species predictions. These applications illustrate how multimodal reasoning has moved from experimental systems to operational tools, reflecting the practical viability of integrating heterogeneous data streams.

Multimodal integration also extends to textual taxonomic resources. Species descriptions constitute foundational components of taxonomic knowledge, yet their traditionally unstructured format has limited computational use. Recent LLM-based approaches have transformed descriptive texts into structured character databases (Marcos et al. 2025; Nainia et al. 2024; Schmidt-Lebuhn and Knerr 2025), enabling integration with image-derived features. For example, Sahraoui et al. (2023) developed a herbarium specimen image retrieval system based on a fine-tuned Contrastive Language–Image Pre-training (CLIP; Radford et al. 2021) model, allowing retrieval of specimens corresponding to queried morphological descriptions. Similarly, Lesperance et al. (2025) proposed an interpretable multimodal framework for rare arthropods that combines image-derived captions with extensive morphological text datasets. Together, these studies reflect a transition from unimodal prediction toward integrative representation learning, in which visual, textual, ecological, and genetic information are embedded within shared representation spaces.

Despite their promise, multimodal approaches face significant technical and conceptual challenges. Image data remain non-standardized across institutions, limiting cross-domain generalization. Integrating heterogeneous data streams increases model complexity and may exacerbate overfitting, particularly when training datasets are limited (Blanco et al. 2025; Javed et al. 2025; Kösters et al. 2025; Lhayani et al. 2025). Multimodal systems also require substantial computational resources, which constrain large-scale deployment. Furthermore, interpretability remains limited: although multimodal embeddings improve predictive performance, disentangling the relative contribution of each modality to final decisions remains difficult. These considerations underscore the need for rigorous evaluation frameworks, transparent model design, and robust cross-domain validation.

Reflecting the biological complexity underlying species boundaries, Karbstein et al. (2024) advocate an AI-based integrative taxonomy that combines genomic, morphological, ecological, and behavioral evidence to reduce subjectivity in species delimitation. In herbarium-based research, multimodal AI therefore represents a shift from specimen-level label prediction toward integrative taxonomic reasoning. By embedding image-derived morphology, structured label metadata, and genetic and ecological information within shared representation spaces, these systems enable cross-validation among heterogeneous lines of evidence. The broader implications of this transition—for taxonomy, collection curation, and biodiversity knowledge infrastructures—are examined in the following section.

Implications for taxonomy, curation, and biodiversity knowledge infrastructure: toward *Integrative Taxonomic AI*

Recent developments reviewed in previous sections—including automated species identification, segmentation-based preprocessing, automated trait extraction, label structuring, detection of potentially misidentified specimens, and multimodal integration—collectively signal a shift in how AI outputs are positioned within taxonomic research and collection management. Rather than functioning solely as tools for specimen-level label prediction, AI systems increasingly contribute to higher-level inference across sets of specimens. As summarized in Table 3, this shift reflects a transition from traditional species identification AI toward a broader framework of taxonomic interpretation within AI-assisted representation spaces. In this emerging perspective, identification results are not treated as final decisions but as intermediate evidence within broader analytical workflows.

A central enabling concept underlying this shift is the embedding of specimens within learned morphological representation spaces (morphospace *sensu lato*; Budd 2021; Cavanagh et al. 2021; Raup 1966). When specimens are situated in shared representation spaces, AI systems can assist in visualizing morphological structure, examining species boundaries, detecting outliers, and prioritizing material for expert review. This multi-specimen framework supports quantitative assessment of morphological variation and taxonomic stability while maintaining interpretation within expert-guided contexts. Multimodal integration further strengthens this approach by combining image-derived morphology with structured label metadata and, increasingly, genetic and ecological information (Javed et al. 2025; Kösters et al. 2025). The goal of such integration is not to eliminate uncertainty, but to generate richer and more comparable lines of evidence across specimens. In this way, AI may facilitate the discovery and formal description of overlooked taxa by identifying coherent but atypical specimen clusters that warrant further examination (Bebber et al. 2010).

These developments also affect biodiversity knowledge infrastructures. Interactive online floras and continuously updated regional checklists (Renner 2025) show how AI can support identification, quality control, and data consistency in community-based platforms. In such systems, AI assists curation and interpretation rather than merely assigning species names, contributing to more responsive biodiversity documentation.

At the same time, the increasing accessibility of AI-based identification tools highlights the continued importance of taxonomic expertise. Although predictive performance has improved substantially, automated systems remain susceptible to misclassification, domain-specific bias, and overconfidence (Arno et al. 2025; Wäldchen et al. 2018). Reliable species delimitation, evaluation of morphologically ambiguous specimens, validation of automated outputs, and refinement of taxonomic concepts continue to depend on expert judgment grounded in comparative morphology and nomenclature. AI, therefore, functions most effectively within human-guided workflows that integrate computational efficiency with disciplinary expertise (Guo et al. 2025).

Taken together, these developments suggest a reorientation of AI in taxonomy: from isolated label prediction toward multi-specimen, evidence-based taxonomic interpretation embedded within

shared representation spaces and curated metadata infrastructures. For clarity, we refer to this integrated perspective as *Integrative Taxonomic AI* (Table 3). The term synthesizes ongoing methodological and conceptual advances rather than introducing a separate technological category. Within this framework, AI augments taxonomic reasoning by organizing, visualizing, and cross-validating heterogeneous evidence, while final interpretive authority remains grounded in expert evaluation.

Conclusion

Advances in image-based, trait-based, label-based, and multimodal AI have transformed herbarium specimens from static archives into active analytical resources. More importantly, these developments signal a shift from specimen-level label prediction toward multi-specimen, decision-support-oriented interpretation embedded within taxonomic workflows. In this emerging framework, AI outputs are treated not as conclusive determinations but as intermediate inputs that support comparative reasoning across collections.

Physical specimens, therefore, remain indispensable as authentic and verifiable references for training, validating, and refining AI-driven analyses. The perspective synthesized in this review—here referred to as *Integrative Taxonomic AI*—describes the integration of learned representation spaces, curated specimen data, and professional taxonomic judgment. Rather than replacing expertise, this framework reorganizes the relationship between computational systems and human interpretation in biodiversity research.

Acknowledgement

A. Takano and H. Moon contributed equally to this work as first authors. A. Takano, J. Lee, T. Akihiro, and S. Kim contributed equally as corresponding authors. The specimen shown in Fig. 2 was originally produced at Kangwon National University (KWNU) and gifted to the Korea National Arboretum (KNA). We thank both institutions for allowing us to use these specimen images. This research was supported by the bilateral program between the Japan Society for the Promotion of Science (Grant No. JSPS120258806 and JSPS24K09576 to A. Takano) and the National Research Foundation of Korea (Grant No. RS-2025-00508625 and RS-2021-NR063105 to S. Kim).

Declarations

Conflict of interest

No conflicts of interest declared.

References

- Al-Shamayleh AS, Ibrahim RW (2025) Grapevine disease detection using (q, τ) -nabla calculus quantum deformation with deep learning features. *MethodsX* 15:103619. doi: 10.1016/j.mex.2025.103619
- Ariouat H, Pimpare EP, Chenin E et al (2025a) Interpretable deep learning for botanical traits: a comparative study on the role of segmentation in herbarium image analysis. *Procedia Comput Sci* 270:3738–3747
- Ariouat H, Sklab Y, Prifti E et al (2025b) Enhancing plant morphological trait identification in herbarium collections through deep learning-based segmentation. *Appl Plant Sci* 13:e70000. doi: 10.1002/aps3.70000
- Arno J, Morel J, Rasaminirina F et al (2025) A pipeline to compile expert-verified datasets of digitised herbarium specimens for automated plant identification to accelerate taxonomy. *Plants People Planet*. doi: 10.1002/ppp3.70149
- Baek Y, Lee B, Han D et al (2019) Character region awareness for text detection. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit*, 9365–9374
- Bateux Q, Koss J, Sweeney PW et al (2025) Improving the accuracy of automated labeling of specimen images datasets via a confidence-based process. *PLoS Comput Biol* 21:e1013650
- Bebber DP, Carine MA, Wood JPI et al (2010) Herbaria are a major frontier for species discovery. *Proc Natl Acad Sci USA* 107:22169–22171. doi: 10.1073/pnas.1011841108
- Belhumeur PN, Chen D, Feiner S et al (2008) Searching the world's herbaria: a system for visual identification of plant species. In: Forsyth D, Torr P, Zisserman A (Eds.): *ECCV 2008*, part IV, LNCS 5305, 116–129
- Bissacco A, Cummins M, Netzer Y, Neven H (2013) Photoocr: reading text in uncontrolled conditions. In: *Proc IEEE Int Conf Comput Vis*, 785–792
- Blanco M, Ruiz-Santaquiteria J, Cristóbal G et al (2025) Multimodal deep learning for cyanobacteria classification: a fusion of CNN and transformer architectures. *Aquat Ecol* 59:1319–1339
- Budd GE (2021) Morphospace. *Curr Biol* 31:R1141–R1185
- Carranza-Rojas J, Goëau H, Bonnet P et al (2017) Going deeper in the automated identification of herbarium specimens. *BMC Evol Biol* 17:181
- Carranza-Rojas J, Joly A, Goëau H et al (2018) Automated identification of herbarium specimens at different taxonomic levels. In: *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*, Springer, 151–167
- Cavanagh H, Mosbach A, Scalliet G et al (2021) Physics-informed deep learning characterizes morphodynamics of Asian soybean rust disease. *Nat Commun* 12:6424
- Chagi K, Kawai H (2025) Deciphering handwritten herbarium specimen labels using vision-language models. Poster session abstract (P2-184), *Ecol Soc Japan 72nd annual meeting*. <https://esj.ne.jp/meeting/abst/72/P2-184.html> (in Japanese)

- Chen X, Wang X, Changpinyo S et al (2021) How to train your ViT? Data, augmentation, and regularization in vision transformers. *arXiv:2106.10270* doi: 10.48550/arXiv.2106.10270
- Christiansen P, Nielsen LN, Steen KA et al (2016) DeepAnomaly: combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors* 16:1904. doi: 10.3390/s16111904
- Chulif S, Lee SH, Chang YL, Chai KC (2023) A machine learning approach for cross-domain plant identification using herbarium specimens. *Neural Comput Appl* 35:5963–5985
- Colutto S, Kahle P, Guenter H, Muehlberger G (2019) Transkribus. A platform for automated text recognition and searching of historical documents. In: *Proc 15th Int Conf eScience*, 463–466
- d'Ascoli S, Touvron H, Leavitt M et al (2021) ConViT: Improving vision transformers with soft convolutional inductive biases. *arXiv:2103.10697* doi: 10.48550/arXiv.2103.10697
- Davis CC, Champ J, Park DS et al (2020) A new method for counting reproductive structures in digitized herbarium specimens using Mask R-CNN. *Front Plant Sci* 11:1129
- de Lutio R, Park JY, Watson KA et al (2022) The Herbarium 2021 Half-Earth Challenge dataset and machine learning competition. *Front Plant Sci* 12:787127
- Deng J, Guo J, Liu T et al (2020) Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In: Vedaldi A, Bischof H, Brox T, Frahm JM (eds) *European Conference on Computer Vision*. Springer, Cham, pp 741–757
- Dosovitskiy A, Lucas B, Kolesnikov A et al (2021) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv:2010.11929* doi: 10.48550/arXiv.2010.11929
- Drinkwater R, Cubey RWN, Haston E (2014) The use of Optical Character Recognition (OCR) in the digitisation of herbarium specimen labels. *PhytoKeys* 38:15–30. doi: 10.3897/phytokeys.38.7168
- Edwards M, Morse DR (1995) The potential for computer-aided identification in biodiversity research. *Trends Ecol Evol* 10:153–158
- Fadilah N, Mohamad-Saleh J, Abdul Halim Z et al (2012) Intelligent Color Vision System for Ripeness Classification of Oil Palm Fresh Fruit Bunch. *Sensors* 12:14179–14195. doi: 10.3390/s121014179
- Geronimo JO, Arguelles EDLR, Abriol-Santos KJM (2023) Automated Classification and Identification System for Freshwater Algae Using Convolutional Neural Networks. *Philipp J Sci* 152:325–335. doi: 10.56899/152.01.25
- Goëau H, Mora-Fallas A, Champ J et al (2020) A new fine-grained method for automated visual analysis of herbarium specimens: a case study for phenological data extraction. *Appl Plant Sci* 8:e11368. doi: 10.1002/aps3.11368
- Goyal SR, Kulkarni VS, Choudhary R, Jain R (2025) A comparative analysis of efficacy of machine learning techniques for disease detection in some economically important crops. *Crop Prot* 190:107093. doi: 10.1016/j.cropro.2024.107093
- Guo YY, Cai H, Bramley GL et al (2025) A review of artificial intelligence in herbarium specimen image

- analysis. *Art Intell Rev* 58:402
- Guralnick RP, Lafrance R, Allen JM, Denslow MW (2025) Ensemble automated approaches for producing high-quality herbarium digital records. *Appl Plant Sci* 13:e11623
- Hackel J, Imhof S, Zizka A (2026) Teaching plant identification at a university in the age of artificial intelligence. *Plants People Planet* 8:20–25
- Han S, Moon H, Lee S et al (2026) A survey of neural network segmentation and validation on plant specimen images of Korean Violaceae. *Expert Syst Appl* 296:128909
- Hart R, Salick J, Ranjitkar S, Xu J (2014) Herbarium specimens show contrasting phenological responses to Himalayan climate. *Proc Natl Acad Sci USA* 111:10615–10619. doi: 10.1073/pnas.1403376111
- He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. In: *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit (CVPR 2016)*, 770–778.
- Heberling JM, Prather LA, Tonsor SJ (2019) The changing uses of herbarium data in an era of global change: an overview using automated content analysis. *Bioscience* 69:812–822. doi: 10.1093/biosci/biz094
- Hollister JD, Martin G, Cai X et al (2025) A computer vision method for finding mislabelled specimens within natural history collections. *Ecol Evol* 15:e71648.
- Hu Y, Hongchen Z, Changming L et al (2025) Classification of maize seed hyperspectral images based on variable-depth convolutional kernels. *Front Plant Sci* 16:1599231
- Hussein BR, Malik OA, Ong W-H, Slik JWF (2022) Application of computer vision and machine learning for digitized herbarium specimens: a systematic literature review. *Ecol Inform* 69:101641. doi: 10.1016/j.ecoinf.2022.101641
- Hussein BR, Malik OA, Ong WH, Slik JWF (2021) Automated extraction of phenotypic leaf traits of individual intact herbarium leaves from herbarium specimen images using deep learning based semantic segmentation. *Sensors* 21:4549. doi: 10.3390/s21134549
- Javed H, Fahad LG, Tahir SF et al (2025) Enhanced Plant Species Identification through Metadata Fusion and Vision Transformer Integration. *Comput Mater Contin* 85:3981–3996. doi: 10.32604/cmc.2025.064359
- Joly A, Pierre B, Goëau H et al (2016) A look inside the Pl@ntNet experience. *Multimed Syst* 22. doi: 10.1007/s00530-015-0462-9
- Kajihara AY, Queiroz GAD, Caxambú MG et al (2025) A database for automatic identification of herbarium specimens in Piperaceae family. *Multimed Tools Appl* 1–29
- Karbstein K, Kösters L, Hodač L et al (2024) Species delimitation 4.0: integrative taxonomy meets artificial intelligence. *Trends Ecol Evol* 39:771–784
- Katal N, Rzanny M, Mäder P, Wäldchen J (2022) Deep learning in plant phenological research: a systematic literature review. *Front Plant Sci* 13:805738
- Kösters LM, Karbstein K, Hofmann M et al (2025) Data fusion for integrative species identification using

deep learning. *Syst Biol* 61:e026. doi: 10.1093/sysbio/syaf026

Kumar N, Belhumeur PN, Biswas A et al (2012) Leafsnap: a computer vision system for automatic plant species identification. In: Fitzgibbon A, Lazbnik S, Perona P et al (eds) *Computer Vision – ECCV 2012*. LNCS 7573. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-33709-3_36

Lee S, Moon H, Kim S, Lee J (2025) Multiple kernel-enhanced encoder for effective herbarium image segmentation. *Electron Lett* 61:e70155. doi: 10.1049/ell2.70155

Lesperance N, Ratnasingham S, Taylor G (2025) Taxonomic Reasoning for Rare Arthropods: Combining Dense Image Captioning and RAG for Interpretable Classification. arXiv:2503.10886 doi: 10.48550/arXiv.2503.10886

Lhayani K, Rabeh K, Medraoui L (2025) Deep learning for Fabaceae identification by integrating molecular and morphological data and a solution for barcode selection. *J Syst Evol*. doi: 10.1111/jse.70021

Li B, Xia R, Li J et al (2025) Multimodal deep learning with hyperspectral imaging for accurate origin classification of wolfberries. *Food Chem X* 31:103166

Li M, Lv T, Chen J et al (2023) Trocr: transformer-based optical character recognition with pre-trained models. In: *Proc AAAI Conf Artif Intell*, 37(11): 13094–13102. doi: 10.1609/aaai.v37i11.26538

Li W, Tojibaev KS, Hisoriev H et al (2020) Mapping Asia Plants: Current status of floristic information for Central Asian flora. *Glob Ecol Conserv* 24:e01220

Li Y, Zhao D, Qiao T et al (2025) MetaWild: A Multimodal Dataset for Animal Re-Identification with Environmental Metadata. In: *Proc ACM Int Conf Multimed*, 13009–13015. doi: 10.1145/3746027.3758249

Li Y, Zou D, Shrestha N et al (2020) Spatiotemporal variation in leaf size and shape in response to climate. *J Plant Ecol* 13:87–96

Little DP, Aguero B, Shaw AJ, Tessler M (2025) AI for difficult herbarium specimens: identification of peat mosses (subgenus *Sphagnum*) without dissection. *New Phytol*. doi: 10.1111/nph.70461

Little DP, Tulig M, Tan KC et al (2020) An algorithm competition for automatic species identification from herbarium specimens. *Appl Plant Sci* 8:e11365

Liu C, Liu W, Lu X et al (2014) Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. *PLoS ONE* 9:e87818. doi: 10.1371/journal.pone.0087818

Lorieul T, Pearson KD, Ellwood ER et al (2019) Toward a large-scale and deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras. *Appl Plant Sci* 7:e01233. doi: 10.1002/aps3.1233

Luo H, Zhang X, Yuan F et al (2025) MOSSNet: a lightweight dual-branch multi-scale attention neural network for bryophyte identification. *Symmetry* 17:347. doi: 10.3390/sym17030347

Mäder P, Boho D, Rzanny M et al (2021) The flora incognita app—interactive plant species identification. *Methods Ecol Evol* 12:1335–1342

- Manolache G, Schouten G, Vanschoren J (2025) CrypticBio: A Large Multimodal Dataset for Visually Confusing Species. arXiv:2505.14707 doi: 10.48550/arXiv.2505.14707
- Marcos D, van de Vlasakker R, Athanasiadis IN et al (2025) Fully automatic extraction of morphological traits from the web: Utopia or reality?. Appl Plant Sci 13:e70005. doi: 10.1002/aps3.70005
- Masuda H, Jukei T, Hasegawa T (2020) Fish species identification using a CNN-based multimodal learning method. In: Proc International Conference on Image, Video and Signal Processing (IVSP), 15–19. doi: 10.1145/3388818.3389164
- Meineke EK, Classen AT, Sanders NJ, Davies TJ (2019) Herbarium specimens reveal increasing herbivory over the past century. J Ecol 107:105–117. doi: 10.1111/1365-2745.13057
- Meineke EK, Tomasi C, Yuan S, Pryer KM (2020) Applying machine learning to investigate long-term insect–plant interactions preserved on digitized herbarium specimens. Appl Plant Sci 8:e11369
- Mir TA, Bharany S, Gupta R et al (2025) Hybrid Deep Learning Model for Date Palm Disease Classification: A Fusion of HybridConv Mixer and Vision Transformer. Food Sci Nutr 13:e71086. doi: 10.1002/fsn3.71086
- Morse LE (1971) Specimen identification and key construction with time-sharing computers. Taxon 20:269–282
- Myers N, Mittermeier RA, Mittermeier CG et al (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858
- Nainia A, Vignes-Lebbe R, Chenin E et al (2024) FloraNER: A new dataset for species and morphological terms named entity recognition in French botanical text. Data Brief 56:110824. doi: 10.1016/j.dib.2024.110824
- Nelson G, Ellis S (2019) The history and impact of digitization and digital data mobilization on biodiversity research. Phil Trans R Soc B 374:2017391. doi: 10.1098/rstb.2017.0391
- Ong CA, Tay FS, Then YL, McCarthy C (2025) An evaluation of a pre-trained transformer-based self-distillation model (DINOv2) for cross-domain plant species identification. Neural Comput Appl 37:21969–21995
- Ott T, Palm C, Vogt R, Oberprieler C (2020) GinJinn: an object-detection pipeline for automated feature extraction from herbarium specimens. Appl Plant Sci 8:e11351
- Pankhurst RJ (1991) Practical taxonomic computing. Cambridge University Press, Cambridge
- Park J, de Lutio R, Rappazzo B et al (2024) NAFlora-1M: continental-scale high-resolution fine-grained plant classification dataset. J Data-Centric Mach Learn Res 1:1–21
- Radford A, Kim J, Hallacy C (2021) Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning (ICML 2021), PMLR 139: 8748–8763
- Raman R, Jayaraman S (2025) Artificial intelligence for sustainable farming with dual branch convolutional graph attention networks in rice leaf disease detection. Sci Rep 15:10595. doi:

10.1038/s41598-025-94891-5

Raup DM (1966) Geometric analysis of shell coiling: general problems. *J Paleontol* 40:1178–1190

Renner S (2025) Linking to images and AI-based identification tools- The only way for flora projects to survive. *Plants People Planet*. doi: 10.1002/ppp3.70131

Pearson KD, Nelson G, Aronson MF et al (2020) Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70:610–620

Russakovsky O, Deng J, Su H et al (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115:211–252. doi: 10.1007/s11263-015-0816-y

Sadek J, Vlachidis A, Pickering V et al (2024) Leveraging OCR and HTR cloud services towards data mobilisation of historical plant names. *Int J Digit Humanit* 6:237–261. doi: 10.1007/s42803-024-00091-4

Sahraoui M, Sklab Y, Pignal M et al (2023) Leveraging multimodality for biodiversity data: exploring joint representations of species descriptions and specimen images using CLIP. *Biodiver Inf Sci Stand* 7:e112666

Sastry S, Khanal S, Dhakal A et al (2025) TaxaBind: A unified embedding space for ecological applications. In: *IEEE Winter Conf Appl Comput Vis*, 1765–1774. doi: 10.1109/WACV61041.2025.00179

Schmidt-Lebuhn AN, Knerr N (2025) Large Language Models can extract morphological data from taxonomic descriptions, but their stochastic nature makes automation challenging: a test on Australian Asteraceae. *PhytoKeys* 261:189–210

Selvaraju RR, Cogswell M, Das A et al (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proc IEEE Int Conf Comput Vis*, 618–626

Shirai M, Takano A, Kurosawa T et al (2022) Development of a system for the automated identification of herbarium specimens with high accuracy. *Sci Rep* 12:8066

Sklab Y, Castanet F, Ariouat H et al (2025) PlantSAM: an object detection-driven segmentation pipeline for herbarium specimens. *Appl Plant Sci* e70034. doi: 10.1002/aps3.70034

Sladojevic S, Arsenovic M, Anderla A et al (2016) Deep neural networks based recognition of plant diseases by leaf image classification. *Comp Intell Neurosci* 2016:3289801

Soltis PS, Nelson G, Zare A et al (2020) Plants meet machines: prospects in machine learning for plant biology. *Appl Plant Sci* 8:e11371. doi:10.1002/aps3.11371

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45: 427–437. doi: 10.1016/j.ipm.2009.03.002

Stenhouse A, Thrall PH (2025) SpeciMate: improving metadata extraction from digitised biological specimens. *Biodiv Data J* 13:e160553

Sun J, Liu B, Rustiami H et al (2024) Mapping Asia plants: Plant diversity and a checklist of vascular plants in Indonesia. *Plants* 13:2281

Szegedy C, Liu W, Jia Y et al (2015) Going deeper with convolutions. In: Proc IEEE Conf Comput Vis Pattern Recognit, 1–9

Takano A, Cole TCH, Konagai H (2024) A novel automated label data extraction and database generation system from herbarium specimen images using OCR and NER. *Sci Rep* 14:112

Takano A, Horiuchi Y, Fujimoto Y et al (2019). Simple but long-lasting: A specimen imaging method applicable for small-and medium-sized herbaria. *PhytoKeys* 118:1–14

Tan KC, Liu Y, Ambroose B et al (2019) The Herbarium Challenge 2019 dataset. arXiv:1906.05372 doi: 10.48550/arXiv.1906.05372

Tan M, Le QV (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc Mach Learn Res, 97:6105–6114

Tessler M, Little DP (2025) On herbarium specimen images and artificial intelligence. *New Phytol.* doi: 10.1111/nph.70312

Thiers BM (2024) Strengthening partnerships to safeguard the future of herbaria. *Diversity* 16:36

Thiers BM (2025) The World's herbaria 2024: a summary report based on data from Index Herbariorum. Technical Report, William and Lynda Steere Herbarium, New York Botanical Garden. <https://sweetgum.nybg.org/science/wp-content/uploads/2025/01/TheWorldHerbaria2024-.pdf> (Accessed November 21 2025)

Tietje M, Antonelli A, Forest F et al (2023) Global hotspots of plant phylogenetic diversity. *New Phytol* 240:1636–1646

Topaloglu F, Kiziloluk S, Sert E, Yildirim M (2025) Automatic detection of harmful cyanobacterial genera using deep CNN models and artemisinin optimization. *Sci Rep* 15:33824

Triki A, Bouaziz B, Mahdi W (2022a) A deep learning-based approach for detecting plant organs from digitized herbarium specimen images. *Ecol Inform* 69:101590

Triki A, Bouaziz B, Mahdi W et al (2022b) Deep learning based approach for digitized herbarium specimen segmentation. *Multimed Tools Appl* 81:28689–28707. doi: 10.1007/s11042-022-12711-6

Turnbull R, Fitzgerald E, Thompson KM, Birch JL (2025) Hespri: A pipeline for automatically detecting information from herbarium specimen sheets. *BioScience* 75:637–648

Vaishnav M, Fel T, Rodríguez IF, Serre T (2022) Conviformers: convolutionally guided vision transformer. arXiv:2208.08900 doi: 10.48550/arXiv.2208.08900

Wäldchen J, Rzanzy M, Seeland M, Mäder P (2018) Automated plant species identification—Trends and future directions. *PLoS Comput Biol* 14:e1005993

Wang H, Soejima A, Chang Ks, Ma K (2020) Mapping Asia plants: Current status of floristic information for Northeast Asia. *Glob Ecol Conserv* 24:e01321

Wang SQ, Dong XY, Ye L et al (2023) Flora of Northeast Asia. *Plants* 12:2240

Weaver WN, Ng J, Laport RG (2020) LeafMachine: using machine learning to automate leaf trait extraction from digitized herbarium specimens. *Appl Plant Sci* 8:e11367

- Weaver WN, Smith SA (2023) From leaves to labels: building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. *Appl Plant Sci* 11:e11548
- Wen X, Liu Q, Xuanyuan T et al (2025) A lightweight convolutional neural network for tea leaf disease and pest recognition. *Plant Methods* 21. doi: 10.1186/s13007-025-01452-y
- White AE, Dikow RB, Baugh M et al (2020) Generating segmentation masks of herbarium specimens and a data set for training segmentation models using deep learning. *Appl Plant Sci* 8:e11352. doi: 10.1002/aps3.11352
- White S, Feiner S, Kopylec J (2006) Virtual vouchers: Prototyping a mobile augmented reality user interface for botanical species identification. In: *Proc IEEE Symp on 3DUI*, 119–126
- Wieczorek J, Bloom D, Guralnick R et al (2012) Darwin core: an evolving community-developed biodiversity data standard. *PLoS ONE* 7: e29715. doi: 10.1371/journal.pone.0029715
- Wilde BC, Bragg JG, Cornwell W (2023) Analyzing trait–climate relationships within and among taxa using machine learning and herbarium specimens. *Am J Bot* 110:e16167
- Williamson DR, Prestø T, Westergaard KB et al (2025) Long-term trends in global flowering phenology. *New Phytol*. doi: 10.1111/nph.70139
- Wright IJ, Dong N, Maire V et al (2017) Global climatic drivers of leaf size. *Science* 357:917–921
- Wu H, Xiao B, Codella N. et al (2021) CvT: Introducing convolutions into vision transformers. *arXiv:2103.15808* doi: 10.48550/arXiv.2103.15808
- Xiao C, Zhang Z, Ma K, Lin Q (2023) Mapping Asia plants: Historical outline and review of sources on floristic diversity in South Asia. *Plants* 12:1617
- Xie G, Xuan J, Luo HR et al (2025) AI-based identification of Chinese vascular plants from herbarium specimens: a tool for all herbaria with Chinese holdings. *New Phytol*. doi: 10.1111/nph.70443
- Xu X, Naqinezhad A, Ghazanfar SA et al (2020) Mapping Asia plants: Current status on floristic information in Southwest Asia. *Glob Ecol Conserv* 24:e01257
- Xue J, Chepinoga VV, Liu Y, Ma K (2020) Mapping Asia Plants: Historical outline and review of sources on floristic diversity in North Asia (Asian Russia). *Glob Ecol Conserv* 24:e01287
- Yang B, Zhang Z, Yang CQ et al (2022) Identification of species by combining molecular and morphological data using convolutional neural networks. *Syst Biol* 71:690–705
- Younis S, Schmidt M, Weiland C et al (2020) Detection and annotation of plant organs from digitised herbarium scans using deep learning. *Biodivers Data J* 8:e57090
- Younis S, Weiland C, Hoehndorf R et al (2018) Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Bot Lett* 165:377–383
- Zhang Y, Ruofei B, Mengxin G et al (2025) A lightweight deep convolutional neural network development for soybean leaf disease recognition. *Front Plant Sci* 16. doi: 10.3389/fpls.2025.1655564
- Zhou X, Yao C, Wen H et al (2017) East: an efficient and accurate scene text detector. In: *Proc IEEE Conf*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

925

Comput Vis Pattern Recognit, 5551–5560

Table 1. Examples of deep learning–based plant classification using herbarium images (multimodal-based methods not included)

Reference	Dataset						Input image size (pixels)	Model architecture		Model performance ^c	
	Taxonomic or geographic category	Source of dataset	# of images	# of taxa (Class)	Taxonomic rank	# of images per class		Architecture category ^b	Baseline architecture	Top-1 accuracy (%)	Macro F1 score
Carranza-Rojas et al. (2017)	-	Herbarium 1K-iDigBio	253,733	1,204	-	-	224 x 224	CNN	Inception v1 (GoogleNet)	79.6	-
Younis et al. (2018)	-	GBIF	830,408	1,000	-	532≤n≤5494	196 x 292	CNN	Modified ResNet	82.4	-
Tan et al. (2019); Little et al. (2020)	Melastomataceae	Herbarium 2019 (FGVC6)	46,469	683	1 family	20≤n≤865	448 x 448	CNN Ensemble	SeResNeXt-101, ResNet-50	89.8	-
de Lutio et al. (2022)	American, Oceanian, and Pacific species	Herbarium 2021 (FGVC8): Half-Earth	2,500,779	64,500	-	n≥3 (test set: n≤100)	256×256, 256×256, 352×352	CNN Ensemble	ResNest-101, ResNeXt-101-IBN-a, ResNeXt-101	84.5	0.757
Park et al. (2024)	90% of North American vascular plants (Canada, Greenland, United States of America)	Herbarium 2022 (FGVC9): NAFLora-1M	105,000	15,501	-	10<n≤100	384 x 384	CNN-ViT Ensemble	ResNeSt-101, EfficientNet-B6, DeiT-III-B, Swin-B, Swin-L, ConvNeXt-B	-	0.877
Shirai et al. (2022)	33.9 % of Japanese vascular plants ^a	Japanese herbaria (author collection, download)	500,554	2,171	192 families, 2064 species, 9 subspecies, 88 varieties, 10 forms	n≥50	299 x 299	CNN	Inception-ResNet-v2	96	0.923
Xie et al. (2025)	53.9% of Chinese vascular plants	Chinese Virtual Herbarium (CVH)	6,500,000	21,589	309 families, 3144 genera, 19723 species, 1866 infraspecific taxa	16≤n≤200	299 x 299	CNN	EfficientNetV2	67.3	-
Kajihara et al. (2025)	Brazilian Piperaceae	speciesLink	10,503	236	1 family, 3 genera	5≤n<240	512 x 512	ViT	ViT	-	0.37
			9,977	160		10≤n<240				-	0.4
			9,237	106		20≤n<240				-	0.45
Arno et al. (2025)	<i>Bulbostylis</i>	GBIF-Royal Botanic Gardens, Kew	4,936	87	1 genus	20≤n≤100	224 x 224	ViT	ViT-H/14	65.38	-
	<i>Ziziphus</i>		1,654	27	1 genus				ViT-B/16	72.61	-

^a % calculation based on the Flora of Japan (reviewed in Wang et al. 2023).^b CNN, Convolutional Neural Network; ViT, Vision Transformer.^c Only the best classification results were included.

Table 2. A list of the digital herbarium image portals in Asia (including transcontinental regions)

Country	Portal or herbarium name ^a	Number of images (Accessed on Feb., 2026)	URL
China	Chinese Virtual Herbarium (CVH)	8,559,422	https://www.cvh.ac.cn/
Russia	Moscow University Herbarium (MW)	1,403,266	https://plant.depo.msu.ru/
Japan	National Museum of Nature and Science (TNS)	388,916	https://db.kahaku.go.jp/webmuseum_en/
	Shimane University Digital Herbarium	332,278	http://tayousei.life.shimane-u.ac.jp/herbarium/
	Makino Herbarium Type Specimen Image Database (MAK)	ca. 740	http://ameba.i.hosei.ac.jp/BIDP/MakinoCD/makino/html_j/index0.html
	Kobe Gakuin University Botanical Specimens in Saikei Collection	ca. 20,000	https://db.kobegakuin.ac.jp/plant/index.html
	Museum of Nature and Human Activities, Hyogo (HYO)	262,461	https://nathist-net.jp/
	The Kagoshima University Museum, Herbarium (KAG)	171,184	https://www.museum.kagoshima-u.ac.jp/hyouhonsitu.html
	The University of Tokyo, type collection database (TI)	^b	https://umdb.um.u-tokyo.ac.jp/DShokubu_en/TShokubu.php
Taiwan	Herbarium of Taiwan Forestry Research Institute (TAIF)	ca. 579,000	https://taif.tfri.gov.tw/en/index.php
	Plants of Taiwan , National Taiwan University Herbarium (TAI)	187,843	https://tai2.ntu.edu.tw/
	Herbarium, Academia Sinica, Taiwan (HAST)	126,135	https://hast.biodiv.tw/en/
India	Indian Virtual Herbarium (IVH)	155,531 ^c	https://ivh.bsi.gov.in/
Nepal	National Herbarium and Plant Laboratories (KATH)	164,471	http://plantdatabase.kath.gov.np/
Korea	Korea Biodiversity Information System, Korea National Arboretum (KH)	^b	https://www.nature.go.kr/kbi/plant/smpl/selectPlantSmplGnrIsrc1.do
	Plant DNA Bank in Korea, Korea University (KUS)	22,000	http://pdbk.korea.ac.kr/index.asp
Singapore	Singapore Botanic Gardens (SING)	46,985	https://herbaria.plants.ox.ac.uk/bol/sing
Thailand	Forest Herbarium (BKF)	13,827	https://botany.dnp.go.th/index_eng.html

^a the characters in parentheses are abbreviations for the portal or herbarium^b not indicated on the website^c including vascular plants as well as algae, bryophytes, lichens, and fungi

Table 3. Differences between traditional species identification (ID) AI and integrative Taxonomic AI

Category	ID AI (Traditional species identification AI)	Integrative Taxonomic AI (Paradigm of the present study)
Definition	Species identification assisted by AI	Taxonomic interpretation within an AI-assisted morphospace, integrating AI-driven taxonomic data curation
Core question	How accurately can a species label be predicted from a given specimen image?	How can species boundaries, stability, and morphological variation be quantitatively represented, explored, and validated within a morphological representation space?
Role of AI	An automated decision-making system that directly outputs the final species identification	A higher-level decision-support framework that integrates identification outputs with AI-driven taxonomic data curation to interpret morphological structure and evaluate species boundaries
Scale of inference	Individual specimen-level inference	Multi-specimen and species-level inference within a morphospace

FIGURE

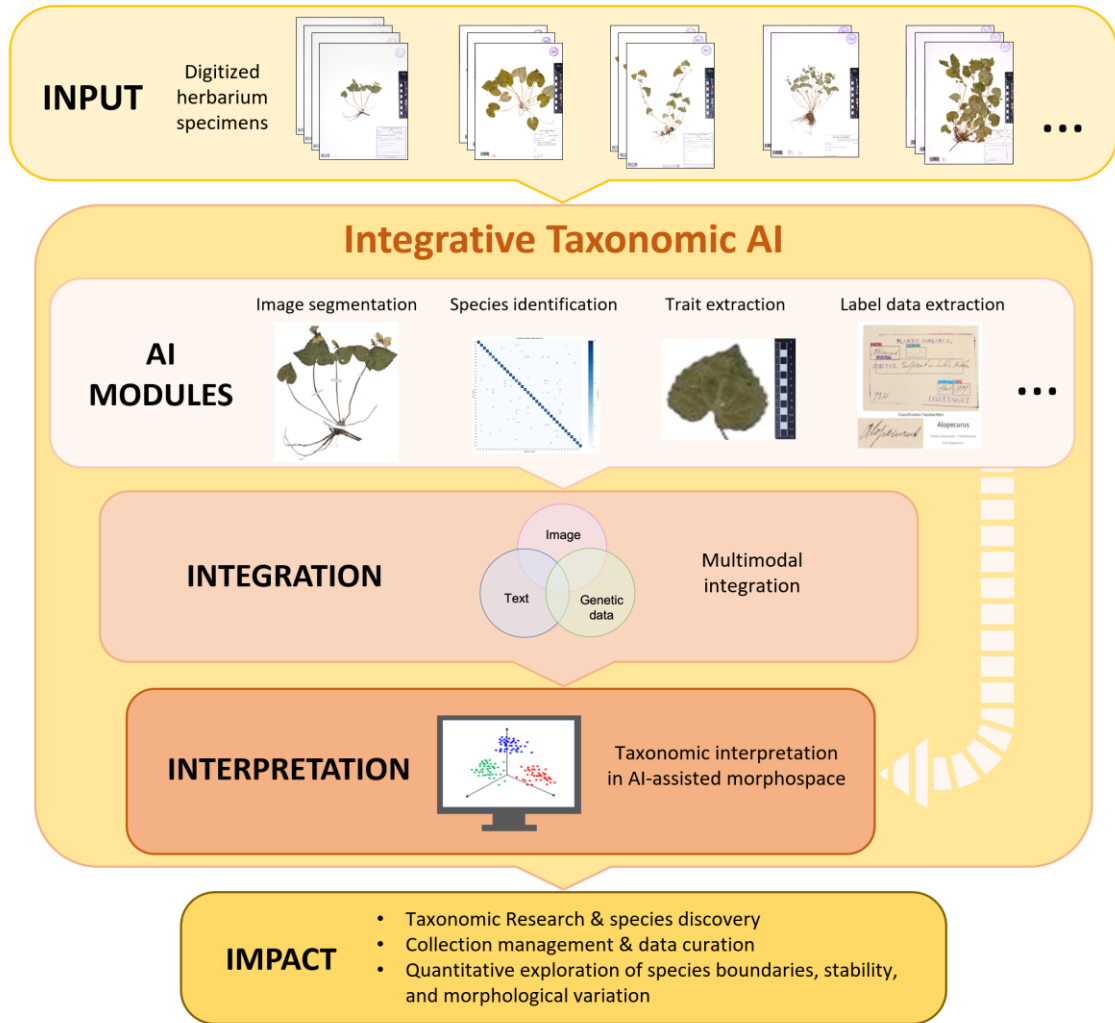


Fig. 1. Conceptual framework of Integrative Taxonomic AI. Digitized herbarium specimens are processed through AI modules, followed by multimodal integration and taxonomic interpretation within an AI-assisted morphospace. The dashed arrow indicates an alternative pathway in which interpretation is performed directly from individual module outputs without multimodal integration.

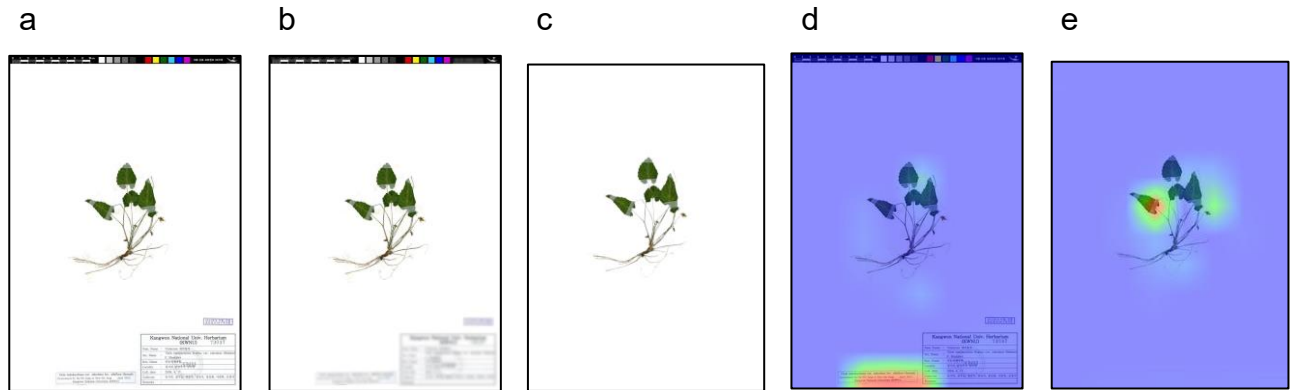


Fig. 2. Image preprocessing strategies and Grad-CAM–based interpretation in herbarium image analysis.

(a) Raw herbarium specimen image with non-plant components. (b) Text blurring. (c) Segmentation-based preprocessing. (d-e) Grad-CAM visualization highlighting model attention on raw (d) and segmented (e) images. In the raw image, the classification model often focuses on non-plant regions, whereas in the segmented image, it only focuses on plant structures.