Interpretable Deep Learning for Botanical Traits: A Comparative Study on the Role of Segmentation in Herbarium Image Analysis

Hanane Ariouat, Eva Perez Pimpare, Eric Chenin, Marc Pignal, Régine Vignes Lebbe, Souhila Arib, Edi Prifti, Jean-Daniel Zucker, Youcef Sklab

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 270 (2025) 3738–3747

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Interpretable Deep Learning for Botanical Traits: A Comparative Study on the Role of Segmentation in Herbarium Image Analysis

Hanane Ariouat[a], Eva Perez Pimpare[b], Eric Chenin[a], Marc Pignal[c], Régine Vignes Lebbe[c], Souhila Arib[d], Edi Prifti[a,e], Jean-Daniel Zucker[a,e], Youcef Sklab[a,*]

[a]*IRD, Sorbonne Université, UMMISCO, F-93143, Bondy, France*

[b]*Infrastructure Récolnat, Direction générale déléguée aux collections, Muséum national d'histoire naturelle. Paris, France*

[c]*Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, Université des Antilles, EPHE, 57, rue Cuvier, CP39, Paris 75005, France*

[d]*Laboratoire Etis, UMR 8051 CNRS, CY Cergy Paris université*

[e]*Sorbonne Université, INSERM, Nutrition et Obesities; systemic approaches, NutriOmique, AP-HP, France*

Abstract

The large-scale digitization of herbarium specimens—over 10 million in France via the Récolnat portal led by MNHN, and hundreds of millions worldwide—offers unprecedented opportunities for advancing plant biodiversity research. However, extracting reliable morphological traits from these images remains challenging due to the presence of non-vegetal elements (e.g., labels, rulers, envelopes) and visual background noise that can mislead model predictions. In this work, we introduce a comparative study of three deep learning models—YOLOv8, ViT, and ResNet101—applied to botanical trait classification from digitized herbarium images. We evaluate these models on both raw images and segmented versions where non-plant elements are removed. Although models trained on raw images can sometimes yield high accuracy, interpretability analysis (e.g., Grad-CAM) reveal that their predictions often rely on irrelevant background features. In contrast, segmentation consistently drives the models to focus on the plant itself, leading not only to improved performance for several morphological traits, but also laying the groundwork for explainable and scientifically meaningful AI. We argue that segmentation is not merely a preprocessing step but a prerequisite for trustworthy trait recognition and a necessary condition for enabling AI-driven discovery in botanical sciences. This work establishes a methodological foundation for interpretable plant-focused deep learning applied to biodiversity research.

Keywords: Deep learning, Herbarium Collections, Trait classification, Botanical recognition;

* Corresponding author.
E-mail address: youcef.sklab@ird.fr

1877-0509
10.1016/j.procs.2025.09.499

1. Introduction

Plants play a fundamental role in maintaining biodiversity, shaping ecosystems, and regulating atmospheric gases [1]. Herbarium collections, which document plant diversity over centuries, provide invaluable data for studying environmental changes and biodiversity loss [2, 3, 4]. With the large-scale digitization of collections, millions of specimens are now accessible online, on portals such as GBIF[1] and Recolnat[2], enabling the use of Deep Learning (DL) for automated plant analysis. These computational methods have been applied to extract morphological traits [26] (e.g., leaf size [5, 6]), detect plant organs, and classify diseases [5, 7, 8, 9, 10, 11]. However, despite these advances, digitized herbarium collections remain underutilized for trait-based research, largely due to the lack of trait-related annotations.

The National Museum of Natural History (MNHN) in Paris initiated a large-scale digitization project that led to the creation of Recolnat, a dataset comprising over 10 million high-resolution images. While this portal provides access to metadata such as species names, collector information, and sometimes geolocation, it lacks functionalities for searching based on taxon-related traits or specimen-specific traits. This limitation is particularly restrictive for researchers studying specific characteristics, such as soil residues on roots or leaf area variations. Given the impracticality of manually annotating such a vast collection, artificial intelligence (AI) presents a promising solution for automating metadata extraction. However, herbarium images pose unique challenges: labels, scale bars, and color charts introduce visual noise, while aging paper can resemble plant material, complicating DL-based classification (*see* Figure 1).

In recent years, CNN-based approaches have achieved remarkable success in species classification [12], leaf trait measurement [2, 5, 6], and segmentation [13]. Tools like LeafMachine2, trained on extensive herbarium datasets, have significantly enhanced the automated extraction of leaf traits. However, existing methods still struggle to detect finer morphological characteristics, such as leaf margin variations (smooth, serrated, crenulated) or the presence of thorns on stems. This lack of precision is a major limitation for detailed specimen analysis. Ontology-based approaches, such as the Flora Phenotype Ontology (FLOPO) [14], have been developed to structure plant trait data, while text-mining techniques have been explored for automated trait extraction. Younis et al. [15] and Lorieul et al. [16] demonstrated the potential of convolutional neural networks (CNNs) for plant trait classification. However, these studies have significant limitations. Firstly, in Younis et al.'s study, trait annotations were derived solely from textual descriptions, without verifying their actual presence in the images. This absence of visual validation raises doubts about the reliability of the results. Secondly, Lorieul et al.'s study is limited to examining three specimen-specific traits: the presence or absence of flowers, the presence or absence of fruits, and the distinction between fertile and non-fertile individuals. This restricted approach neglects a large portion of plant morphological diversity.

Fig. 1. Examples illustrating the diversity in paper color, quality and the non-plant elements in herbarium sheets (non-exhaustive).

To directly address these limitations, this study introduces a comparative evaluation of deep learning models for botanical trait recognition in herbarium specimens. Our main hypothesis is that the visual background present in raw images misleads model attention, particularly in small datasets, and that segmentation—by isolating the plant—enables models to learn on biologically relevant structures. We evaluate three deep learning architectures—YOLOv8, ViT, and ResNet101—on both raw and segmented images. Although models trained on raw images

[1] https://www.gbif.org/fr/
[2] www.recolnat.fr

can achieve reasonable accuracy, interpretability analyses (e.g., Grad-CAM) reveal that predictions often rely on irrelevant background regions. In contrast, segmentation significantly shifts the model's focus to the plant itself, leading to more trustworthy and explainable outcomes. This observation is not merely technical. As highlighted by Roscher et al. [17], explainability, interpretability, and scientific consistency are foundational prerequisites for deriving novel insights from machine learning in natural sciences. If the model's decisions are influenced by visual noise, no matter how accurate, they cannot support scientific interpretation. This study is the first to combine trait-specific classification, segmentation-based input cleaning, and attention-based analysis to examine how deep learning models behave on botanical data. These contributions establish a foundation for interpretable, biologically grounded AI pipelines tailored for biodiversity research. Segmentation is a key prerequisite for aligning AI perception with botanical expertise, enabling interpretable and reproducible discoveries.

2. Approach

Our methodological approach is designed not only to evaluate model performance but also to foster the development of explainable and scientifically grounded deep learning pipelines for botanical trait recognition. As illustrated in Figure 2, our goal is to align the entire image processing chain—from segmentation to classification—with the four key pillars of trustworthy AI in science: transparency, interpretability, explainability, and scientific consistency. Transparency is ensured through the use of open, well-documented architectures (YOLOv8, ViT, ResNet101) and clearly defined preprocessing procedures. Interpretability is addressed through Grad-CAM[18] visualizations, which highlight the image regions that most influence trait predictions. Explainability emerges from the alignment of these attention maps with expert-validated morphological traits, providing human-understandable justifications for the model's decisions. Scientific consistency extends beyond predictive accuracy, requiring that model outputs be biologically plausible and grounded in observable plant structures. This alignment is essential to unlock the potential of deep learning as a credible tool for hypothesis generation and scientific discovery in plant sciences.

To evaluate this framework, we constructed two parallel datasets from the same collection of 4,005 digitized herbarium specimens: one comprising raw, unprocessed images, and the other composed of segmented images where non-plant elements (e.g., labels, rulers, barcodes) were removed using a dedicated segmentation pipeline [19, 20]. This dual-dataset design enables a controlled comparison between models trained on noisy versus purified visual inputs. We evaluated three deep learning architectures—YOLOv8, Vision Transformer (ViT), and ResNet101—selected for their relevance to image classification tasks and architectural diversity. Each model was trained separately on both datasets and fine-tuned for five distinct botanical traits. Our pipeline includes image standardization to 640×640 pixels, zoom-based augmentation, and domain-informed binary trait labeling. Performance is assessed using standard metrics, while interpretability is evaluated via attention visualization. The core hypothesis guiding our work is that segmentation enhances both model performance and scientific validity by helping models focus on biologically meaningful structures. In particular, it improves the spatial localization of model attention and reduces reliance on irrelevant background features. As illustrated in the figure, predictions based on unsegmented images may achieve accuracy by focusing on non-biological cues, undermining their scientific interpretability. In contrast, segmentation enables a biologically grounded association between visual input, model attention, and trait prediction—ultimately leading to more robust, interpretable, and scientifically consistent outcomes.

2.1. Data Preparation

To effectively train deep learning models for trait identification in herbarium specimen images, we initiated a dedicated annotation campaign on the Les Herbonautes platform[3], under the project entitled *"Ce 'robotaniste', quel caractère?"*. This collaborative initiative, involving international contributors, focused on the annotation of visually observable morphological traits. In close coordination with botanical experts, we defined a set of 29 distinct traits encompassing key diagnostic features, including leaf margin type (smooth or toothed), stem structure (herbaceous or ligneous), and leaf size categories (less than 1 cm, between 1 cm and 10 cm, or greater than 10 cm). Additional

[3] http://herbonautes.mnhn.fr/

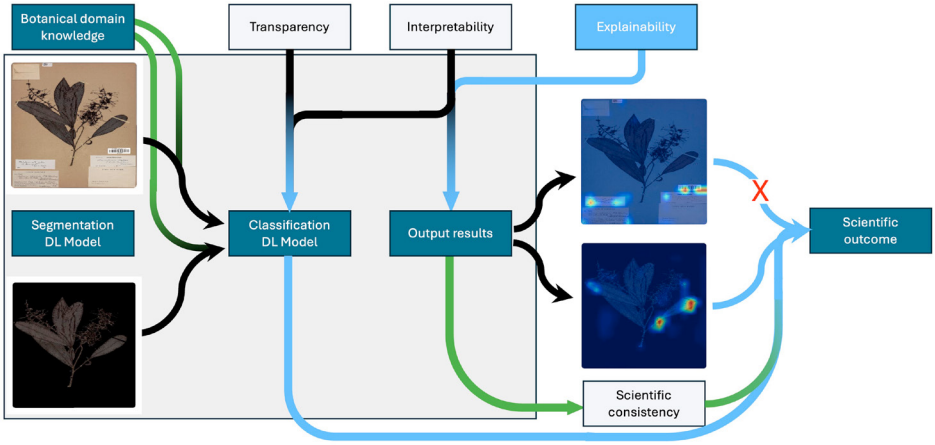
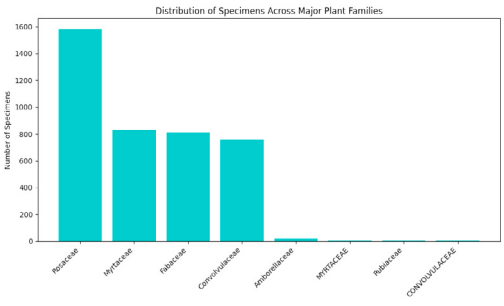
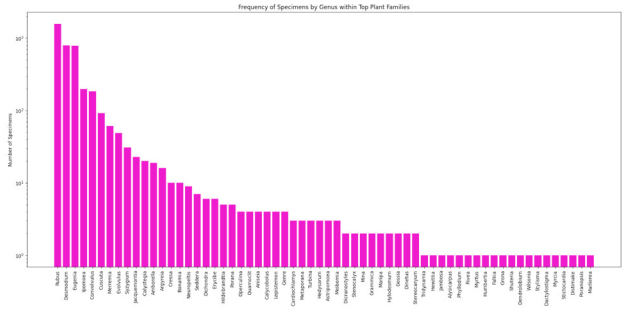


Fig. 2. Integrative pipeline linking DL learning predictions (dark gray) to scientifically valid outcomes in the context of herbarium collections. This framework highlights the foundational role of image segmentation in enabling explainability (blue) and scientific consistency (green). Starting from raw herbarium images, the segmentation model removes non-plant elements, producing a refined input for the classification model. When combined with botanical domain knowledge (green) and guided by the principles of transparency, interpretability (light gray), and explainability (blue), this pipeline supports the generation of biologically meaningful outputs. As illustrated by Grad-CAM heatmaps, models trained on unsegmented images may yield high accuracy while attending to spurious background regions—compromising scientific reliability. In contrast, segmentation redirects model attention toward plant structures, reinforcing the link between visual evidence, model decisions, and botanical traits.

features such as flower structure (simple or compound) and leaf shape (lobed, obovate, ovate) were also included. The complete list of traits is publicly available at http://lesherbonautes.mnhn.fr/missions/23879004. In total, contributors annotated 4,005 digitized specimens from the vascular plant collection of the Muséum national d'Histoire naturelle (MNHN) in Paris [21]. The resulting dataset spans 10 plant families and 61 genera, offering a high degree of botanical diversity. This initiative provides a precisely curated, expert-validated dataset that serves as a robust foundation for future research on automated trait recognition from herbarium images. Figures 3(a) and 3(b) illustrate the distribution of specimens across families and genera, respectively. For the purposes of this study, we selected a subset of these annotations, focusing on five specific traits. Among the 29 defined traits, we selected the five that had the highest number of annotations. The remaining traits had significantly fewer examples, with an average of fewer than 1500 annotated images per trait.



(a) Distribution of Specimen Counts Across Major Plant Families.



(b) Frequency of Specimens by Genus within the Top Plant Families.

Fig. 3. (a) Distribution of specimen counts across major plant families; (b) Frequency of specimens by genus within the top plant families.

The dataset generated through this effort holds significant potential for advancing research on phenotypic plasticity and adaptive trait variation across ecological gradients. It also opens avenues for integrating phylogenetic and genomic information to investigate the evolutionary trajectories of specific morphological traits. Table 1 summarizes the five selected traits, indicating the corresponding number of annotated specimens for each binary classification task: general

Table 1. Distribution of annotated specimens for each selected trait. Values indicate the number of images labeled as exhibiting (**Yes**) or not exhibiting (**No**) the corresponding trait.

Trait	Yes (# scans)	No (# scans)
General shape: obovate	897	3108
General shape: ovate	1128	2877
Occurrence of thorns	2444	1534
Presence of infructescence	1470	1965
Presence of flowers	3133	872

leaf shape obovate (yes or no), general leaf shape ovate (yes or no), presence or absence of thorns, presence or absence of flowers, and presence or absence of infructescences.

Given the observed class imbalances for each trait (as detailed in Table 1), we adopted a trait-specific, binary classification strategy. This approach enables the application of targeted optimization techniques to address the unique distribution and visual characteristics associated with each trait. By processing traits independently, we enhance both predictive accuracy and model robustness, while allowing for finer adaptation to the complex variability present in botanical data. Despite the extensive morphological diversity found in herbarium collections, the relatively modest size of our annotated dataset presents inherent limitations in terms of representativity. To compensate for this constraint and increase training diversity, we evaluated several data augmentation techniques and determined that random zoom augmentation yielded the most consistent performance gains. A 70%/30% training/validation split was adopted for each trait-specific dataset. For models generalization assessment, we curated independent test sets. These test sets consist exclusively of specimens belonging to families and genera that were excluded from the training phase. This experimental design ensures a good evaluation of the models' generalization capacity across unseen botanical taxa.

2.2. Image Preprocessing

To conduct our study, we prepared two parallel datasets based on the same set of 4,005 annotated herbarium images. The first dataset comprises unsegmented (raw) images, retaining all visual elements present on the herbarium sheets—including barcodes, envelopes, color palettes, text boxes, and measurement scales—which are generally considered as visual noise in the context of image-based trait classification (see Figure 4(a)). Although these elements are useful for human experts, particularly for cataloging and referencing, they can introduce confounding signals during model training, leading deep learning networks to inadvertently focus on non-biological features rather than the morphological characteristics of the plant specimen. To mitigate this issue, we constructed a second dataset composed of segmented images, where non-plant regions were masked or removed using a segmentation pipeline developed in our previous work [19, 20] (see Figure 4(b)). This pipeline was evaluated and demonstrated good performance, achieving an average F1 score of 96.6% and IoU values above 0.90 across most plant families. This segmentation step aims to isolate the vegetative structures and reduce the visual noise that may bias model learning. It helps redirect model attention toward botanically relevant regions. This dual-dataset design allows for a rigorous comparison of deep learning models under two preprocessing scenarios. It enables us to quantify not only the differences in predictive performance but also the extent to which model focus aligns with biologically meaningful structures. All images were standardized to a resolution of 640×640 pixels to ensure consistency across the dataset. To increase data variability and enhance models robustness, we initially explored several data augmentation strategies, including random rotations, zooming, and horizontal flipping. However, empirical evaluations revealed that combining multiple augmentation techniques led to a decline in model performance. As a result, we adopted a more targeted strategy, applying only zoom-based augmentation.

2.3. Selected Architectures for Botanical Trait Analysis

The selected architectures for this study represent three complementary paradigms in deep learning: convolutional networks, transformer-based models, and detection-oriented frameworks adapted for classification. The Vision Transformer (ViT) [22] leverages a transformer architecture originally developed for natural language processing, applying

(a) (b)

Fig. 4. Visual comparison between a non-segmented and segmented herbarium image. (a) The original image includes both plant and non-plant elements; (b) the segmented version isolates the plant, reducing background noise.

it to image patches to capture long-range dependencies—a property particularly useful for identifying global morphological patterns in botanical specimens. We used the pretrained *google/vit-base-patch16-224-in21k* model, adapted via transfer learning to our task. ResNet101 [23] is a deep convolutional neural network with 101 layers and residual connections that address the vanishing gradient problem. Its depth and robustness make it effective for extracting hierarchical features from herbarium images. We fine-tuned a pretrained version by modifying its classification head to align with our binary trait tasks. Finally, YOLOv8 [24], originally designed for object detection, was repurposed in its classification mode. We employed its medium variant, which offers a balance between performance and efficiency. This configuration was particularly suited for learning localized features from segmented images, especially in the context of traits like thorns or leaf shape.

3. Results and Discussion

This section presents a comparative analysis of the selected models, examining how segmentation influences classification performance, model generalization, and interpretability.

3.1. Model Training Configuration and Setup

To ensure optimal performance and robust generalization, we performed systematic hyperparameter tuning for each of the evaluated models. This process involved a series of controlled experiments examining the influence of key training parameters—such as batch size, learning rate, and weight regularization—on model stability and convergence dynamics. The final configuration was selected based on its ability to improve both training efficiency and validation accuracy. Across all models, we standardized the batch size to 12 and adopted a learning rate schedule with an initial rate (*lr0*) of 0.01 and a final rate (*lrf*) also set to 0.01, combined with a momentum coefficient of 0.95 and a weight decay of 0.0001. To stabilize training in the early epochs, we implemented a warm-up phase over 10 epochs, using a warm-up momentum of 0.5 and a warm-up bias learning rate of 0.1. The Adam optimizer was employed due to its ability to manage sparse gradients and adaptively adjust learning rates, thereby facilitating stable and rapid convergence. Notably, these configurations were particularly important for YOLOv8, which showed sensitivity to learning rate variations. The inclusion of a warm-up phase proved critical in preventing early divergence and enhancing parameter stability. We also applied a cosine learning rate scheduler (*Cos_lr*) and set a patience threshold of 25 epochs for early stopping. Each model was trained for a maximum of 300 epochs, with validation performance monitored periodically to avoid overfitting and ensure training consistency. Given the modest dataset size (4,005 annotated images), we further incorporated regularization strategies such as zoom-based data augmentation and label smoothing. These techniques improved training robustness and reduced overfitting.

Table 2. Accuracy (%) of YOLOv8, ViT, and ResNet101 models on segmented and unsegmented validation datasets across five botanical traits. The last column for each model reports the difference (Δ) computed as: $\Delta = \text{Accuracy}_{\text{seg}} - \text{Accuracy}_{\text{unseg}}$. Positive values indicate performance improvement due to segmentation, while negative values reflect performance degradation.

Trait	YOLOv8 (%)			ViT (%)			ResNet101 (%)		
	Seg.	Unseg.	Δ	Seg.	Unseg.	Δ	Seg.	Unseg.	Δ
General shape: obovate	86.8	86.5	+0.3	78.0	75.0	+3.0	78.9	77.0	+1.9
General shape: ovate	88.0	84.7	+3.3	72.0	70.0	+2.0	76.8	76.0	+0.8
Occurrence of thorns	96.6	96.2	+0.4	85.0	85.1	-0.1	96.0	94.0	+2.0
Presence of infructescence	65.0	71.3	-6.3	56.0	64.0	-8.0	55.6	63.8	-8.2
Presence of flowers	87.8	89.6	-1.8	77.0	79.1	-2.1	78.0	80.0	-2.2

3.2. Trait-Specific Evaluation of Model Performance

The results presented in Table 2 compare model performance on segmented and unsegmented validation datasets. These results provide insight into how image segmentation influences the accuracy of botanical trait classification across different architectures and morphological categories. For vegetative traits, particularly leaf shape, segmentation consistently led to improved performance. YOLOv8 demonstrated the largest gains, with an increase of +3.3% in accuracy for the ovate trait and +0.3% for the obovate trait. Similar trends were observed with ViT (+2.0% and +3.0%, respectively), while ResNet101 showed more modest improvements (+0.8% and +1.9%). These findings suggest that segmentation is effective in suppressing visual noise, thereby helping models to better focus on relevant plant structures.

In contrast, segmentation had a detrimental effect on traits that rely on broader contextual information, such as reproductive structures. All models exhibited decreased accuracy in predicting infructescence: -6.3% for YOLOv8, -8.0% for ViT, and -8.2% for ResNet101. A similar, albeit less pronounced, decline was observed for flower presence. While this performance degradation is partly due to the loss of contextual cues during segmentation, we hypothesize that it is also linked to the current limitations in segmentation quality. In several cases, we observed that the segmentation process inadvertently removed portions of the plant—particularly fine structures—leading to a loss of essential morphological details. This issue is especially critical for traits such as flowers or infructescences, where localized information plays a decisive role in classification. Improving segmentation fidelity, particularly in preserving complete plant morphology, may therefore offer a promising direction for mitigating these effects and enhancing overall model reliability. For thorn detection, segmentation had a negligible to moderate impact depending on the model: YOLOv8 recorded a slight improvement (+0.4%), ViT showed a minimal decrease (-0.1%), and ResNet101 improved moderately (+2.0%). These results collectively underscore that the effect of segmentation is highly trait-dependent and that refining segmentation quality and granularity could be a key lever for improving the consistency and accuracy of trait-specific predictions.

Although models trained on unsegmented images sometimes achieved higher raw accuracy, interpretability analyses using Grad-CAM revealed a critical shortcoming. These models frequently concentrated on irrelevant visual elements such as labels, barcodes, or background textures, which are not biologically meaningful. In contrast, segmentation helped redirect model attention toward valid morphological structures, thereby improving not only classification reliability but also scientific interpretability. These findings reinforce the importance of adopting selective and context-aware preprocessing techniques to balance the benefits of visual denoising with the preservation of biologically informative context (see Figure 5).

3.3. Comparison of Heatmap Focus Areas

To further investigate the impact of segmentation on model behavior, we analyzed Grad-CAM heatmaps generated from test images belonging to taxa excluded from both the training and validation phases, including Plumeria, Prunus, Bougainvillea, and Citrus. This evaluation focused on three representative traits: presence of flowers, obovate leaves, and ovate leaves. Each heatmap was manually examined in collaboration with a botanical expert to ensure accurate biological interpretation. The results reveal a clear shift in model attention following segmentation. For the flower presence trait, the proportion of attention focused on the background dropped markedly from 63.6% in the raw,

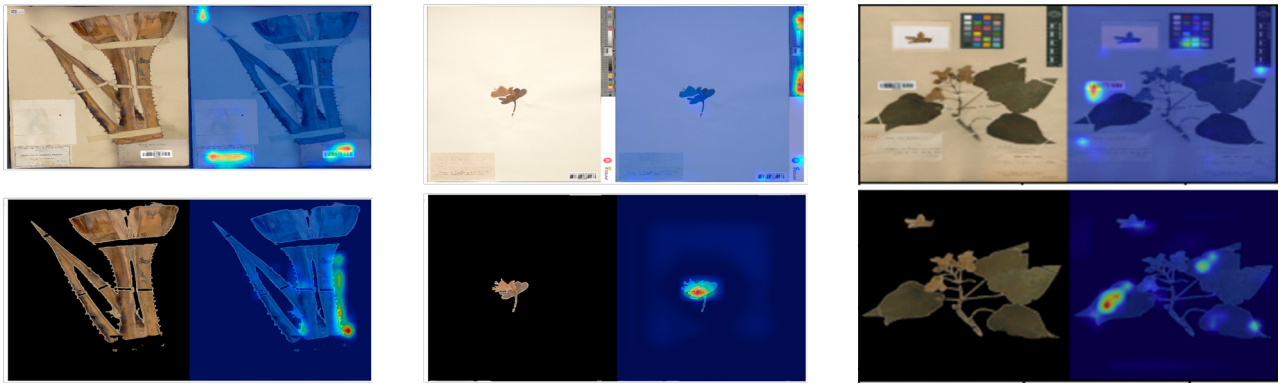


Fig. 5. Grad-CAM heatmaps illustrating model attention for three botanical traits: (a) presence of thorns, (b) leaf margin morphology, and (c) presence of flowers. For each trait, the top row shows the original (unsegmented) images and their corresponding attention maps, while the bottom row presents the segmented versions. These visualizations demonstrate how segmentation reduces irrelevant focus on background elements (e.g., labels, rulers) and redirects attention toward biologically meaningful plant structures, thereby enhancing interpretability and scientific consistency of model predictions.

unsegmented images to just 11.5% in the segmented versions. Simultaneously, attention directed toward floral or vegetative structures increased substantially, confirming that segmentation encourages the model to focus on botanically meaningful regions. Moreover, segmentation led to a notable improvement in the spatial precision of the attention maps: the proportion of focalized (i.e., well-localized and trait-specific) attention areas rose from 68.6% in the unsegmented case to 77.6% post-segmentation (Figure 6). Leaf shape traits followed a comparable trend (See Fig. 7). For the obovate category, segmentation completely suppressed background attention—reducing the number of heatmaps focusing on non-plant areas from 22 to 0—while simultaneously increasing the number of instances where attention was directed toward leaf structures from 40 to 49. Similarly, in the ovate category, background focus decreased from 11 to just 1 instance, accompanied by an increase in leaf-directed attention from 21 to 24. These results support our hypothesis that segmentation improves both interpretability and scientific consistency by filtering out irrelevant visual cues and directing model attention toward biologically meaningful structures.

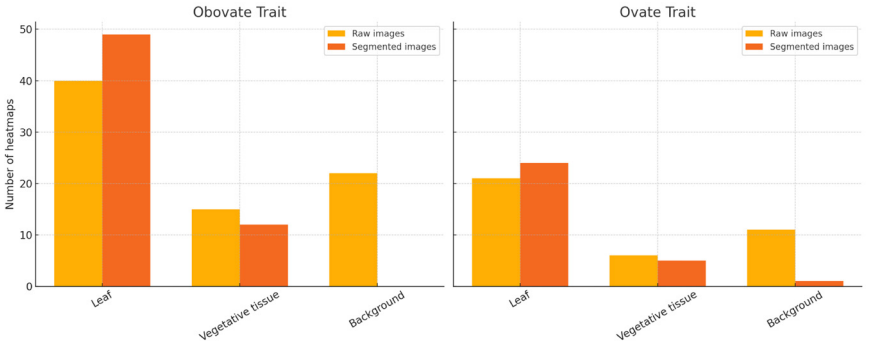


Fig. 6. Comparison of Grad-CAM focus areas for obovate (left) and ovate (right) leaf traits across raw and segmented images. Each bar represents the number of heatmaps in which the model's attention was primarily directed toward the leaf, other vegetative structures, or background elements. Segmentation notably increases model focus on leaf regions while substantially reducing attention to background noise, thereby improving the biological relevance and interpretability of trait predictions.

4. Conclusion

This study provides a comprehensive evaluation of three deep learning models—YOLOv8, ViT, and ResNet101—for automated botanical trait recognition in digitized herbarium specimens. The central focus is the

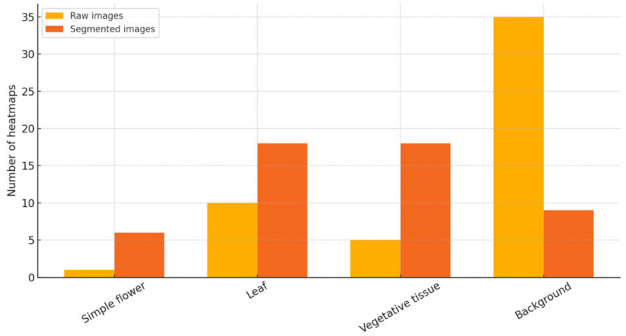


Fig. 7. Grad-CAM focus area comparison for the flower trait across raw and segmented images. The chart displays the number of heatmaps where the model's attention was concentrated on simple flowers, leaves, vegetative tissues, or background elements. Segmentation significantly reduced background attention while increasing focus on biologically relevant structures, notably floral and vegetative parts—highlighting its role in improving interpretability and scientific consistency of model predictions.

impact of image segmentation on classification accuracy, interpretability, and scientific reliability. Our results show that segmentation significantly improves performance by removing non-informative background elements and directing model attention to biologically relevant regions. However, segmentation may hinder performance for traits requiring broader context, such as reproductive organs (e.g., flowers, infructescences), due to the inadvertent loss of important visual details. This highlights the need for more precise and context-aware segmentation methods to support nuanced trait recognition. Among the tested architectures, YOLOv8 benefited most from segmentation, particularly for morphology-based traits. ResNet101 was more stable but less sensitive to refined inputs, while ViT delivered intermediate results.

Despite the dataset's modest size compared to major benchmarks, it offers unique, expert-validated annotations focused on visible morphological traits—making it particularly valuable. To our knowledge, no other publicly available dataset provides this level of visual trait detail in herbarium images. As such, our approach emphasizes not only accuracy but also explainability through model interpretation techniques like Grad-CAM. Models trained on segmented images consistently focused on plant structures, whereas those trained on raw images often relied on background artifacts like labels or barcodes. This reinforces our claim that segmentation enhances both the biological relevance and interpretability of predictions—key requirements for trustworthy AI in plant sciences.

Segmentation should thus be considered not just a preprocessing step but a core component of interpretable, reproducible research in biodiversity informatics. Its integration with attention-based explainability methods opens new avenues for hypothesis generation and large-scale trait discovery from herbarium data. Future work should prioritize the development of more accurate segmentation pipelines capable of preserving fine contextual and morphological information. The integration of models—such as Mask2Former [25]—represents a promising avenue for improving both accuracy and contextual awareness in trait-level recognition. In parallel, the adoption of semi-supervised annotation techniques could help reduce manual labeling efforts while enabling the expansion of trait-specific datasets across diverse botanical taxa. Expanding the diversity of training data—both taxonomically and morphologically—will be essential to ensure generalization across a wider range of plant lineages. As digital herbarium resources grow, the framework introduced here provides a solid foundation for scalable, interpretable, and biologically meaningful trait recognition in support of biodiversity research and conservation.

Acknowledgements

This work was partially funded by the French National Research Agency (Agence Nationale de la Recherche; ANR) in the context of the e-Col+ project (ANR-21-ESRE-0053). High-performance computing and storage resources were provided through the 2023-A0150114385 grant by Grand Equipement National de Calcul Intensif (GENCI) at the Institute for Development and Resources in Intensive Scientific Computing (IDRIS).

References

[1] Srivastava B, Reddy, P. RECENT ADVANCES IN BIODIVERSITY RESEARCH. 2023

[2] Wilde, Brendan C., Bragg, Jason G., Cornwell, William Analyzing trait-climate relationships within and among taxa using machine learning and herbarium specimens. American Journal of Botany, volume: 110, Number 5, 2023.

[3] Drew, Joshua and Moreau, Corrie and Stiassny, Melanie. Digitization of museum collections holds the potential to enhance researcher diversity. Nature Ecology and Evolution, 2017, doi: 10.1038/s41559-017-0401-6

[4] Youcef Sklab and Hanane Ariouat and Youssef Boudjydah and Yassine Qacami and Edi Prifti and Jean-Daniel Zucker and Régine Vignes Lebbe and Eric Chenin Towards a Deep Learning-Powered Herbarium Image Analysis Platform Biodiversity Information Science and Standards. 2024. https://doi.org/10.3897/biss.8.135629

[5] Weaver, William N, Ng, Julienne, Laport, Robert G. LeafMachine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens. Applications in Plant Sciences, volume: 8, number 6, url: https://doi.org/10.1002/aps3.11367.

[6] Weaver, William N. and Smith, Stephen A. From leaves to labels: Building modular machine learning networks for rapid herbarium specimen analysis with LeafMachine2. Applications in Plant Sciences,2022.

[7] Zhang, Wenli and Wang, Jiaqi and Liu, Yuxin and Chen, Kaizhen and Li, Huibin and Duan, Yulin and Wu, Wenbin and Shi, Yun and Guo, Wei. Deep-learning-based in-field citrus fruit detection and tracking. Horticulture Research, 2022.

[8] Ariouat H. Sklab Y., Pignal M, Jabbour F, Vignes Lebbe R, Prifti E, Zucker J-D, and Chenin E. Enhancing YOLOv7 for Plant Organs Detection Using Attention-Gate Mechanism. In Advances in Knowledge Discovery and Data Mining. PAKDD 2024. Lecture Notes in Computer Science, vol 14646. Springer, Singapore. https://doi.org/10.1007/978-981-97-2253-2_18

[9] Youcef Sklab and Hanane Ariouat and Edi Prifti and Eric Chenin and Jean-Daniel Zucker Identification of Non-Plant Elements in Herbarium Images Using YOLO In Proceedings of the Conférence Africaine sur la Recherche en Informatique et en Mathématiques. CARI 2024. https://doi.org/10.1007/978-3-031-88226-5_10

[10] Y. Jiang and C. Li. Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. Plant Phenomics, 2020.

[11] Yasamin Borhani and Javad Khoramdel and Esmaeil Najafi. A deep learning based approach for automated plant disease classification using vision transformer. Scientific Reports, 2022.

[12] Carranza-Rojas, J., Goeau, H., Bonnet, P. et al. Going deeper in the automated identification of Herbarium specimens. BMC Evol Biol, 2017.

[13] Abdelaziz Triki and Bassem Bouaziz and Jitendra Gaikwad b and Walid Mahdi Deep leaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images. Pattern Recognition Letters 2021.

[14] Hoehndorf R, Alshahrani M, Gkoutos GV, Gosline G, Groom Q, Hamann T, Kattge J, de Oliveira SM, Schmidt M, Sierra S, Smets E, Vos RA, Weiland C. The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. J Biomed Semantics, 2016.

[15] Younis, Sohaib and Weiland, Claus and Hoehndorf, Robert and Dressler, Stefan and Hickler, Thomas and Seeger, Bernhard and Schmidt, Marco. Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks Botany Letters, 2018, url: http://dx.doi.org/10.1080/23818107.2018.1446357.

[16] Lorieul, T., Pearson, K. D., Ellwood, E. R., Goeau, H., Molino, J.-F., Sweeney, P. W. Joly, A. (2019). a large-scale and deep phenological stage annotation of herbarium specimens : case studies from temperate, tropical, and equatorial floras. Applications in Plant Sciences, 7, e1233 [14 p.]. Retrieved from https://www.documentation.ird.fr/hor/fdi:010075504 doi: 10.1002/aps3.1233

[17] Roscher Ribana and Bohn Bastian and Duarte Marco F. and Garcke, Jochen Explainable Machine Learning for Scientific Insights and Discoveries IEEE Access 2020 doi=10.1109/ACCESS.2020.2976199

[18] Selvaraju Ramprasaath R. and Cogswell Michael and Das Abhishek and Vedantam Ramakrishna and Parikh Devi and Batra Dhruv. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization The vascular plants collection (P) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris) 2017 IEEE International Conference on Computer Vision (ICCV).

[19] Hanane, Ariouat and Youcef, Sklab and Marc, Pignal and Régine, Vignes Lebbe and Jean-Daniel, Zucker and Edi, Prifti and Eric, Chenin. Extracting Masks from Herbarium Specimen Images Based on Object Detection and Image Segmentation Techniques Biodiversity Information Science and Standards, 2023, url: https://doi.org/10.3897/biss.7.112161.

[20] Ariouat, H., Y. Sklab, E. Prifti, J.-D. Zucker, and E. Chenin. 2025. Enhancing plant morphological trait identification in herbarium collections through deep learning–based segmentation. *Applications in Plant Sciences* e70000. https://doi.org/10.1002/aps3.70000.

[21] MNHN, Chagnoux S. The vascular plants collection (P) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris) MNHN - Museum national d'Histoire naturelle., GBIF.org on 2024-03-20.

[22] A. Dosovitskiy and L. Beyer and A. Kolesnikov and D. Weissenborn and X. Zhai and T. Unterthiner and M. Dehghani and M. Minderer and G. Heigold and S. Gelly and J. Uszkoreit and N. Houlsby An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale Computer Vision and Pattern Recognition, 2021.

[23] K. He and X. Zhang and S. Ren and J. Sun Deep Residual Learning for Image Recognition Computer Vision and Pattern Recognition, 2015.

[24] Dillon R.,Jordan K., Jacqueline H., Ahmad D Real-Time Flying Object Detection with YOLOv8. Computer Vision and Pattern Recognition(2023).

[25] Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., & Girdhar, R. (2022). Masked-attention Mask Transformer for Universal Image Segmentation. CVPR 2022. arXiv:2112.01527

[26] Sklab, Y., Ariouat, H., Chenin, E., Prifti, E., & Zucker, J.-D. (2025). SIM-Net: A Multimodal Fusion Network Using Inferred 3D Object Shape Point Clouds from RGB Images for 2D Classification. *IET Computer Vision*, e70036. https://doi.org/10.1049/cvi2.70036