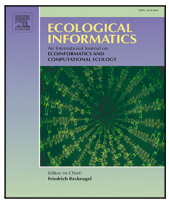
XAI tools to predict biological invasiveness from phenotypes: A case study in plants

Ecological Informatics 97 (2026) 103877

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



XAI tools to predict biological invasiveness from phenotypes: A case study in plants

Guido Spina [a,1], Barbara Frittella [a,1], Riccardo Ciarle [b], Simone Monaco [a,*]

[a] Politecnico di Torino, Department of Control and Computing Engineering, Torino, Italy
[b] Te Herenga Waka - Victoria University of Wellington, School of Biological Sciences, Wellington, New Zealand

ARTICLE INFO

Dataset link: https://github.com/inaturalist/inaturalist-open-data

Keywords:
Invasive species
Lythrum
Machine learning
Functional traits
Image-based models
Morphology

ABSTRACT

Determining which functional traits relate to invasion success is crucial for understanding how biological invasions develop and for predicting which species are likely to become invasive. Yet, studies relating morphological traits to invasiveness remain limited in scope and scale, and evidence for consistent morphological differences between invasive species and their non-invasive relatives is still sparse. Here, by focusing on the plant genus *Lythrum* as a case study, we develop a deep learning pipeline to distinguish invasive species from their non-invasive relatives by identifying discriminative morphological traits from images. We first developed an image-based classification model. We then applied an explainability pipeline to identify the image regions contributing to model predictions. Finally, we integrated prediction outcomes with region-level attributions to determine which morphological traits influenced invasiveness classification. Results showed that classification was driven primarily by morphological trait combinations characteristic of non-invasive species, rather than by the presence of invasive-specific traits. This supports the idea that invasiveness is not consistently associated with a fixed set of morphological traits, but rather reflects a relaxation or disruption of trait combinations characteristic of their non-invasive relatives. In general, this study showed that image-based models can recover biologically meaningful information from heterogeneous plant photographs and that their predictions are shaped by interactions among morphological traits rather than by single diagnostic features. By integrating an explainability pipeline, embedding analysis, and trait-pair attribution, these findings highlighted both the potential and limits of image-based approaches for invasion assessment, offering a scalable and interpretable method for linking deep learning to trait-based invasion ecology.

1. Introduction

Human-mediated dispersal of species has become a key feature of the Anthropocene (Essl et al., 2019) and around 4% of all extant vascular plants have become established outside their native range, with numbers still increasing worldwide (Seebens et al., 2018; van Kleunen et al., 2015). An increasing proportion of these species have become invasive (Shackleton et al., 2019), negatively impacting ecosystems (Seebens et al., 2017; Essl et al., 2020; Chai et al., 2025), economies (Haubrock et al., 2021; Holmes et al., 2009) and human lives (Hulme, 2015; Shackleton et al., 2019). Much research has focused on the processes and drivers shaping the geographic distributions and richness of invasive species (Catford et al., 2009; Dawson et al., 2017; Essl et al., 2019). In contrast, comparatively little attention has been given to whether invasive species are morphologically distinct from their non-invasive relatives and to what functional traits could explain and predict invasiveness.

Understanding how functional traits relate to invasion success is key for understanding how biological invasions develop and for predicting which species are likely to become invasive in the future (Palma et al., 2021). Several studies indicate that invasive and non-invasive species can differ predictably in their morphological traits (Palma et al., 2021; Pyšek and Richardson, 2007; Van Kleunen et al., 2010). For example, invasive species are typically characterized by higher seed production, germination rates, and overall reproductive success (Bora and Padial, 2023). Invasive species can also be associated with faster growth rates (Van Kleunen et al., 2010), lower plant height and smaller leaf area (Mathakutha et al., 2019), the ability of vegetative reproduction and short life cycles (Hussner et al., 2021). Despite this, generalized relationships between morphological traits and species invasiveness, as well as evidence for consistent morphological differences between invasive species and their non-invasive relatives, remain limited.

* Corresponding author.
E-mail address: simone.monaco@polito.it (S. Monaco).
[1] Authors contributed equally.

https://doi.org/10.1016/j.ecoinf.2026.103877
Received 24 February 2026; Received in revised form 8 June 2026; Accepted 9 June 2026
Available online 17 June 2026

Most studies relating morphological traits to invasiveness remain limited in scope and scale, relying on manually measured traits and on a limited number of species. These limitations can be effectively overcome by deep learning techniques that automatically extract and process vast, high-dimensional datasets from various sources and identify complex biological patterns. Employing deep learning with a systematic approach allows these studies to scale up the amount of data and species, enabling comparisons of multiple elements that would be slow and difficult, if not unfeasible, for human researchers. The vast number of images shared on platforms such as iNaturalist (iNaturalist, 2026) provides an unprecedented opportunity for large-scale biological analysis. Using computer vision techniques, deep learning models can employ scalable computational pipelines to extract biological knowledge from thousands of images and relate it to invasiveness, reducing the need for extensive in situ analyzes by human experts. This new paradigm has recently given rise to a growing research direction within the machine learning community known as *Imageomics*,[2] which seeks to make biological traits computable from visual data, thereby enabling new insights into the diversity, evolution, and functional biology of living organisms.

In parallel, recent work has increasingly focused on extracting interpretable relationships among traits derived from visual data, bridging computer vision and explainable machine learning (Granier and Vile, 2014). For example, association-rule-based approaches such as Fister et al. (2026) provide a complementary framework for discovering co-occurring patterns among numerical features and visualizing their interactions, aligning closely with the goal of identifying meaningful trait combinations. Besides these advancements, traditional machine learning approaches based on hand-crafted descriptors have also been proposed for morphological classification, such as SVMorph (Teng et al., 2021). However, these methods rely on predefined features that may not capture complex trait interactions. Similarly, studies in image-based phenotyping and trait inference have explored correlations among morphological characteristics using deep learning embeddings and post-hoc interpretability techniques, highlighting the growing importance of linking predictive performance with human-understandable trait relationships in ecological and biological applications (Carranza-Rojas et al., 2017). To the best of our knowledge, no work in the literature has explicitly bridged computer vision and post-hoc explainability analysis to extract morphological trait combinations from images to investigate and explain patterns of plant invasiveness.

In this study, we hypothesize that invasiveness is associated with identifiable combinations of morphological traits detectable from images, and that explainable ML can recover these trait combinations. To test this, we focus on the plant genus *Lythrum* as a case study and employ a deep learning framework to distinguish invasive species from their non-invasive relatives by identifying discriminative morphological traits from images. We devised a pipeline articulated in three main steps. First (I), we developed an image-based classification model to predict species invasiveness; then (II), we applied an explainability pipeline to identify the image regions contributing most to model predictions; finally (III), we integrated prediction outcomes with region-level attributions to determine which morphological traits, alone or in combination, influenced invasiveness classification within *Lythrum*. An overview of the full pipeline is shown in Fig. 1.

2. Methods

2.1. Data collection

We used the plant genus *Lythrum* (Lythraceae) as a case study. This taxon was chosen because it is relatively small, with 40 accepted species (*Lythrum* L. | Plants of the World Online | Kew Science, n.d.). More importantly, the taxon includes several invasive species, varying in their degree of invasiveness. *Lythrum salicaria* is a highly invasive herbaceous perennial plant native to Europe, Asia, northern Africa, and eastern Australia, which has been included in the IUCN list of the 100 World's Worst Invasive Alien Species. *Lythrum virgatum* and *Lythrum hyssopifolia* are invasive herbaceous perennial plants listed as species of concern in several regions outside their native range (Mattingly et al., 2023; New Zealand Plant Conservation Network, 2026). The genus is also well represented on iNaturalist (iNaturalist, 2026). This allowed us to investigate similarities among invasive *Lythrum* species and differences between invasive species and their non-invasive relatives. For each species, we gathered image data from iNaturalist. Species with no records at the time of download (June 20th, 2025) were removed from the analyses. After download, each image was associated with a label indicating whether the represented species was invasive or not.

2.2. Model classification and evaluation

The curated dataset of images and corresponding labels naturally formulates a binary image classification problem, which can be effectively addressed using deep learning-based computer vision models. A common and well-established strategy consists of leveraging pretrained convolutional or vision transformer architectures (e.g., ResNet, He et al. (2016), Vision Transformers, Dosovitskiy (2020)), which act as feature encoders, followed by task-specific fine-tuning of a lightweight classification head on the learned embeddings. This transfer learning paradigm is known to accelerate convergence, improve generalization, and reduce the amount of labeled data required for downstream tasks (Yosinski et al., 2014; Tan et al., 2018).

Rather than relying on a general-purpose visual representation model, we adopt BioCLIP-2 (Gu et al., 2025), a large-scale foundation model trained using hierarchical contrastive learning to align images with taxonomic and ecological labels. The authors demonstrate that this training strategy induces strong inter-species ecological separation while preserving meaningful intra-species variation, resulting in representations that are particularly well-suited for biological visual tasks. Consequently, BioCLIP-2 provides more informative, biologically grounded image embeddings in zero-shot and few-shot settings than general-purpose vision models trained on natural image datasets. Moreover, BioCLIP-2 provides general-purpose biological image representations trained on a large and taxonomically diverse organismal dataset. Other relevant alternatives include PlantCLEF-derived models (Goëau et al., 2023; Goeau et al., 2025) for multilabel plant recognition, or plant-specific models optimized for crop, leaf, or disease-recognition tasks (Nahian et al., 2025; Han et al., 2026). We therefore do not claim that BioCLIP-2 is optimal for invasion analysis, but use it as a biologically informed backbone suitable for extracting transferable visual representations.

Finally, we built a classifier composed of two fully connected layers with a hidden dimension of 512 and ReLU activation function on the extracted embeddings to predict whether an image represented an invasive or non-invasive species, optimizing the model with a cross-entropy loss function using the Adam optimizer with a learning rate of 0.001, a batch size of 16, and trained for 15 epochs. With our computational resources, the full parallel training of the models in a run of the cross-validation took around 85 min in training time and 7 min of inference time (14 $\pm$ 8 s for each model). A summary of the main architectural choices and training hyperparameters is reported in Table 1.

The classification model was evaluated using a Leave-One-Species-Out (LOSO) cross-validation scheme (Fig. 2). Namely, we trained one model for each species. For each iteration, all images belonging to a single species were held out as the test set, while images from all remaining species were used for training, following an 80/20 split between training and validation without. This procedure was repeated for each species in the dataset. LOSO cross-validation prevents data leakage among images of the same species and assesses the model's ability to learn visual features associated with invasiveness rather than memorizing species identity.

[2] https://imageomics.osu.edu/.

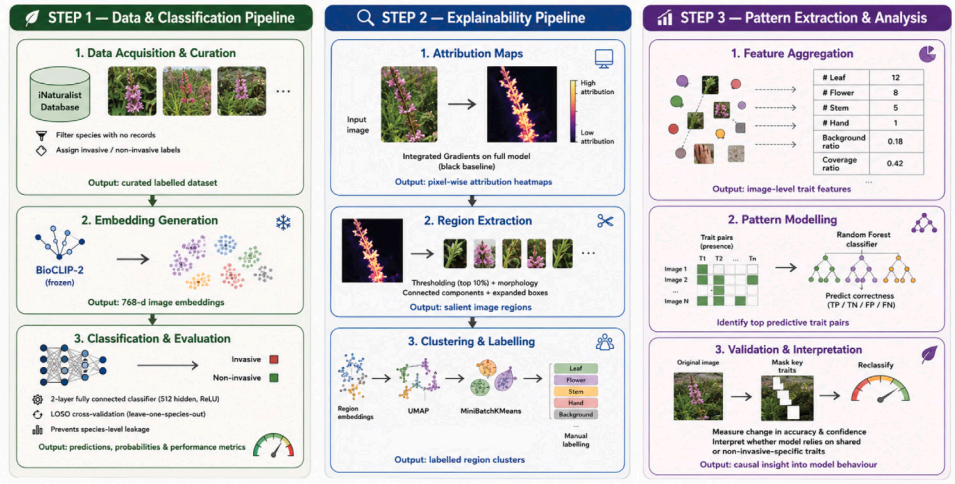


Fig. 1. Pipeline overview. The first step involves data acquisition and subsequent ingestion into the classification model via subsets defined by Leave-One-Species-Out (LOSO) cross-validation scheme. The second step introduces the explainability process to extract the most relevant regions and subsequently label them with the corresponding plant structure. Finally, the last step investigates trait attribution by combining invasiveness and trait occurrences using pattern-extraction algorithms.

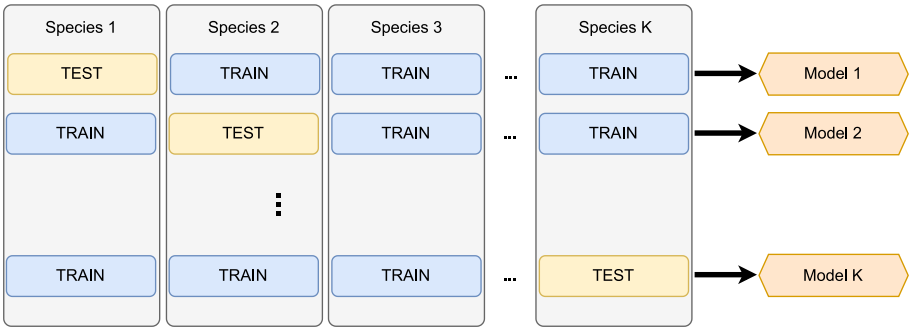


Fig. 2. LOSO Cross-validation scheme.

Table 1
Summary of classifier architecture and training hyperparameters.

Component	Description
Embedding model architecture	BioCLIP-2
Pretraining	TreeOfLife-200M (Gu et al., 2025)
Embedding dimension	768
Classifier type	Fully connected layers
Number of layers	2
Hidden dimension	512
Activation function	ReLU
Output dimension	2 (invasive vs non-invasive)
Loss function	Cross-entropy
Optimizer	Adam
Learning rate	0.001
Batch size	16
Number of epochs	15
Weight decay	0.0
Input image size	$\leq 500 \times 500$
Train/validation split	80/20 (and LOSO schema for test)
Hardware	GPU NVIDIA L4

During model evaluation, one invasive species, *Lythrum hyssopifolia*, consistently showed near-zero classification performance. To investigate this behavior, we applied Uniform Manifold Approximation and Projection (UMAP) to the BioCLIP-2 image embeddings, reducing them to two dimensions for visualization. We then computed pairwise interspecies distances as the average Euclidean distance between all points representing each species pair in the reduced feature space, excluding under-represented species for statistical robustness. In this representation, *L. salicaria* and *L. virgatum* clustered closely, whereas *L. hyssopifolia* was positioned closer to non-invasive species, suggesting limited visual separability from that group. Because the objective of this study was to test whether invasiveness within *Lythrum* is associated with identifiable image-derived morphological patterns, we decided to exclude *L. hyssopifolia* from subsequent analyses to keep a subset of invasive species for which a coherent morphology-based signal was detectable, allowing us to investigate the specific trait patterns underlying successful classification without conflating fundamentally distinct morphological regimes.

2.3. Explainability pipeline

To interpret the model's predictions, we implemented an explainability pipeline designed to assess whether invasiveness classifications relied on biologically meaningful plant traits rather than spurious visual cues (e.g., background or non-plant elements). The pipeline comprised four steps: (i) attribution map generation, (ii) extraction of salient image regions, (iii) embedding and clustering of regions, and (iv) manual cluster annotation.

Pixel-wise attribution maps were generated using Integrated Gradients (Sundararajan et al., 2017), a gradient-based method that attributes predictions to input features by integrating gradients from a baseline image to the original input. A finer analysis on the ratio behind this choice is provided in Appendix A.4. Attributions were computed for the complete classification architecture, including the frozen BioCLIP-2 image encoder and the classifier head. A black image was used as

the baseline, and attributions were aggregated across RGB channels to obtain a single normalized heatmap per image.

Salient image regions were extracted from the attribution maps using a standardized procedure. Attribution maps were resized to the input image resolution, normalized, and thresholded by retaining pixels above the 90th percentile of the attribution values. Morphological opening and closing were applied to reduce noise and enforce spatial coherence, identifying connected components while removing small, noisy regions. Bounding boxes were then expanded to capture the full visual structure of each identified pixel cluster. Extracted bounding boxes were then collected for each original image and stored as individual regions.

Then, we passed the extracted regions to BioCLIP-2 to get semantically meaningful embeddings. Embeddings were projected to a lower-dimensional space using UMAP to preserve local neighborhood structure (McInnes et al., 2018) and subsequently clustered. Due to the large amount of data, we adopted MiniBatchKMeans, a scalable variant of KMeans algorithm that updates the centroids using small random batches of data rather than the entire dataset. This approach greatly reduces computational cost while maintaining comparable clustering quality. with KMeans into a fixed number of visually coherent groups. Clustering was performed on a representative subset of 2000 images, and the resulting centroids were used to assign regions to the remaining dataset samples. We repeated the experiment with a variable number of clusters and selected the optimal value using the silhouette score. Resulting labels were finally obtained via majority vote of five repetitions of the clustering algorithm on different initialization points.

Clusters were manually annotated through visual inspection of the regions assigned to them. Each cluster received one or more labels describing the dominant visual content. We organized the labels into two main categories:

- *Plant structures*:
  - **Leaf**: clusters primarily showing leaves;
  - **Flower**: clusters characterized by the presence of flowers;
  - **Stem**: clusters clearly depicting stems.
- *Spurious or non-informative features*:
  - **Hand**: clusters containing primarily parts of human hands or skin;
  - **Background/undefined**: clusters dominated by background fragments or visually ambiguous regions.

We focused on these plant structures because they were present in most images and could be reliably distinguished by the model. These structures are also ecologically relevant, as they are known to vary among species and have been associated with differences between invasive and non-invasive taxa (Palma et al., 2021; Van Kleunen et al., 2010).

Labels within the same category were not mutually exclusive (e.g., a cluster may simultaneously be labeled *Leaf* and *Flower*), whereas labels from different categories were mutually exclusive. For instance, a cluster cannot be labeled both *Leaf* and *Hand*. This annotation enabled structured analysis of whether model attention was concentrated on biologically relevant traits.

Using the clustering model trained on the representative subset of images, we applied the same embedding and clustering procedure to the full dataset. Each region was assigned to its nearest cluster centroid and inherited the corresponding annotation, resulting in a set of labeled regions for each image. Region-level information was then aggregated to the image level to support higher-level analyses. Cluster labels were further refined to enable more detailed biological interpretation by incorporating species-specific morphological traits (Appendix B) in addition to the general categories Leaf, Flower, and Stem.

2.4. Traits attribution

Building on the explainability pipeline, we analyzed model predictions together with region-level attributions to identify which morphological traits, alone or in combination, influenced invasiveness classification. As an initial exploratory step, we performed analyses based on single traits and image-level summary metrics derived from the explainability pipeline, including region counts, coverage ratios, and proportions of non-informative regions. These analyses aimed to identify simple associations between individual traits or summary metrics and prediction outcomes. As no strong or consistent patterns emerged, these results are not discussed further in the main text. Methods and results are reported in Appendix C.

Given the lack of signal in single-trait analyses, we next investigated whether model predictions depended on trait combinations. We conducted a pairwise characteristic trait analysis by generating all possible pairs of cluster-level traits and grouping them according to their taxonomic specificity: traits present only in non-invasive species, traits present only in invasive species, and shared traits. Each image was labeled by prediction correctness (i.e. correct or incorrect), with additional analyses distinguishing true positives, true negatives, false positives, and false negatives. Using the presence of trait pairs as input features, Random Forest classifiers were trained to predict prediction correctness under these different settings. The five most relevant trait pairs were selected for further analysis.

For each selected trait pair, all images containing both traits were identified, and the corresponding regions were localized in the original images. These regions were then masked (replaced with blank regions), and the modified images were reclassified to quantify changes in prediction accuracy and confidence after removing the most predictive traits.

Finally, by comparing model behavior for trait pairs classified as Shared versus Non-invasive only (no Invasive-only pairs were present in the dataset), we assessed whether predictions relied primarily on general morphological patterns shared across taxa or on traits distinctive of non-invasive species.

3. Results

3.1. Data collection and preprocessing

The final dataset comprised 27 non-invasive species and 3 invasive ones. The total number of images available per species and their distributions worldwide is shown in Fig. 3.

Images were collected from iNaturalist (iNaturalist, 2026), without filtering them by quality grade to avoid excluding underrepresented species. Approximately 75% of the collected images were at *research* level. Images were retained regardless of the accuracy of their acquisition dates or geographic coordinates, as these parameters did not impact our study.

Prior to model training, all acquired images were resized (if necessary) to a maximum dimension of 500 pixels and then normalized per channel. No other additional preprocessing or data augmentation was applied. Although the dataset included few invasive species, it was fairly balanced in the number of samples per class: of the 45 361 total images, 20 528 belonged to non-invasive species (45.25%) and 24 833 to invasive ones (54.75%).

3.2. Classification and evaluation

Fig. 4 reports classification accuracy for each target species, ordered by F1 score. To quantify the variability in model performance, the figure shows mean values with 95% confidence intervals across LOSO evaluation repetitions. Note how *L. hyssopifolia*, one of the most represented species in the dataset, shows very low accuracy and F1 score. Conversely, the other two invasive species exceed the random-choice

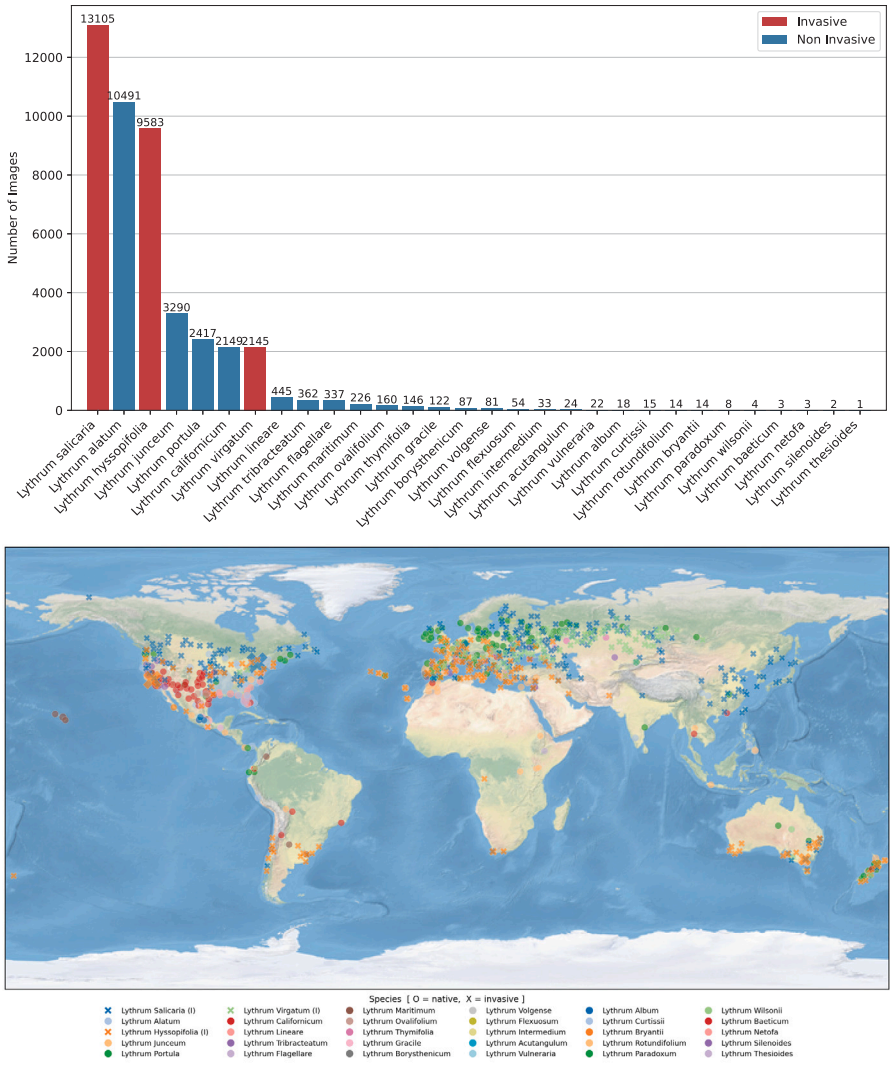


Fig. 3. Total number of images retrieved per species (top) and their global spatial distribution (bottom). Invasive species are shown in red in the bar chart and marked with '×' on the map, while non-invasive species are shown in blue in the bar chart and marked with '○' on the map.

baseline in both accuracy and F1 score, although these remain lower than those of non-invasive species. To assess the potential impact of species-level imbalance, we performed an additional sensitivity analysis excluding all low-represented species (Appendix A.2).

UMAP results showed a global mean distance of 6.584 ± 3.073 units across all species pairs. *L salicaria* and *L. virgatum* clustered closely together, while *L. hyssopifolia* positioned closer to non-invasive species. Specifically, for *Lythrum salicaria*, the nearest neighbor in feature space was *Lythrum intermedium* at a distance of 2.867 units, followed by *Lythrum virgatum* at 4.203 units. This proximity in the embedding space suggests shared morphological characteristics between the two invasive species. Conversely, *L. hyssopifolia* ranked as the fourth-most distant species from *L. salicaria*, with a separation of 11.608 units. Analysis of distances from *L. hyssopifolia* revealed an inverted pattern: the five closest species were all non-invasive, with distances ranging from 2.480 to 4.045 units, while *L. virgatum* and *L. salicaria* appeared instead among the most distant species at 10.825 and 11.608 units, respectively. The furthest species from *L. hyssopifolia* was *L. intermedium* at 13.321 units.

This indicates that *L. hyssopifolia* occupies a well-separated region of the inferred morphological space, explaining the classification failures observed during the Leave-One-Species-Out cross-validation procedure, in which the model achieved near-zero accuracy (0.0057) when predicting its invasiveness. Consistently, removing *L. hyssopifolia* from the model led to a substantial improvement in both accuracy and F1 scores across the remaining species, as reported in Fig. 5. Based on this evidence, *L. hyssopifolia* was excluded from all subsequent analyses. The UMAP representation of the species and the parameters used in the analysis are reported in Appendix A.3.

To provide a more complete evaluation of model performance, we report global accuracy, precision, recall, and F1-score for both experimental settings, including and excluding *L. hyssopifolia*. As shown in Fig. 6, excluding *L. hyssopifolia* improved all metrics, with accuracy increasing from 0.52 to 0.80, precision from 0.48 to 0.77, recall from 0.74 to 0.91, F1-score from 0.58 to 0.83, and AUC from 0.54 to 0.77. The improvement in recall indicates that the model produced fewer false negatives, which is particularly relevant in invasion-risk assessment because failing to identify potentially invasive species may have important ecological consequences. At the same time, the significant increase in precision suggests a reduction in false positives, limiting the risk of incorrectly classifying non-invasive species as invasive. Overall, these results show that the exclusion of *L. hyssopifolia* leads to a more balanced and reliable classification performance.

3.3. Explainability and morphological trait clustering

In most cases, attribution heatmaps generated via Integrated Gradients highlighted regions corresponding to biologically meaningful plant

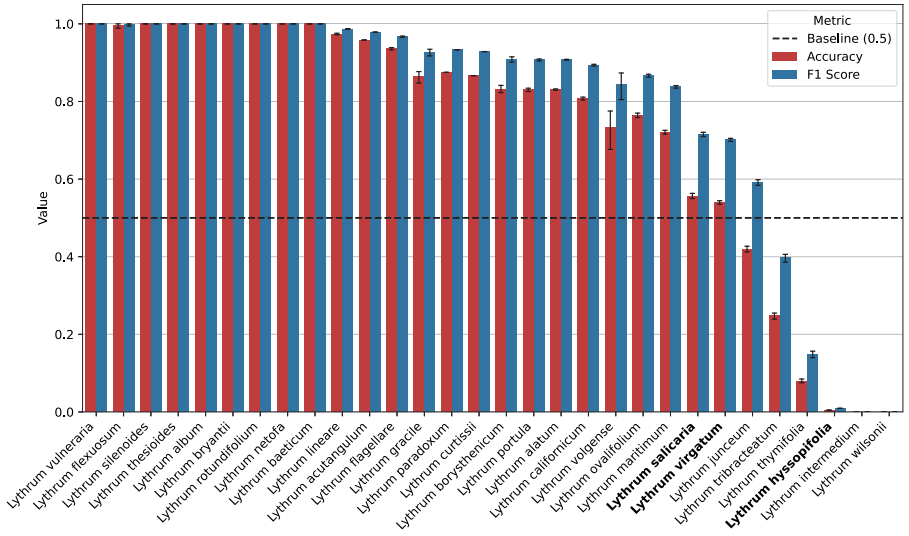


Fig. 4. Model accuracy and F1 score for the Lythrum genus in the LOSO scheme. Invasive species are highlighted in bold. Results are reported by mean values with 95% confidence intervals.

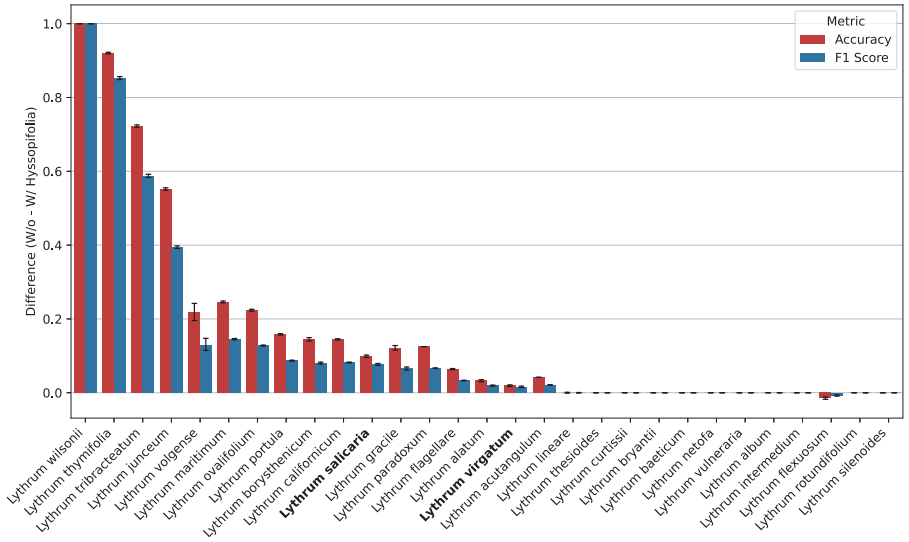


Fig. 5. Difference in performance for accuracy (red) and F1 score (blue) after removing L. hyssopifolia from the training data. Invasive species are highlighted in bold. Results are reported by mean values with 95% confidence intervals.

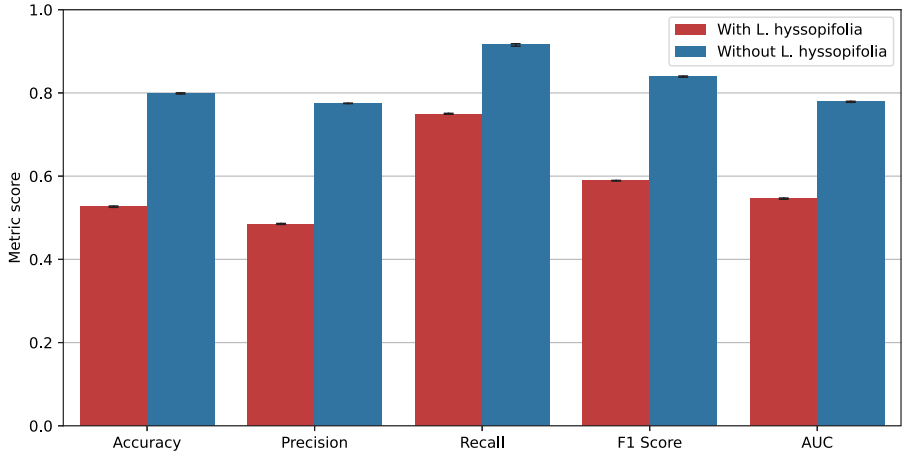


Fig. 6. Aggregated performance of experiments including and excluding *L. hyssopifolia*. Results are reported in terms of mean and 95% confidence intervals over 5 repetitions of the LOSO scheme.

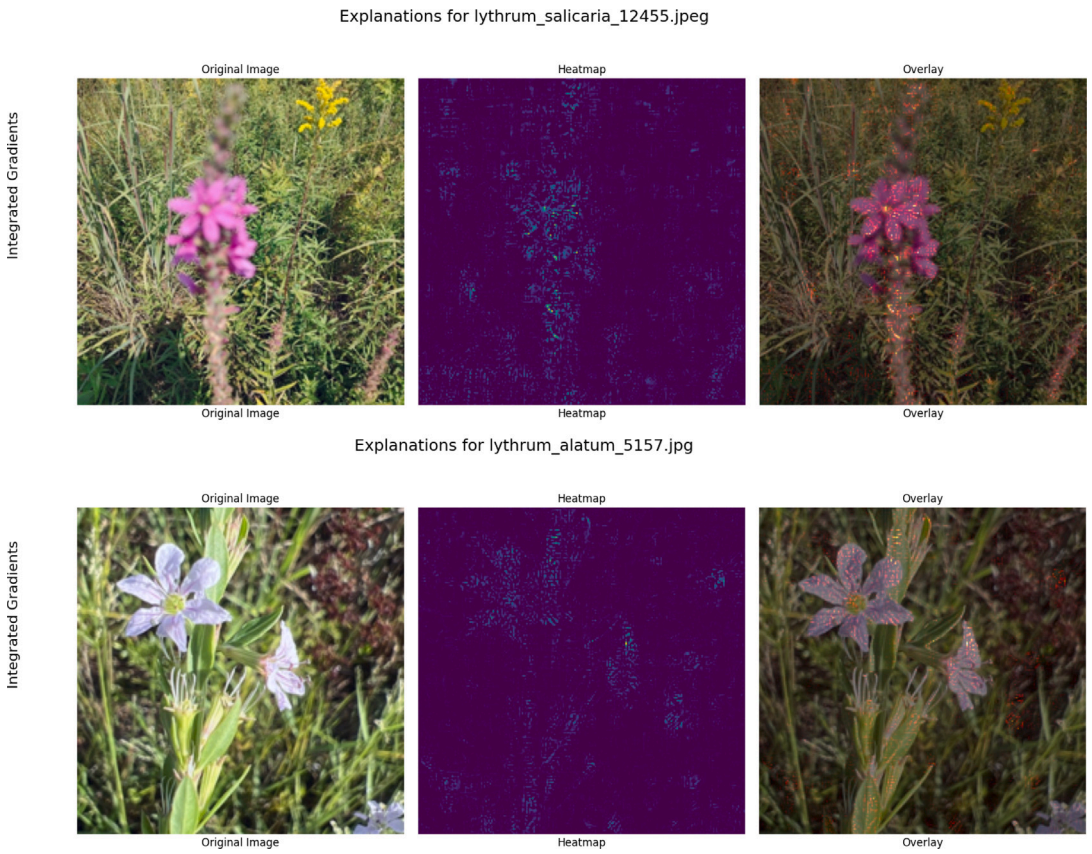


Fig. 7. Examples of attribution heatmaps generated using Integrated Gradients. For each sample (a, b), the original image, generated heatmap, and overlay are shown from left to right. In most cases, highlighted regions correspond to biologically meaningful structures such as flowers, leaves, or stems.

structures, as shown in Fig. 7. However, certain samples exhibited attributions focused on non-biological features, including portions of human hands or diffuse patterns lacking clear correspondence to any recognizable feature.

The extraction procedure yielded 132 128 regions extracted from the full dataset of 35 678 images. For the clustering stage, the optimal number of clusters was set to 30. Cluster size in the preliminary subset ranged from 157 to 329, with a mean of 250.3 ± 48.7 regions per cluster. The resulting structural labels distribution across all clusters, and therefore all extracted regions, is presented in Fig. A.17.

The majority of extracted regions corresponded to biologically relevant structures. Across the 2000-image subset, 5324 regions (70.9%) were assigned to clusters labeled with at least one plant structure, with *Leaf* being the most frequent label (3876 regions, 51.6%). Non-biological features accounted for the remaining regions, with 1589 (21.2%) being classified as *Background/undefined*, and 596 (7.9%) as human hands. Most clusters were associated with a single dominant label, while a smaller subset exhibited mixed annotations reflecting regions containing multiple structures. A further validation on the clustering procedure is provided in Appendix A.6.

3.4. Pairwise trait analysis and occlusion-based validation

To identify which combinations of morphological traits strongly influenced model predictions, a Random Forest classifier was trained using pairwise trait co-occurrences as input features. This analysis differs from standard feature importance evaluation by explicitly encoding trait interactions rather than treating each trait independently. Feature importance scores were calculated both globally and separately for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) prediction categories.

The 15 most important trait pairs are presented in Table 2. To provide a quantitative scale for interpreting the reported values, we compared the observed feature importances with a uniform-importance baseline. Since the analysis considered 64 unique traits (as in Appendix B), leading to $\tilde{2}000$ possible trait pairs. Under a uniform and non-informative distribution of importance across trait pairs, each pair would be expected to contribute approximately $1/2016 = 4.96 \times 10^{-4}$ of the total normalized importance. The top-ranked pair, *Erect–Sessile*, was more than 600 times higher than the uniform expectation, and even the 15th-ranked pair remained above this baseline. Together, the top 15 pairs accounted for approximately 99.7% of the total normalized importance, indicating that the attribution signal was concentrated in a small set of recurrent trait combinations.

Based on these results, five trait pairs were selected for further investigation via occlusion experiments: *Erect-Sessile*, *Alternate–Subsessile*, *Opposite–Sessile*, *Erect–Opposite*, and *Linear–Opposite*. These pairs were chosen for their high importance in TP and TN categories and their biological interpretability. Trait pairs were classified into two categories: *Common* (traits shared by both invasive and non-invasive species) and *Non-Invasive only* (traits exclusive to non-invasive species). No *Invasive only* pairs were identified in the dataset. The selected pairs and their classifications are summarized in Table 3.

Fig. 8 and Table 4 compare the prediction performance of models using unaltered images with the performance of models using images where regions containing at least one of the selected traits were masked. For trait pairs in the *Common* category, masking produced only minor effects: accuracy decreased by less than 2%, and label flip rates, i.e. the proportion of samples whose predicted label changes after masking, remained near 17%. Changes in True Positive and True Negative counts were minimal, indicating that these trait combinations, while frequently observed, are not essential for correct classification. In contrast, masking trait pairs from the *Non-Invasive only* category

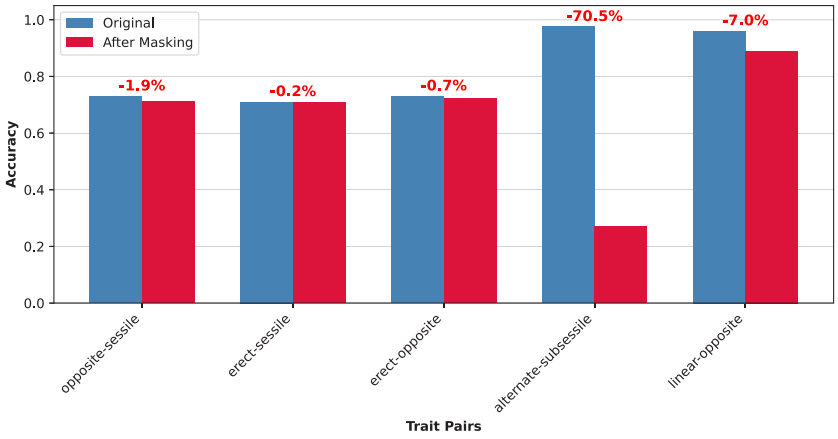


Fig. 8. Accuracy comparison before (blue) and after (magenta) masking for each trait pair. Red values above bars indicate accuracy change (Δ). Trait pairs from the *Non-Invasive only* category exhibit substantial accuracy decreases, while *Common* pairs show minimal change.

Table 2
Top 15 trait pairs ranked by Random Forest feature importance. Importance values are reported globally and separately for each prediction category (TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative). Since 2016 possible unordered trait pairs were considered, the expected importance under a uniform non-informative distribution is 4.96×10^{-4}. Values above this baseline indicate trait pairs that contribute disproportionately to the prediction outcomes.

Feature A	Feature B	Importance				
		Global	TP	TN	FP	FN
Erect	Sessile	0.3112	0.3051	0.3005	0.3055	0.3091
Opposite	Sessile	0.1811	0.1755	0.1823	0.1326	0.1908
Alternate	Subsessile	0.1777	0.1945	0.1860	0.1026	0.1846
Erect	Opposite	0.1675	0.1508	0.1668	0.2016	0.1522
Linear	Opposite	0.0574	0.0733	0.0610	0.0218	0.0672
Erect	Linear	0.0521	0.0532	0.0554	0.0185	0.0518
Opposite	Petiolated	0.0267	0.0217	0.0235	0.0716	0.0202
Opposite	Subsessile	0.0065	0.0065	0.0064	0.0095	0.0063
Opposite	Prostrate	0.0041	0.0044	0.0041	0.0040	0.0038
Obovate	Sessile	0.0027	0.0026	0.0026	0.0038	0.0023
Prostrate	Subsessile	0.0025	0.0024	0.0025	0.0034	0.0025
Alternate	Sessile	0.0024	0.0024	0.0017	0.0769	0.0022
Alternate	Erect	0.0019	0.0016	0.0017	0.0144	0.0014
Erect	Obovate	0.0018	0.0031	0.0020	0.0026	0.0029
Alternate	Creeping	0.0017	0.0009	0.0012	0.0126	0.0009

Table 3
Selected trait pairs for occlusion analysis. For each trait, the associated plant structure and category (Common or Non-Invasive only) are indicated.

Feature A		Feature B		Category
Name	Plant Structure	Name	Plant Structure	
Erect	Stem	Sessile	Leaf	Common
Alternate	Leaf	Subsessile	Leaf	Non-Invasive only
Opposite	Leaf	Sessile	Leaf	Common
Erect	Stem	Opposite	Leaf	Common
Linear	Leaf	Opposite	Leaf	Non-Invasive only

resulted in substantial performance degradation. Removing regions associated with *Alternate–Subsessile* caused accuracy to drop from 0.973 to 0.268, with 71.3% of images experiencing label flips. The mean change in predicted probability for the true class was −0.694, indicating a strong shift toward invasive predictions when this trait pair was absent. The pair *Linear–Opposite* exhibited a similar but weaker trend, with accuracy decreasing by 7% and a mean probability shift of −0.046.

To assess whether the masking effects were simply driven by trait frequency, we considered the representativeness of the highest-attribution trait pairs directly in the text. Among the five pairs reported in Fig. 8, *Alternate–Subsessile* was the least frequent, occurring in approximately 30% of images when considering the presence of at least one of the two traits. By contrast, the other highlighted pairs occurred in approximately 70%–90% of images. Therefore, the relatively limited representation of the *Alternate–Subsessile* pair, combined with its large impact on classification, likely represents a real biological signal.

Table 4
Performance comparison after region masking for selected trait pairs. Flip Rate reports the proportion of samples whose predicted label changes after masking, while the TP Ratio and TN Ratio quantify the fraction of true invasive and true non-invasive samples correctly identified after masking. The rightmost columns report changes in model confidence. For samples that are truly invasive or truly non-invasive, we show the proportion of samples in each group and the mean change in predicted probability for the invasive class ($(\Delta P = P_{\text{after}} - P_{\text{before}})$). Positive ($\Delta P$) indicates increased confidence in invasive predictions following masking, while negative values indicate decreased confidence. Non-invasive-only (NI) trait pairs contain no invasive samples and therefore omit invasive-class statistics. (NI): Non-Invasive only category; dashes indicate no invasive examples.

Trait pair	Flip rate	True invasive		True non-invasive	
		n (%)	Mean ΔP	n (%)	Mean ΔP
Erect–Sessile	17.3%	57.2%	+0.023	42.8%	−0.021
Alternate–Subsessile (NI)	71.3%	–	–	100.0%	+0.694
Opposite–Sessile	17.6%	52.1%	+0.022	47.9%	+0.015
Erect–Opposite	16.7%	52.7%	+0.023	47.3%	−0.010
Linear–Opposite (NI)	10.1%	–	–	100.0%	+0.046

The pronounced effect of masking *Non-Invasive only* trait pairs indicates that the classifier relies heavily on these visual combinations as diagnostic markers for non-invasive species. Their removal systematically biases predictions toward the invasive class, confirming their role as class-specific discriminative features. In contrast, *Common* trait pairs encode redundant morphological information that does not uniquely distinguish between classes, resulting in minimal impact on prediction outcomes when removed.

4. Discussion

Our results demonstrate that species invasiveness can be recognized through morphological traits retrieved from images using a deep learning embedding model such as BioCLIP-2, although this may not be possible for all species as shown by *L. hyssopifolia*. Importantly, the LOSO evaluation protocol ensures that each test species is entirely unseen during training. This design prevents the model from relying on species identity cues and instead requires it to learn morphological patterns that generalize beyond the training taxa. Therefore, the observed performance under this stringent setting supports the hypothesis that

invasiveness can be inferred from external morphological traits, thereby providing a foundation for subsequent deep learning analyses.

The fact that *L. hyssopifolia* was consistently misclassified despite good representation, combined with UMAP results showing the species embedded among non-invasive species, suggests that the model behaved coherently and classified species based on true morphological similarity rather than species labels. *L. hyssopifolia*'s visual characteristics overlapped more with those of non-invasive taxa rather than with those of the other 2 invasive species. In this context, poor classification performance reflects a biologically meaningful mismatch between invasiveness labels and morphology, rather than a failure of the classification model.

From a data science perspective, we interpreted *L. hyssopifolia*'s behavior as a biologically informative exception in which invasiveness was not recoverable from the visual trait space captured by the model. From an ecological standpoint, this discrepancy suggests that its invasiveness may not be driven by clearly distinguishable morphological traits, or at least not by traits that are reliably captured in image data. Unlike *L. virgatum* and *L. salicaria*, *L. hyssopifolia* may rely on non-morphological mechanisms such as phenological timing, plasticity, reproductive ecology or dispersal ability to achieve invasion success. This interpretation aligns with broader findings in invasion ecology showing that some invasive species succeed without exhibiting clear morphological differentiation from non-invasive relatives (García et al., 2013; Divíšek et al., 2018). This shows how image-based models can successfully identify invasiveness when this is associated with consistent morphological patterns, but they may fail when invasiveness arises from non-morphological mechanisms, or when invasiveness is not visually encoded. But even in these cases, they can provide valuable ecological insight, highlighting species whose invasiveness is context-driven rather than trait-driven.

The trait-pair analysis did not find a unique visual signature for invasive species, but rather trait combinations that reliably characterize non-invasive ones, namely alternate–subsessile leaves and linear–opposite leaves, indicating that invasiveness classification in *Lythrum* is driven less by the presence of distinctive invasive traits than by the absence of morphological combinations characteristic of non-invasive species. Occlusion results showed how trait pairs exclusive to non-invasive taxa played a disproportionate role in correct classification, as their removal caused substantial drops in accuracy and frequent label reversals toward the invasive class. In other words, the model classified invasiveness through negative evidence, i.e., lack of trait pairs. This suggests that image-based models can provide useful information for invasiveness assessment by identifying stabilizing trait configurations typical of non-invasive species, rather than relying on a universal invasive phenotype. Such an approach may be particularly valuable in groups where invasive and non-invasive species are morphologically similar (García et al., 2013), and where traditional single-trait diagnostics have limited predictive power.

From an ecological perspective, the non-invasive trait combinations reflect distinct aspects of plant architecture and resource-use strategy. Alternate phyllotaxis combined with subsessile leaves is consistent with moderately flexible shoot organization and intermediate structural investment (Nicotra et al., 2011; Reinhardt and Gola, 2022), while the association between linear leaves and opposite phyllotaxis is indicative of reduced leaf surface area and more conservative, structurally regular growth forms often associated with stress tolerance or resource limitation (Wright et al., 2004; Nicotra et al., 2011). Importantly, these trait combinations should not be interpreted as diagnostic traits of non-invasive species per se, but rather as a constrained morphological syndrome shared by a subset of non-invasive species in the dataset. The absence of invasive-specific trait pairs, together with the strong effect of masking non-invasive combinations, suggests that model predictions are driven by the presence of these constrained morphological patterns rather than by unique features of invasive taxa. This aligns with ecological theory emphasizing that invasiveness is rarely associated with unique morphological traits, but instead reflects broader functional strategies and trait trade-offs (Van Kleunen et al., 2010; Díaz et al., 2016).

More broadly, these results support the idea that invasiveness is not consistently associated with a fixed set of morphological traits, but instead reflects a relaxation or disruption of trait combinations that characterize non-invasive relatives. The absence of invasive-only trait pairs, coupled with the strong influence of non-invasive trait combinations, suggests that invasiveness, at least in *Lythrum*, may be better understood as a loss of morphological constraint rather than the emergence of a distinct invasive syndrome. This interpretation aligns with long-standing challenges in invasion biology in identifying universal trait predictors and highlights the value of integrative approaches based on trait-combination (Van Kleunen et al., 2010; Gallagher et al., 2015). By linking model behavior to biologically interpretable trait interactions, our results demonstrate how explainable machine-learning frameworks can contribute to ecological theory as well as practical risk assessment. The failure of single-trait analysis was expected, as most traits considered in this study are individually common, weakly diagnostic on their own, and highly context dependent. In contrast, trait combinations successfully captured morphological syndromes, identifying ecological and morpholofical constraints in *Lythrum*. That said, our results should be interpreted cautiously in terms of ecological causation. The identified trait combinations do not demonstrate that particular morphological syndromes directly cause invasiveness. Rather, the model captures statistical regularities within the image-derived morphological space. In particular, the strong influence of non-invasive-associated trait pairs may indicate that invasive species occupy a broader or less constrained region of morphospace, rather than exhibiting unique invasive-specific morphologies.

In the UMAP embedding space, *Lythrum intermedium* exhibited an average distance of 11.332 ± 2.934 units from all other species. Notably, its closest neighbors were the invasive species *L. salicaria* (2.867 units) and *L. virgatum* (3.878 units), while its five most distant neighbors ranged from 13.155 to 13.870 units, and included *L. hyssopifolia* (13.321 units). The close association of *Lythrum intermedium* with invasive species likely reflects its taxonomic status, as it is frequently treated in botanical databases and the published literature as a subspecies or synonym of *L. salicaria* rather than as a distinct species (Plants of the World Online, 2026). Rather than reflecting misclassification, this result further suggests that the model captures meaningful morphological similarities among species, supporting its ability to learn biologically relevant patterns beyond the imposed class labels.

The substantial variation in species-level F1 scores may indicate that classification difficulty was not determined solely by representation within the dataset. Several non-invasive species remained difficult to classify despite relatively high image availability, suggesting that performance was influenced more strongly by morphological overlap with invasive taxa than by sample size alone. This interpretation is consistent with the embedding analyses and explainability results, which indicate that invasive and non-invasive species occupy partially overlapping regions of morphospace and do not exhibit completely distinct morphological syndromes. In this context, low-performing non-invasive taxa may represent species whose architectural and/or foliar traits resemble those also associated with invasive species, thereby increasing classification ambiguity. More broadly, these results reinforce the idea that invasiveness in *Lythrum* is associated with distributed trait patterns rather than simple diagnostic features, and that image-based classification reflects a continuous morphological landscape rather than discrete ecological categories.

Although this study focuses solely on the genus *Lythrum*, the proposed methodology is inherently scalable to broader taxonomic groups and different ecological contexts, as it relied on general image representations and did not require species-specific model training. The use of a pretrained vision–language model (BioCLIP) enabled transferability across plant taxa, while the classification, explainability, and clustering

components could be applied to larger, more diverse datasets without modification. The core morphological categories considered here, i.e., leaf, stem, and flower, are broadly applicable across Angiosperms, providing a common basis for analysis, and the framework can be readily extended to incorporate taxon-specific traits where needed. The main practical requirements for scaling are the availability and quality of annotated image data to support training and evaluation. In larger, more heterogeneous datasets, additional considerations, such as variation in image background, phenological stage, or acquisition conditions, may require preprocessing steps or trait annotation schemes. Nonetheless, the modular structure of the pipeline makes it adaptable, and its use of intermediate features that can be directly interpreted in biological terms supports its application to other genera and ecological questions beyond *Lythrum*.

Moreover, a limitation of this study is the absence of direct comparisons with traditional machine learning models (e.g., Random Forest, SVM) based on hand-crafted morphological descriptors. While such baselines could be informative, they require explicit feature extraction, which is beyond the scope of this work. Future studies could explore these comparisons to further contextualize the proposed approach.

From a technical perspective, our analysis relied on a community-driven dataset that inherently comprises images acquired under highly variable conditions in terms of resolution, focus, lighting, and framing. While this variability enables large-scale data collection, it substantially limits the detection of fine-grained morphological traits beyond coarse structures such as flowers, stems, or leaves. Subtle characteristics are often not visible or not consistently captured, thus limiting the final trait analysis. Additionally, While the genus *Lythrum* serves as a viable case study due to the presence of different invasive species and high sample availability on iNaturalist, the dataset contains a significant taxonomic imbalance: 27 non-invasive species compared to only three invasive ones. This disparity is exacerbated by the exclusion of *L. hyssopifolia*, although the total number of individual samples remained balanced across categories. Also, the lack of invasive-only trait pairs reduces the possibilities of exploration of the model's behavior.

In the present study, the classification and explainability analyses primarily relied on organ-level structures such as stems, leaves, and flowers, rather than fine-scale diagnostic characters such as venation patterns, leaf margins, or floral parts. Consequently, resizing images likely did not substantially affect the biological signals driving model predictions. Similarly, the clustering stage relied on manually assigned labels corresponding to morphologically explicit and readily identifiable structures within *Lythrum*, reducing ambiguity in trait annotation. Nevertheless, we acknowledge that extending this framework to broader Angiosperm datasets or to taxa requiring finer-scale morphological discrimination may increase the importance of image resolution, expert-guided annotation protocols, and formal reproducibility assessments. In such cases, dedicated sensitivity analyses and multi-expert validation procedures would likely become necessary.

In our study, plants were classified as either invasive or non-invasive. However, species can become invasive in different ways, often relying on different combinations of traits (Palma et al., 2021). Furthermore, biological invasions are context-dependent, and a species can be invasive in one ecosystem but not in another (Catford et al., 2022; Muth and Pigliucci, 2006; Fristoe et al., 2021). We encourage future research to expand our deep learning approach to a larger number of species to implement a wider array of morphological traits, geographical distributions, and context-dependent definitions of invasiveness. We believe it would be particularly valuable to test whether image-based deep learning models can correctly tell apart the same invasive species in different areas (e.g., its native range vs. its invasive one).

In conclusion, this study showed that image-based models can recover biologically meaningful information from heterogeneous plant photographs and, crucially, that their predictions are shaped by interactions among morphological traits rather than by single diagnostic features. By integrating an explainability pipeline, embedding analysis, and trait-pair attribution, we showed that, in *Lythrum*, successful classification of invasiveness depended strongly on trait combinations of non-invasive species, while no uniquely invasive morphological signature was identified. Overall, these results highlight both the potential and limits of image-based approaches for invasion assessment, offering a scalable and interpretable method for linking deep learning to trait-based invasion ecology.

CRediT authorship contribution statement

Guido Spina: Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Barbara Frittella:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation. **Riccardo Ciarle:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Simone Monaco:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M.S. acknowledges travel support from ELIAS (GA no 101120237). The SmartData@PoliTO research centre of Politecnico di Torino, Italy, has partially funded this work.

Appendix A. Additional results and validation

A.1. Assessing phylogenetic effects on embedding similarity

To assess whether the learned visual representations were primarily structured by phylogenetic relatedness, we performed a correlation analysis on the subset of species for which phylogenetic information was available, following Vieira et al. (2022). For these species, we compared pairwise phylogenetic distances among species with pairwise cosine similarity between species-level mean BioCLIP 2 embeddings. The analysis showed no significant correlation between phylogenetic distance and embedding distance, regardless of the invasiveness of the species under comparison (Fig. A.9). This suggests that the model's image representations were not primarily driven by phylogenetic proximity.

A.2. Sensitivity analysis to under-represented species

To assess the potential impact of species-level imbalance, we performed an additional sensitivity analysis excluding all species represented by fewer than 20 images. The results, reported in Fig. A.10, were consistent with those obtained on the full dataset. Performance generally grew among the remaining species. This general slight improvement, common to both invasive and non-invasive species, suggests that the observed patterns were not driven by species with very small sample sizes. It is worth noting that *L. junceum*, *L. tribracteatum*, and *L. thymifolia* gained significant improvement, reaching nearly optimal classification performance. This could be explained by considering that, in embedding-based classifiers, low-sample species can introduce unstable or high-variance feature estimates, which may distort nearby decision boundaries and reduce separability for morphologically adjacent species. Importantly, this effect was not restricted to either invasive or non-invasive taxa, suggesting that the observed improvement reflects a general stabilization of the feature space rather than systematic bias toward a particular ecological category.

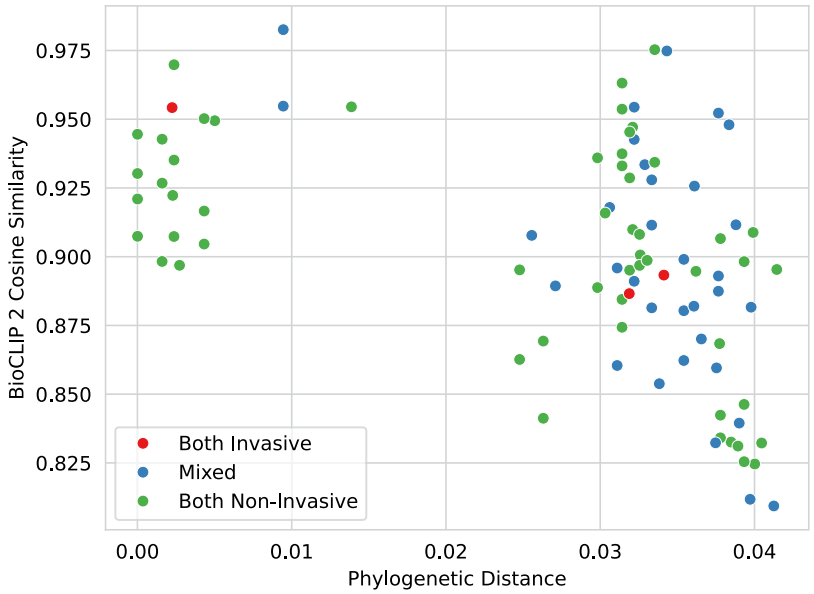


Fig. A.9. Scatter plot representing the phylogenetic distance against the model's embedding distance. The color of each point indicates whether the distance is between two invasive species, one invasive and one non-invasive (mixed), or two non-invasive species.

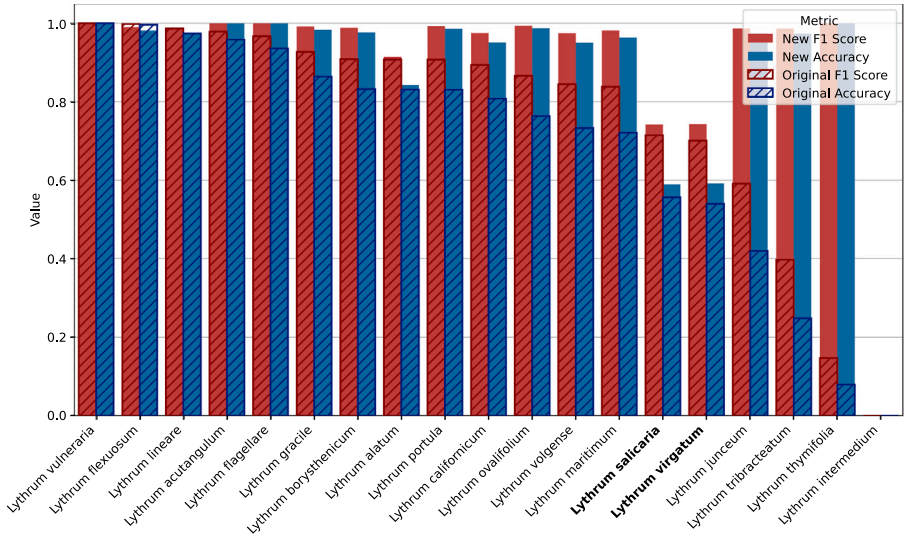


Fig. A.10. Species-level classification performance for the models trained in the LOSO validation scheme. For each species, the mean Accuracy and F1 Score across seeds are shown. Solid bars correspond to the training using all the available species except *L. hyssopifolia*, and the hatched bars overlay the models trained after removing the under-represented species too. Species associated with invasion are highlighted in bold on the *x*-axis.

A.3. Two-dimensional representation of species

To better study the difference in representation between the images of the different species, we applied Uniform Manifold Approximation and Projection (**UMAP**) to the multidimensional embeddings of the image data extracted with BioCLIP 2, reducing and mapping them to a 2-dimensional space (Fig. A.11). The parameters used in this phase for UMAP can be found in Table A.5.

Table A.5
Parameters used to map the embeddings in two dimensions with UMAP.

Parameter	Value
`n_neighbors`	15
`min_dist`	0.01
`metric`	euclidean
`random_state`	42

A.4. Explainability model selection

To assess the robustness of the attribution step, we compared Integrated Gradients with Gradient SHAP as an alternative explanation method. Integrated Gradients was retained as the main attribution method because it provided more spatially focused and biologically interpretable regions, consistent with its established use for input-level attribution in deep neural networks. In contrast, Gradient SHAP generated a substantially larger, more fragmented set of regions: from 2000 randomly selected images, it produced 20,021 extracted regions, compared with 7509 with Integrated Gradients. Although this suggests greater sensitivity, the resulting patches were less specific and less stable for downstream trait annotation. We therefore used Integrated Gradients for the main analysis and report Gradient SHAP as a sensitivity analysis.

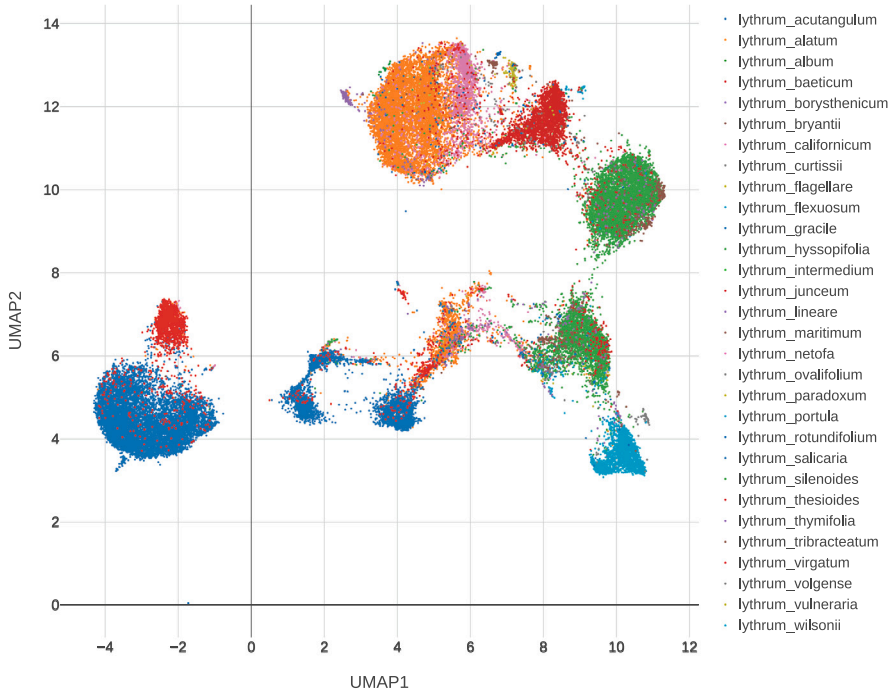


Fig. A.11. Two-dimensional UMAP projection of image embeddings. The invasive species are represented as follows: Lythrum salicaria (blue) occupies the region between UMAP1: −5 to 5 and UMAP2: 3 to 6.5. Lythrum virgatum (red) spans UMAP1: −5 to 6 and UMAP2: 4 to 8. Lythrum hyssopifolia (green) is located between UMAP1: 7 to 12 and UMAP2: 5 to 11.

Table A.6
Hyperparameter configurations explored during the clustering phase.

Hyperparameter		Values
Explainability method		Integrated gradients, Gradient SHAP
KMeans	n_clusters	[30, 35, 40, … , 100]
	n_neighbors	[5, 10, 15, 20, 30, 50]
UMAP	min_dist	[0.01, 0.025, 0.05, 0.075, 0.1]
	distance_metric	euclidean, manhattan

Table A.7
Hyperparameters and silhouette scores for the five best performing configurations tested for the clustering pipeline.

Explainability method	n clusters	n neighbors	Min dist	Dist metric	Silhouette score
Integrated Gradients	**30**	**5**	**0.025**	**manhattan**	**0.428**
Integrated Gradients	30	5	0.010	manhattan	0.427
Integrated Gradients	80	5	0.010	manhattan	0.425
Integrated Gradients	90	5	0.010	euclidean	0.425
Integrated Gradients	95	5	0.010	euclidean	0.425

Table A.8
Clustering results for the configuration with the best silhouette score.

Cluster size				Silhouette
min	max	mean	entropy	score
88	441	250.3	0.984	0.428

Fig. A.12 shows some examples, highlighting how attribution maps often align closely with biologically relevant structures, such as distinct plant organs or characteristic morphological features (Figs. A.12(a)–A.12(d)). In other instances, however, the highlighted areas do not correspond to meaningful biological traits. For example, in Fig. A.12(e), the method focuses on a human hand visible in the image, while in Fig. A.12(f), the heatmap is diffuse and fails to emphasize any clearly defined structure. All the examples clearly demonstrate the increased sensitivity of Gradient SHAP, which led to a preference for Integrated Gradients in the following steps.

A.5. Selected clustering configuration

To determine the most suitable explainability method for our workflow, as well as the optimal hyperparameter configuration for both KMeans and UMAP, we conducted a systematic exploration of different parameter combinations. The full range of tested hyperparameters is summarized in Table A.6. Each configuration was evaluated using the silhouette score computed on the resulting clustering assignments, as this metric provides a quantitative measure of both intra-cluster compactness and inter-cluster separation.

Combining all parameter values yields 900 distinct configurations for each explainability method. The Clustering was carried out using the salient regions extracted from a subset of 2000 images: Integrated Gradients produced 7509 regions, whereas Gradient SHAP yielded 20 021 regions. The five configurations with the highest silhouette score are reported in Table A.7. The configuration with the best silhouette score was selected for subsequent analyses. Table A.8 reports some statistics of the clusters with this configuration.

A.6. Clustering validation with HDBSCAN

To validate the reliability of the KMeans-based clustering and the consistency of manual cluster annotations, an independent validation analysis was performed using HDBSCAN. The validation employed the same set of 7509 regions extracted from the preliminary 2000-image subset used for the initial clustering phase. We finetuned the minimum cluster size of HDBSCAN by looking for the optimal value across the range:

$$min\_cluster\_size \in \{5, 6, 7, \dots, 71\}$$

To ensure comparability with the KMeans results, only configurations producing between 10 and 100 clusters were retained. Out of 66 tested configurations, 9 satisfied this criterion. Although this

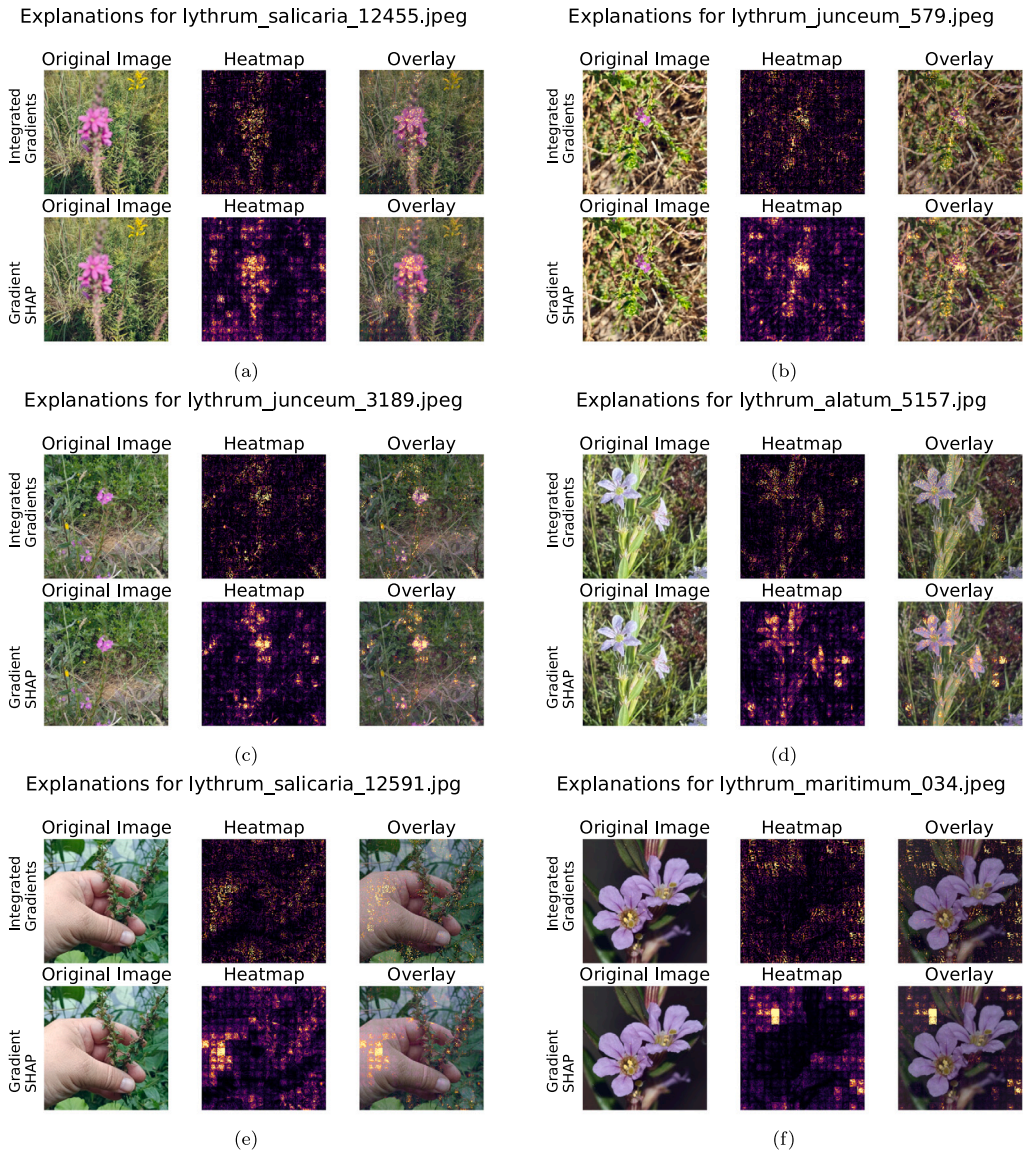


Fig. A.12. Examples of generated heatmaps. For each sample, the original image, the generated heatmap, and the overlay between the two are shown from left to right. Both Integrated Gradients (top row) and Gradient SHAP (bottom row) results are presented for each example. In most cases (a–d), the highlighted regions correspond to biologically meaningful structures, while in others (e–f) they do not align with the expected features.

represents a limited subset, the objective was not to optimize HDBSCAN performance but rather to assess whether an alternative clustering method, operating under different assumptions, would produce partitions consistent with the manual labels derived from KMeans clusters.

For each valid HDBSCAN configuration, cluster quality was evaluated using the *cluster label consistency ratio*, defined as:

$$\text{Consistency Ratio}_c = \frac{n_{\text{dominant}}}{n_{\text{total}}}$$

where n_{dominant} is the number of regions in cluster c that share the cluster's most frequent label set, and n_{total} is the total number of regions in cluster c. This metric quantifies cluster purity with respect to the manually assigned labels. A value of 1.0 indicates perfect homogeneity, while lower values reflect greater label diversity within a cluster.

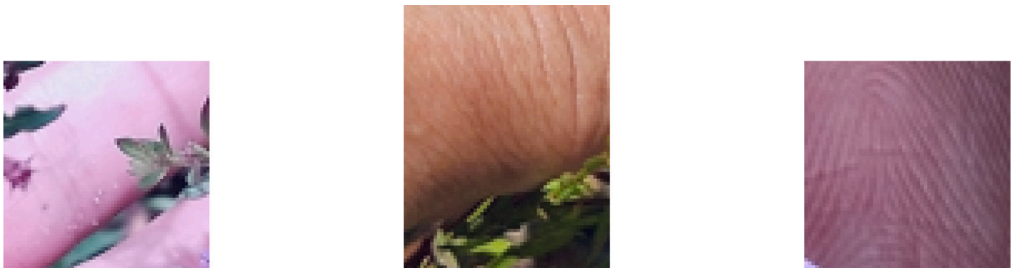
Across the nine valid HDBSCAN configurations, the mean cluster label consistency ratio was 0.75 ± 0.08. The best-performing configuration, corresponding to $min\_cluster\_size = 16$, achieved a mean consistency ratio of 0.92, approaching near-perfect alignment with the manual annotations.

A.7. Cluster-specific region examples

To illustrate the visual characteristics of different clusters, representative examples from four selected clusters are presented below. Each cluster was manually assigned a label set based on visual inspection of its constituent regions. These examples demonstrate the diversity of morphological structures captured by the clustering procedure and provide qualitative support for the assigned annotations.

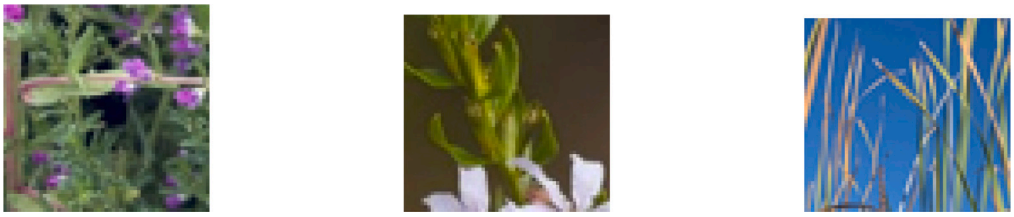
A.8. Cluster 0: Hand

Cluster 0 was assigned the label *Hand*, indicating that its regions primarily contain portions of human hands or skin visible in the original photographs. Three representative regions from this cluster are shown in Fig. A.13. All three examples exhibit skin tones, curved contours resembling fingers or palms, and minimal plant material. These regions likely originated from images where observers held plant specimens during photography, a common occurrence in citizen science datasets such as iNaturalist. The presence of such non-biological features underscores the importance of the clustering and labeling pipeline in



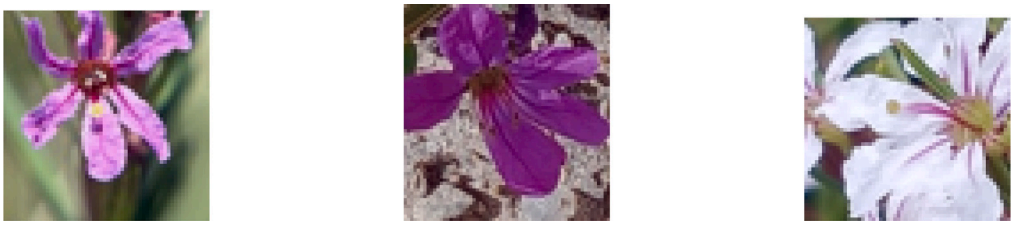
(a) *Lythrum alatum* (b) *Lythrum californicum* (c) *Lythrum alatum*

Fig. A.13. Representative regions from Cluster 0 (*Hand*). All regions display human skin or hand portions visible in the original images.



(a) *Lythrum junceum* (b) *Lythrum alatum* (c) *Lythrum salicaria*

Fig. A.14. Representative regions from Cluster 16 (*Leaf, Stem, Flower*). Regions contain multiple overlapping plant structures.



(a) *Lythrum californicum* (b) *Lythrum junceum* (c) *Lythrum lineare*

Fig. A.15. Representative regions from Cluster 23 (*Flower*). All regions prominently display floral structures.

identifying and filtering spurious visual cues that should not influence invasiveness predictions.

A.9. Cluster 16: Leaf, stem, flower

Cluster 16 was assigned a composite label set comprising *Leaf*, *Stem*, and *Flower*, reflecting its inclusion of regions depicting multiple plant structures simultaneously. Three representative regions are shown in Fig. A.14. Each example contains visible leaf structures, with at least one also displaying stem or floral elements. The regions exhibit green foliage, elongated leaf shapes, and in some cases, small purple or pink flowers characteristic of *Lythrum* species. This cluster represents a common scenario in natural plant images, where multiple organs appear within the same cropped region due to overlapping structures or close spatial proximity.

A.10. Cluster 23: Flower

Cluster 23 was labeled *Flower*, indicating exclusive or dominant presence of floral structures. Three representative regions are presented in Fig. A.15. Each example prominently features purple or pink flowers with visible petals and, in some cases, central stamens. Background elements and foliage are minimal or absent, allowing the floral morphology to dominate the region. These regions likely correspond to close-up photographs focused on inflorescences, which are diagnostic features for species identification in the *Lythrum* genus.

A.11. Cluster 27: Background/Undefined

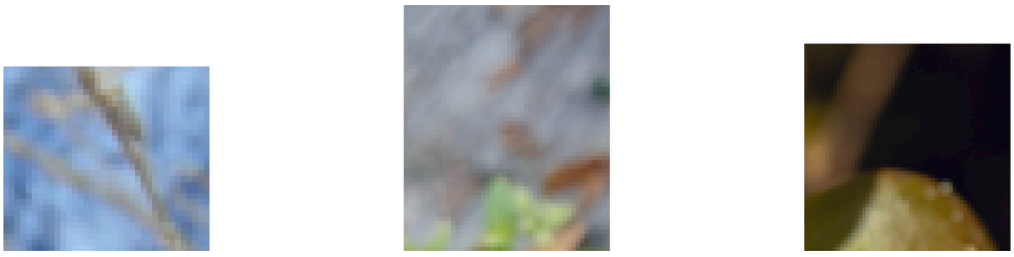
Cluster 27 was assigned the label *Background/undefined*, indicating regions that lack clear biological content or consist primarily of background elements. Three representative regions are shown in Fig. A.16. The examples display blurred or out-of-focus areas, soil, debris, or diffuse patches of color without recognizable plant structures. These regions likely arise from portions of images where the explainability method assigned moderate attribution scores to non-informative areas, possibly due to contextual cues or artifacts in the model's learned representations. The identification and labeling of such clusters is essential for downstream filtering, ensuring that subsequent trait-based analyses focus on biologically meaningful regions.

A.12. Manual cluster labeling results

We report the resulting structural labels distribution across all clusters (see 3.3) as presented in A.17

Appendix B. Labels enrichment with species characteristic traits

We enrich the labels inherited from the cluster assignment (see 3.3) using specific traits for each species and for each label. For each region extracted by images of our dataset, we first identified the species it belonged to. Each region is labeled (after the clustering phase) to indicate the presence of key biological structures *Leaf, Flower and Stem*. Using this information, we assigned to the image the characteristic traits



(a) *Lythrum californicum* (b) *Lythrum salicaria* (c) *Lythrum portula*

Fig. A.16. Representative regions from Cluster 27 (*Background/undefined*). Regions contain non-informative background elements or lack recognizable plant structures.

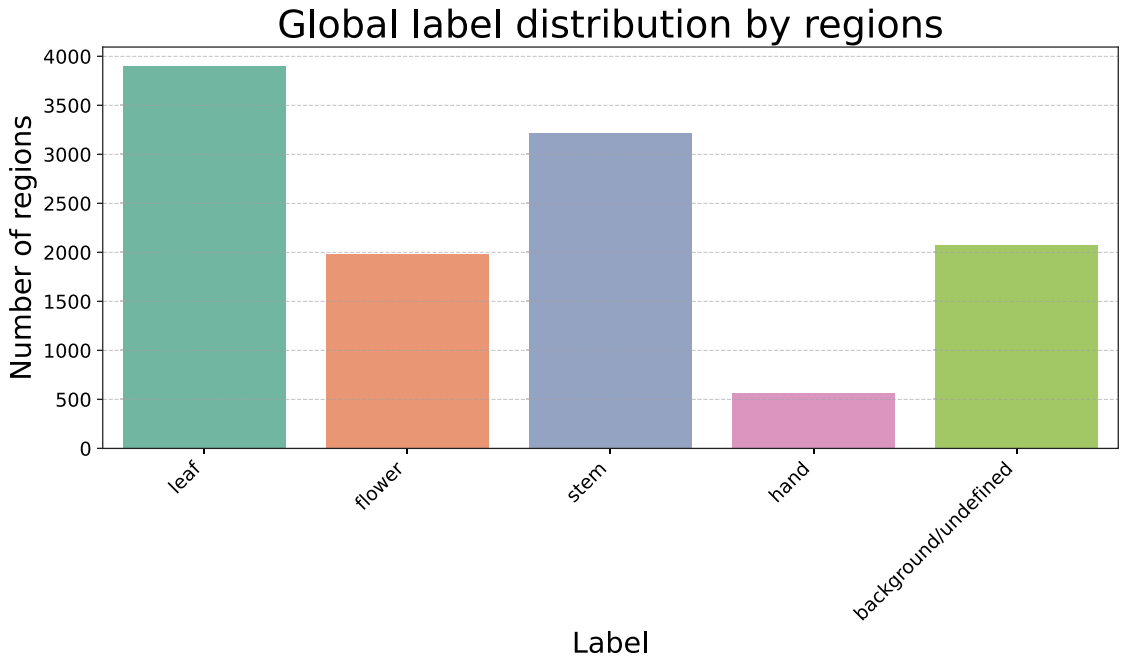


Fig. A.17. Results of manual cluster labeling: global distribution by number of regions (a) and labelset assigned to each cluster (b) are shown.

associated with the present structures for that species, as defined in B.9. For example, a region of *Lythrum anatolicum* labeled as containing both flower and stem would be annotated with the traits *Petals purple, Stamens 12* for the flower and *Erect* for the stem. The final enriched set of features comprises 17 unique Leaf features, 38 unique Flower features, and 9 Stem ones.

Appendix C. Single-trait and metric analysis

In Section 2.4 we mentioned the analysis conducted in an initial exploratory step, here we describe the methodology used and the results obtained.

C.1. Methodology

Region-level outputs were aggregated at the image level. For each image, we computed:

- **Ground truth** and **predicted class** computed by our classification model;
- **Prediction correctness**: binary indicator (if ground truth and predicted class correspond) and error type:
  - **True Positive (TP)** = Invasive plants correctly predicted;
  - **True Negative (TN)** = Non Invasive plants correctly predicted;
  - **False Positive (FP)** = Non Invasive plants predicted as Invasive;
  - **False Negative (FN)** = Invasive plants predicted as Non Invasive.
- **Regions metrics**: number of regions extracted from the image, *Hand* and *Background/undefined* fractions (representing the proportions of regions that contain these labels), coverage fraction (computed as the sum of the areas covered by the regions divided by the total area of the image);
- **Trait metrics**: presence of each characteristic trait (listed in Appendix B), relative frequency within the image, total count of distinct traits present (richness), and the Pielou evenness index, defined as

$$J' = \frac{H'}{\log S}$$

where $H' = -\sum_{i=1}^{S} p_i \log p_i$ is the Shannon diversity of the trait distribution, p_i is the relative frequency of trait i in the image, and S is the number of distinct traits (richness). The index J' ranges from 0 (uneven distribution, dominated by a few traits) to 1 (perfectly even distribution across all traits).

We applied a Random Forest classifier to identify which image-level features are associated with prediction correctness. The input features included both trait presence and frequency, the number of extracted regions, the number of distinct traits (richness), the coverage fraction, the

Table B.9
Characteristic traits for each species for each biological structure.

Species	Leaf	Flower	Stem
Lythrum acutangulum	–	–	–
Lythrum alatum	Opposite Opposite, becoming alternate distally Sessile Attenuate at the base	Inflorescence raceme Petals purple Floral tube cylindrical Stamens 6	Erect
Lythrum album	Alternate	Petals white	Erect
Lythrum anatolicum	Opposite Sessile Cordate at the base	Petals purple Stamens 12	Erect
Lythrum baeticum	–	–	–
Lythrum borysthenicum	Opposite, becoming alternate distally Sessile Obovate	Petals reddish Petals minute Stamens 6	Erect Erect or decumbent
Lythrum bryantii	–	–	–
Lythrum californicum	Opposite Opposite, becoming alternate distally Linear	Inflorescence raceme Petals purple Stamens 5–8	Erect
Lythrum curtissii	Opposite Opposite, becoming alternate distally Sessile or subsessile Attenuate at the base	Inflorescence raceme Petals lilac to pink Usually with a darker midrib Floral tube obconic Stamens 6	Erect
Lythrum flagellare	Opposite Petiolated Rounded at the base A discernible gap between the stem and the base of the blade	Inflorescence raceme Floral tube obconic without red dots Petals purple Stamens 6	Creeping to weakly erect
Lythrum flexuosum	Alternate Sessile	Calyx with alternate long and small teeth Flowers solitary Petals purple	Creeping
Lythrum gracile	Opposite Opposite, becoming alternate distally Rounded at the base	Petals white to pink	Erect
Lythrum hyssopifolia	Alternate Sessile Rounded at the base	Inflorescence raceme Floral tube obconic without red dots Petals pink Calyx with alternate long and small teeth Stamens 4–6	Erect to weakly erect
Lythrum intermedium	Opposite Sessile Opposite, becoming alternate distally Rounded at the base	Inflorescence spikelike Flowers in whorled clusters Calyx with alternate long and small teeth Floral tube cylindrical Petals purple Stamens 12	Erect
Lythrum junceum	Alternate Subsessile Obtuse to truncate at the base	Inflorescence raceme Flowers solitary in leaf axils Floral tube obconic Floral tube red dotted Petals purple Stamens 12	Sprawling or ascending
Lythrum lineare	Opposite Sessile Attenuate at the base	Inflorescence raceme Floral tube cylindrical Petals pale purple or whitish Usually with a darker midrib Stamens 6	Erect
Lythrum maritimum	Opposite Subsessile	Petals pink Usually with a darker midrib	Prostrate
Lythrum netofa	Sessile Broader at the base	Flowers solitary Floral tube campanulate Petals 4 Petals purple with a darker midrib Stamens 6–8	Erect
Lythrum ovalifolium	Alternate Sessile or subsessile Attenuate at the base	Inflorescence raceme Floral tube obconic without red dots Petals purple with a darker midrib Stamens 6	Erect or decumbent

(*continued on next page*)

Table B.9 (*continued*).

Lythrum paradoxum	Alternate Sessile	Petals pink to purple Stamens 10–12	Erect
Lythrum portula	Opposite Sessile	Inflorescence spikelike Floral tube campanulate Petals white to pink Stamens 5–8	Prostrate and spreading
Lythrum rotundifolium	Petiolated	Flowers solitary Petals pink to purple with a darker midrib Stamens 8	Prostrate
Lythrum salicaria	Opposite Sessile Opposite, becoming alternate distally Rounded at the base	Inflorescence spikelike Flowers in whorled clusters Calyx with alternate long and small teeth Floral tube cylindrical Petals purple Stamens 12	Erect
Lythrum silenoides	–	–	–
Lythrum thesioides	Alternate	Petals 4 Petals pink	Erect
Lythrum thymifolia	Needle-like	One or few flowers in the axil of leaves Calyx with alternate long and small teeth Stamens 2–3	Prostrate
Lythrum tribracteatum	Opposite Sessile Attenuate at the base	Inflorescence spikelike Floral tube narrowly cylindrical without red dots Calyx with teeth of the same length Petals lavender Stamens 4–6	Prostrate to weakly erect
Lythrum virgatum	Opposite Sessile Narrower at the base	Inflorescence spikelike to raceme Flowers in whorled clusters Floral tube cylindrical Petals pink to purple Stamens 10–14	Erect
Lythrum volgense	Needle-like	Petals pink Petals minute	Prostrate
Lythrum vulneraria	Opposite Opposite, becoming alternate distally	Petals pink	Erect
Lythrum wilsonii	Alternate Sessile Rounded at the base	Petals pink to purple	Erect

Table C.10
Global feature importance analysis from Random Forest classification. The table reports both impurity based importance and permutation importance (accompanied by the standard deviation across 20 repetitions) for each feature. A dash (−) indicates that the value was not computed or not meaningful for that feature, typically due to low prevalence or insufficient variation in the subset.

Feature	Impurity	Permutation	
	Importance	Importance	Std
coverage_frac	0.6198	0.2224	0.0017
pielou_evenness	0.4101	0.0557	0.0007
background_frac	0.0329	0.0729	0.0009
hand_frac	0.0314	0.0571	0.0009
trait_erect_freq	0.0287	0.0545	0.0005
trait_rounded_at_the_base_freq	0.0216	0.0387	0.0009
richness	0.0021	0.0486	0.0008
trait_rounded_at_the_base_present	0.0152	–	–
n_regions	0.0149	0.0564	0.0008
trait_sessile_freq	0.0125	0.0304	0.0007

Pielou evenness index, the hand and background/undefined fractions. Two complementary analyses were performed:

1. a **global** analysis, ranking predictors of correct vs. incorrect classifications, and
2. a **error-type** analysis, conducted in a one-vs-all manner (TP, TN, FP, FN) to identify features contributing to specific error patterns.

Feature importance was quantified in two ways. First, *impurity-based importance* measured each feature's average contribution to reducing node impurity across the forest, providing a fast estimate. However, it can be biased toward variables with many categories or continuous scales, and it may overestimate the importance of features that produce strong splits for a small subset of samples. Second, *permutation-based importance* assessed the decrease in forest accuracy when the values of a single feature were randomly shuffled, offering a measure of each feature's true predictive influence that is less susceptible to biases from feature's scale or cardinality.

To further investigate the Random Forest analysis, we analyzed the correlation between specific image metrics and prediction correctness. Each metric was binned into discrete categories, chosen either by quantiles (for continuous metrics) or fixed intervals (for counts), and the average prediction accuracy was computed per bin.

Within each bin, species composition was analyzed by computing the distribution of taxa and comparing it to the overall dataset using the Kullback–Leibler (KL) divergence:

$$\mathrm{KL}(P \parallel Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}$$

where $P(i)$ and $Q(i)$ are the proportions of species i in the bin and in the overall dataset, respectively. This analysis allowed us to assess whether bins with distinct accuracy patterns also had broadly comparable species distributions; if a bin's accuracy differed but the distribution was highly skewed, the observed difference might reflect species composition rather than the metric itself.

Differences in accuracy across bins were tested using one-way ANOVA, which evaluates whether the mean accuracy differs significantly between two or more groups (bins). In this context, ANOVA

Table C.11
Feature importance analysis from Random Forest for True Positive (TP) and True Negative (TN) classifications. The table reports both impurity based importance and permutation importance (accompanied by the standard deviation across 20 repetitions) for each feature. A dash (–) indicates that the value was not computed or not meaningful for that feature, typically due to low prevalence or insufficient variation in the subset.

Feature	Impurity		Permutation			
	Importance		Importance		Std	
	TP	TN	TP	TN	TP	TN
coverage_frac	0.4099	0.1650	0.1642	0.0881	0.0013	0.0011
trait_rounded_at the_base_freq	0.0652	0.1015	0.0300	–	0.0006	–
trait_rounded_at the_base_present	0.0613	0.0981	0.0219	–	0.0006	–
trait_flowers_in whorled_clusters_freq	0.0331	0.0542	–	–	–	–
trait_flowers_in_ _whorled_clusters_present	0.0278	0.0426	–	–	–	–
pielou_evenness	0.0272	0.0327	0.0294	0.0327	0.0006	0.0004
trait_attenuate_at_ _the_base_freq	0.0219	0.0320	–	0.0205	–	0.0003
background_frac	0.0218	–	0.0508	0.0446	0.0006	0.0006
trait_erect_freq	0.0209	–	0.0325	0.0220	0.0005	0.0005
hand_frac	0.0193	–	0.0433	0.0276	0.0007	0.0005
richness	0.0185	0.0186	0.0273	0.0345	0.0005	0.0005
n_regions	–	–	0.0413	0.0352	0.0006	0.0008
trait_sessile_freq	–	–	0.0214	0.0222	0.0006	0.0006
trait_opposite_freq	–	–	–	0.0148	–	0.0003
trait_floral_tube cylindrical_freq	–	–	–	0.0147	–	0.0004

determines whether variation in prediction correctness is actually associated with the binned metric, while accounting for within-bin variance.

This approach allowed us to quantify how the chosen metrics are associated with model performance while controlling for potential confounding effects of species composition, completing the feature importance analysis from the Random Forests.

C.2. Results

Tables C.10–C.12 display the results for the global analysis, for True Positive and True Negative analysis, and for False Positive and False Negative analysis, respectively.

Table C.10 summarizes the global feature importance for correct predictions across all images. The fraction of the image covered by extracted regions (`coverage_frac`) is by far the most important predictor (impurity importance= 0.6198; permutation importance= 0.2224), indicating that images with larger region coverage are much more likely to be correctly classified. Secondary features include `pielou_evenness`, `background_frac` and `hand_frac`, which contribute moderately to global prediction accuracy. Permutation importance is generally smaller than impurity, but confirms the rank order of the main predictors. Interestingly, features as `richness` and `n_regions` show low impurity importance but moderate permutation importance, suggesting they might interact with other features to affect model performance.

Overall, the global analysis identifies the strongest single predictors but may obscure feature-specific effects that vary across different types of predictions.

To capture more detailed relationships, we performed a separate analysis for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Tables C.11,C.12). This approach allows us to determine which features drive specific types of correct and incorrect predictions, providing a more nuanced understanding than the global analysis alone.

For TP and TN outcomes, `coverage_frac` remains the dominant predictor (TP: 0.4099 impurity, 0.1642 permutation; TN: 0.1650 impurity, 0.0881 permutation), confirming its strong influence on correct classification. Traits related to leaf base morphology (`rounded_at_the_base`) and to flower morphology (`flowers_in_whorled_clusters`) also show elevated importance, suggesting that these features help distinguish between true positive and true negative classifications. Permutation importance generally aligns with impurity ranking, supporting robustness of the identified predictors. Notably, while the overall patterns are similar, TP outcomes show higher reliance on `coverage_frac`, whereas TN outcomes emphasize certain trait-specific features slightly more, highlighting subtle differences in the drivers of correct positive versus correct negative predictions.

For FP and FN outcomes, `coverage_frac` again dominates (FP: 0.6493 impurity, 0.0684 permutation; FN: 0.5372 impurity, 0.1558 permutation). Other contributors include `pielou_evenness`, `background_frac` and `hand_frac`. Permutation importance highlights additional moderate contributors, such as `n_regions` and `attenuate_at_the_base_freq`, suggesting that feature interactions and nonlinear effects influence error patterns. Notably, FP and FN patterns are broadly similar.

As suggested by the Random Forest analysis, the fraction of the image covered by extracted regions (`coverage_frac`) emerged as the most influential predictor of model correctness. To further investigate this relationship, we analyzed how prediction accuracy varies across different levels of region coverage. Images were grouped into eleven coverage categories, and the mean accuracy, sample size, and corresponding species distribution divergence (KL divergence) were computed for each bin (Table C.13, Fig. C.18).

Overall, results reveal that classification accuracy remains largely stable across most coverage ranges, fluctuating between 0.76 and 0.79 up to 25% of coverage (Fig. C.18(a)). Only the smallest (<1%) and largest (>25%) bins show deviations, although these categories include very few samples (220 and 65 images, respectively), making their estimates less reliable. The extremely high accuracy in the last bin (100%) is based on a single image and therefore not meaningful. Species distribution divergence (KL) remains low across all major bins

Table C.12
Feature importance analysis from Random Forest for False Positive (FP) and False Negative (FN) classifications. The table reports both impurity based importance and permutation importance (accompanied by the standard deviation across 20 repetitions) for each feature. A dash (−) indicates that the value was not computed or not meaningful for that feature, typically due to low prevalence or insufficient variation in the subset.

Feature	Impurity		Permutation			
	Importance		Importance		Std	
	FP	FN	FP	FN	FP	FN
coverage_frac	0.6493	0.5372	0.0684	0.1558	0.0008	0.0014
pielou_evenness	0.0470	0.0344	0.0287	0.0301	0.0004	0.0007
background_frac	0.0357	0.0279	0.0263	0.0482	0.0004	0.0005
hand_frac	0.0344	0.0266	0.0177	0.0400	0.0005	0.0006
trait_erect_freq	0.0274	0.0262	0.0217	0.0313	0.0004	0.0005
trait_attenuate_atthe_base_freq	0.0253	0.0128	0.0305	0.0074	0.0005	0.0001
richness	0.0218	0.0198	0.0249	0.0294	0.0005	0.0005
n_regions	0.0172	0.0124	0.0195	0.0393	0.0004	0.0008
trait_rounded_atthe_base_freq	–	0.0403	–	0.0333	–	0.0007
trait_rounded_atthe_base_present	–	0.0341	–	0.0202	–	0.0005
trait_sessile_freq	0.0102	0.0120	0.0176	0.0166	0.0004	0.0007
trait_opposite_freq	0.0106	–	0.0171	–	0.0004	–
trait_oppositebecoming_alternate distally_freq	0.0116	–	0.0215	0.0122	0.0005	0.0006
trait_stamens_6_freq	0.0125	–	0.0198	–	0.0003	–
trait_floral_tubecylindrical_freq	0.0087	–	0.0134	–	0.0003	–

Table C.13
Results of Region coverage correlation with accuracy analysis. For each region coverage category, mean accuracy for the category, sample size and KL divergence from the dataset distribution is shown.

Category	Accuracy	Sample size	KL divergence
0%–1%	0.759	220	0.009
1%–2%	0.778	2977	0.007
2%–3%	0.776	2134	0.008
3%–4%	0.770	3891	0.005
4%–5%	0.781	3419	0.003
5%–6%	0.783	3754	0.032
6-7.5%	0.783	5086	0.002
7.5%–10%	0.787	6035	0.003
10%–25%	0.769	6807	0.007
25%–50%	0.862	65	0.201
50%–100%	1.000	1	1.22

ANOVA: $F = 1.221$, $p = 0.2712$; the highest mean accuracy is observed in the 50%–100% coverage category (1.000) and the lowest in the 0%–1% category (0.759). No strong statistical evidence of differences across categories.

(<0.01), confirming that these accuracy patterns are not driven by differences in taxonomic composition (Fig. C.18(b)).

These findings indicate that, contrary to what the Random Forest importance might suggest, region coverage alone does not strongly determine prediction accuracy. The Random Forest model likely interpreted the separation between low- and high-coverage samples as a strong discriminative signal, even though the underlying relationship is weak or non-causal. This behavior is consistent with how impurity-based importance in Random Forests operates: features that allow a clear partition of the data, even if only over a small subset of samples, can receive disproportionately high importance. In our case, a few extreme high-coverage samples show higher accuracy and are easily separable from the rest. The model therefore identifies coverage as a useful split criterion, despite its limited general predictive power across the dataset.

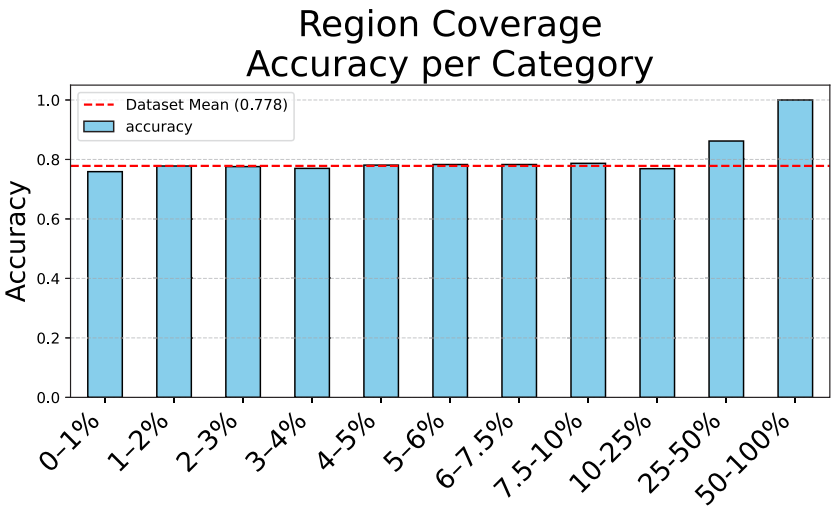
It is also important to note that the vast majority of images in our dataset exhibit very low coverage, while samples with higher coverage values are comparatively rare and display larger KL divergence from overall species distribution. This indicates that these images are not representative of the general dataset and may bias the Random Forest's assessment of importance. Consequently, within available data, there is **no clear evidence of a consistent dependency between coverage and prediction accuracy**, although such a relationship cannot be completely excluded given the strong imbalance in the coverage distribution.

In addition to region coverage, the same analysis was extended to all other metrics identified as potentially relevant by the Random Forests, including the Pielou evenness index, the number of distinct traits (richness), hand and background fractions, and image complexity. The detailed results of these analyses are reported in C.18(b). Overall, no other metric showed a clear or consistent relationship with prediction accuracy, except for weak or dataset-dependent effects. In particular, while some metrics exhibit statistically significant trends, these patterns were often accompanied by strong differences in species composition across bins (high KL divergence), suggesting that the observed variations are likely driven by taxonomic imbalance rather than intrinsic effects of the metrics themselves.

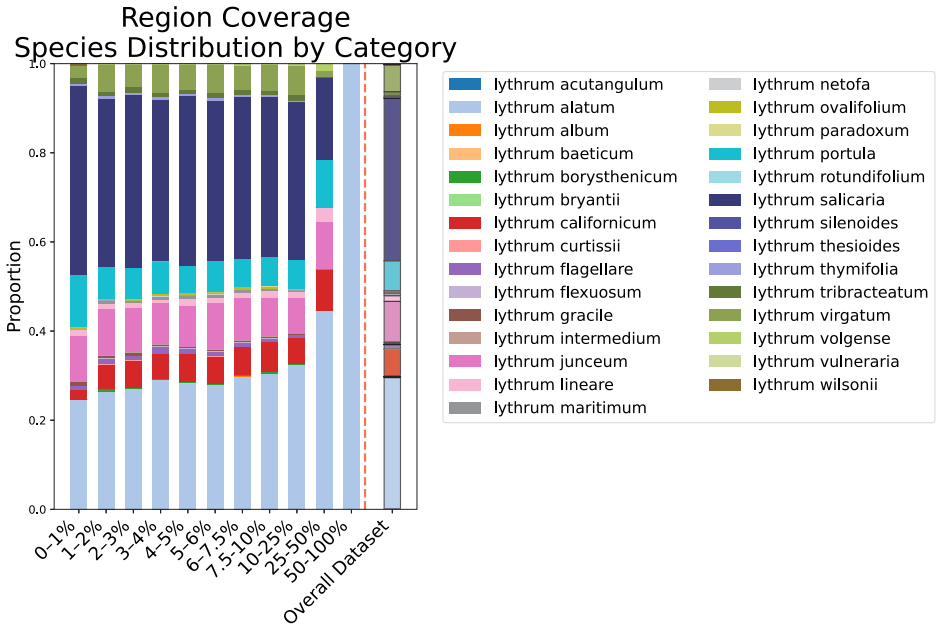
Data availability

The images used in this study were obtained from the iNaturalist Open Dataset(https://github.com/inaturalist/inaturalist-open-data), a publicly available repository of biodiversity observations and associated photographs hosted on Amazon Web Services (AWS) S3. We used a curated subset of images corresponding to selected *Lythrum* taxa, based on observation and photo metadata provided in the official iNaturalist snapshot (CSV files for observations, photos, and taxa).

Due to licensing constraints and the large size of the dataset, the specific image subset analyzed in this study is not redistributed. All images remain publicly accessible through the iNaturalist Open Dataset under their respective Creative Commons licenses. Photo IDs and taxon-level selection criteria used to assemble the dataset are available from the corresponding author upon reasonable request, enabling full reproducibility by direct download from the original source.



(a)



(b)

Fig. C.18. For the Region coverage analysis, accuracy for each category (a) and distribution of species for each category (b) is shown.

References

Bora, L.S., Padial, A.A., 2023. A global review on invasive traits of macrophytes and their link to invasion success. Acta Limnol. Bras. 35, e20. http://dx.doi.org/10.1590/S2179-975X4222, URL: https://www.scielo.br/j/alb/a/JYynHxXSMvJqzDMghRPHyTx/?format=html&lang=en.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A., 2017. Going deeper in the automated identification of herbarium specimens. BMC Evol. Biol. 17, 181.

Catford, J.A., Jansson, R., Nilsson, C., 2009. Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. Diversity and Distributions 15, 22–40. http://dx.doi.org/10.1111/j.1472-4642.2008.00521.x, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1472-4642.2008.00521.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1472-4642.2008.00521.x.

Catford, J.A., Wilson, J.R.U., Pyšek, P., Hulme, P.E., Duncan, R.P., 2022. Addressing context dependence in ecology. Trends Ecol. Evolut. 37, 158–170. http://dx.doi.org/10.1016/j.tree.2021.09.007, URL: https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(21)00258-5.

Chai, Y., Miao, C., AghaKouchak, A., Pokhrel, Y., Fu, Y., Li, X., Ji, J., Zhang, Q., Peñuelas, J., 2025. Flash droughts exacerbate global vegetation loss and delay recovery. Nat. Commun..

Dawson, W., Moser, D., van Kleunen, M., Kreft, H., Pergl, J., Pyšek, P., Weigelt, P., Winter, M., Lenzner, B., Blackburn, T.M., Dyer, E.E., Cassey, P., Scrivens, S.L., Economo, E.P., Guénard, B., Capinha, C., Seebens, H., García-Díaz, P., Nentwig, W., García-Berthou, E., Casal, C., Mandrak, N.E., Fuller, P., Meyer, C., Essl, F., 2017. Global hotspots and correlates of alien species richness across taxonomic groups. Nat. Ecol. Evol. 1, 0186. http://dx.doi.org/10.1038/s41559-017-0186, URL: https://www.nature.com/articles/s41559-017-0186.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Colin Prentice, I., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Chave, J., Joseph Wright, S., Sheremet'ev, S.N., Jactel, H., Baraloto, C., Cerabolini, B., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A., Falczuk, V., Rüger, N., Mahecha, M.D., Gorné, L.D., 2016. The global spectrum of plant form and function. Nature 529, 167–171. http://dx.doi.org/10.1038/nature16489, URL: https://www.nature.com/articles/nature16489.

Divíšek, J., Chytrý, M., Beckage, B., Gotelli, N.J., Lososová, Z., Pyšek, P., Richardson, D.M., Molofsky, J., 2018. Similarity of introduced plant species to native ones facilitates naturalization, but differences enhance invasion success. Nat. Commun. 9,

4631. http://dx.doi.org/10.1038/s41467-018-06995-4, URL: https://www.nature.com/articles/s41467-018-06995-4.

Dosovitskiy, A., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Essl, F., Dawson, W., Kreft, H., Pergl, J., Pyšek, P., Van Kleunen, M., Weigelt, P., Mang, T., Dullinger, S., Lenzner, B., Moser, D., Maurel, N., Seebens, H., Stein, A., Weber, E., Chatelain, C., Inderjit, Genovesi, P., Kartesz, J., Morozova, O., Nishino, M., Nowak, P.M., Pagad, S., Shu, W.S., Winter, M., 2019. Drivers of the relative richness of naturalized and invasive plant species on earth. AoB PLANTS 11, lz051. http://dx.doi.org/10.1093/aobpla/plz051.

Essl, F., Lenzner, B., Bacher, S., Bailey, S., Capinha, C., Daehler, C., Dullinger, S., Genovesi, P., Hui, C., Hulme, P.E., Jeschke, J.M., Katsanevakis, S., Kühn, I., Leung, B., Liebhold, A., Liu, C., MacIsaac, H.J., Meyerson, L.A., Nuñez, M.A., Pauchard, A., Pyšek, P., Rabitsch, W., Richardson, D.M., Roy, H.E., Ruiz, G.M., Russell, J.C., Sanders, N.J., Sax, D.F., Scalera, R., Seebens, H., Springborn, M., Turbelin, A., van Kleunen, M., von Holle, B., Winter, M., Zenni, R.D., Mattsson, B.J., Roura-Pascual, N., 2020. Drivers of future alien species impacts: An expert-based assessment. Global Change Biol. 26, 4880–4893. http://dx.doi.org/10.1111/gcb.15199, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15199, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.15199.

Fister, Jr., I., Emsenhuber, G., Plümer, J.H., Fister, I., Holzinger, A., 2026. Niarules: Advancing interpretable machine learning through numerical association rule mining and 3d coral plot visualization. SoftwareX 33, 102470.

Fristoe, T.S., Chytrý, M., Dawson, W., Essl, F., Heleno, R., Kreft, H., Maurel, N., Pergl, J., Pyšek, P., Seebens, H., Weigelt, P., Vargas, P., Yang, Q., Attorre, F., Bergmeier, E., Bernhardt-Römermann, M., Biurrun, I., Boch, S., Bonari, G., Botta-Dukát, Z., Bruun, H.H., Byun, C., Čarni, A., Carranza, M.L., Catford, J.A., Cerabolini, B.E.L., Chacón-Madrigal, E., Ciccarelli, D., Ćuštarevska, R., de Ronde, I., Dengler, J., Golub, V., Haveman, R., Hough-Snee, N., Jandt, U., Jansen, F., Kuzemko, A., Küzmič, F., Lenoir, J., Macanović, A., Marcenò, C., Martin, A.R., Michaletz, S.T., Mori, A.S., Niinemets, U., Peterka, T., Pielech, R., Rašomavičius, V., Rūsiņa, S., Dias, A.S., Šibíková, M., Šilc, U., Stanisci, A., Jansen, S., Svenning, J.C., Swacha, G., van der Plas, F., Vassilev, K., van Kleunen, M., 2021. Dimensions of invasiveness: Links between local abundance, geographic range size, and habitat breadth in europe's alien and native floras. Proc. Natl. Acad. Sci. 118, e2021173118. http://dx.doi.org/10.1073/pnas.2021173118, URL: https://www.pnas.org/doi/abs/10.1073/pnas.2021173118.

Gallagher, R.V., Randall, R.P., Leishman, M.R., 2015. Trait differences between naturalized and invasive plant species independent of residence time and phylogeny. Conserv. Biol. 29, 360–369. http://dx.doi.org/10.1111/cobi.12399, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.12399, _eprint: https://conbio.onlinelibrary.wiley.com/doi/pdf/10.1111/cobi.12399.

García, Y., Callaway, R.M., Diaconu, A., Montesinos, D., 2013. Invasive and non-invasive congeners show similar trait shifts between their same native and non-native ranges. PLOS ONE 8, e82281. http://dx.doi.org/10.1371/journal.pone.0082281, URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0082281.

Goëau, H., Bonnet, P., Joly, A., 2023. Overview of plantclef 2022: Image-based plant identification at global scale. arXiv preprint arXiv:2509.17632.

Goeau, H., Espitalier, V., Bonnet, P., Joly, A., 2025. Overview of plantclef 2024: multi-species plant identification in vegetation plot images. arXiv preprint arXiv:2509.15768.

Granier, C., Vile, D., 2014. Phenotyping and beyond: modelling the relationships between traits. Curr. Opin. Plant Biol. 18, 96–102.

Gu, J., Stevens, S., Campolongo, E.G., Thompson, M.J., Zhang, N., Wu, J., Kopanev, A., Mai, Z., White, A.E., Balhoff, J., Dahdul, W., Rubenstein, D., Lapp, H., Berger-Wolf, T., Chao, W.L., Su, Y., 2025. Bioclip 2: Emergent properties from scaling hierarchical contrastive learning. http://dx.doi.org/10.48550/arXiv.2505.23883, URL: http://arxiv.org/abs/2505.23883. arXiv:2505.23883 [cs].

Han, Y., Jeon, W., Kim, T., Chung, S., 2026. Leafvision: Self-supervised agricultural vision foundation models for plant disease classification. Eng. Appl. Artif. Intell. 176, 114660. http://dx.doi.org/10.1016/j.engappai.2026.114660, URL: https://www.sciencedirect.com/science/article/pii/S0952197626009425.

Haubrock, P.J., Turbelin, A.J., Cuthbert, R.N., Novoa, A., Taylor, N.G., Angulo, E., Ballesteros-Mejia, L., Bodey, T.W., Capinha, C., Diagne, C., Essl, F., Golivets, M., Kirichenko, N., Kourantidou, M., Leroy, B., Renault, D., Verbrugge, L., Courchamp, F., 2021. Economic costs of invasive alien species across europe. NeoBiota 67, 153–190. http://dx.doi.org/10.3897/neobiota.67.58196, URL: https://hal.science/hal-03329736.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Holmes, T.P., Aukema, J.E., Von Holle, B., Liebhold, A., Sills, E., 2009. Economic impacts of invasive species in forests. Ann. New York Acad. Sci. 1162, 18–38. http://dx.doi.org/10.1111/j.1749-6632.2009.04446.x, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2009.04446.x, _eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2009.04446.x.

Hulme, P.E., 2015. Invasion pathways at a crossroad: policy and research challenges for managing alien species introductions. J. Appl. Ecol. 52, 1418–1424, URL: https://www.jstor.org/stable/43869318.

Hussner, A., Heidbüchel, P., Coetzee, J., Gross, E.M., 2021. From introduction to nuisance growth: a review of traits of alien aquatic plants which contribute to their invasiveness. Hydrobiologia 848, 2119–2151. http://dx.doi.org/10.1007/s10750-020-04463-z.

iNaturalist, 2026. Inaturalist. URL: https://www.inaturalist.org/. (Accessed Jan 2026).

van Kleunen, M., Dawson, W., Essl, F., Pergl, J., Winter, M., Weber, E., Kreft, H., Weigelt, P., Kartesz, J., Nishino, M., Antonova, L.A., Barcelona, J.F., Cabezas, F.J., Cárdenas, D., Cárdenas-Toro, J., Castaño, N., Chacón, E., Chatelain, C., Ebel, A.L., Figueiredo, E., Fuentes, N., Groom, Q.J., Henderson, L., Inderjit, Kupriyanov, A., Masciadri, S., Meerman, J., Morozova, O., Moser, D., Nickrent, D.L., Patzelt, A., Pelser, P.B., Baptiste, M.P., Poopath, M., Schulze, M., Seebens, H., Shu, W.s., Thomas, J., Velayos, M., Wieringa, J.J., Pyšek, P., 2015. Global exchange and accumulation of non-native plants. Nature 525, 100–103. http://dx.doi.org/10.1038/nature14910, URL: https://www.nature.com/articles/nature14910.

Mathakutha, R., Steyn, C., le Roux, P.C., Blom, I.J., Chown, S.L., Daru, B.H., Ripley, B.S., Louw, A., Greve, M., 2019. Invasive species differ in key functional traits from native and non-invasive alien plant species. J. Veg. Sci. 30, 994–1006. http://dx.doi.org/10.1111/jvs.12772, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jvs.12772, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jvs.12772.

Mattingly, K.Z., Braasch, B.N., Hovick, S.M., 2023. Greater flowering and response to flooding in lythrum virgatum than L. salicaria (purple loosestrife). AoB PLANTS 15, lad009. http://dx.doi.org/10.1093/aobpla/plad009, URL: https://doi.org/10.1093/aobpla/plad009.

McInnes, L., Healy, J., Melville, J., 2018. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Muth, N.Z., Pigliucci, M., 2006. Traits of invasives reconsidered: phenotypic comparisons of introduced invasive and introduced noninvasive plant species within two closely related clades. Am. J. Bot. 93, 188–196. http://dx.doi.org/10.3732/ajb.93.2.188, URL: https://bsapubs.onlinelibrary.wiley.com/doi/10.3732/ajb.93.2.188.

Nahian, A., Jaber, M., et al., 2025. Agri-fm+: A self-supervised foundation model for agricultural vision. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5511–5523.

New Zealand Plant Conservation Network, 2026. New zealand plant conservation network. URL: https://www.nzpcn.org.nz/. (Accessed Jan 2026).

Nicotra, A.B., Leigh, A., Boyce, C.K., Jones, C.S., Niklas, K.J., Royer, D.L., Tsukaya, H., 2011. The evolution and functional significance of leaf shape in the angiosperms. Funct. Plant Biol. 38, 535–552. http://dx.doi.org/10.1071/FP11057.

Palma, E., Vesk, P.A., White, M., Baumgartner, J.B., Catford, J.A., 2021. Plant functional traits reflect different dimensions of species invasiveness. Ecology 102, e03317. http://dx.doi.org/10.1002/ecy.3317, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.3317, _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.3317.

Plants of the World Online, 2026. Lythrum L. | plants of the world online | Kew science. URL: http://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:30002863-2. (Accessed Jan 2026).

Pyšek, P., Richardson, D.M., 2007. Traits associated with invasiveness in alien plants: Where do we stand?. In: Nentwig, W. (Ed.), Biological Invasions. Springer, Berlin, Heidelberg, pp. 97–125. http://dx.doi.org/10.1007/978-3-540-36920-2_7.

Reinhardt, D., Gola, E.M., 2022. Law and order in plants – the origin and functional relevance of phyllotaxis. Trends Plant Sci. 27, 1017–1032. http://dx.doi.org/10.1016/j.tplants.2022.04.005, URL: https://www.cell.com/trends/plant-science/abstract/S1360-1385(22)00126-1.

Seebens, H., Blackburn, T.M., Dyer, E.E., Genovesi, P., Hulme, P.E., Jeschke, J.M., Pagad, S., Pyšek, P., van Kleunen, M., Winter, M., Ansong, M., Arianoutsou, M., Bacher, S., Blasius, B., Brockerhoff, E.G., Brundu, G., Capinha, C., Causton, C.E., Celesti-Grapow, L., Dawson, W., Dullinger, S., Economo, E.P., Fuentes, N., Guénard, B., Jäger, H., Kartesz, J., Kenis, M., Kühn, I., Lenzner, B., Liebhold, A.M., Mosena, A., Moser, D., Nentwig, W., Nishino, M., Pearman, D., Pergl, J., Rabitsch, W., Rojas-Sandoval, J., Roques, A., Rorke, S., Rossinelli, S., Roy, H.E., Scalera, R., Schindler, S., Štajerová, K., Tokarska-Guzik, B., Walker, K., Ward, D.F., Yamanaka, T., Essl, F., 2018. Global rise in emerging alien species results from increased accessibility of new source pools. Proc. Natl. Acad. Sci. 115, E2264–E2273. http://dx.doi.org/10.1073/pnas.1719429115, URL: https://www.pnas.org/doi/abs/10.1073/pnas.1719429115.

Seebens, H., Blackburn, T.M., Dyer, E.E., Genovesi, P., Hulme, P.E., Jeschke, J.M., Pagad, S., Pyšek, P., Winter, M., Arianoutsou, M., Bacher, S., Blasius, B., Brundu, G., Capinha, C., Celesti-Grapow, L., Dawson, W., Dullinger, S., Fuentes, N., Jäger, H., Kartesz, J., Kenis, M., Kreft, H., Kühn, I., Lenzner, B., Liebhold, A., Mosena, A., Moser, D., Nishino, M., Pearman, D., Pergl, J., Rabitsch, W., Rojas-Sandoval, J., Roques, A., Rorke, S., Rossinelli, S., Roy, H.E., Scalera, R., Schindler, S., Štajerová, K., Tokarska-Guzik, B., van Kleunen, M., Walker, K., Weigelt, P., Yamanaka, T., Essl, F., 2017. No saturation in the accumulation of alien species worldwide. Nat. Commun. 8, 14435. http://dx.doi.org/10.1038/ncomms14435, URL: https://www.nature.com/articles/ncomms14435.

Shackleton, R.T., Shackleton, C.M., Kull, C.A., 2019. The role of invasive alien species in shaping local livelihoods and human well-being: A review. J. Environ. Manag. 229, 145–157. http://dx.doi.org/10.1016/j.jenvman.2018.05.007, URL: https://www.sciencedirect.com/science/article/pii/S0301479718305267.

Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. http://dx.doi.org/10.48550/arXiv.1703.01365, URL: http://arxiv.org/abs/1703.01365, arXiv:1703.01365 [cs].

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. http://dx.doi.org/10.48550/arXiv.1808.01974, URL: http://arxiv.org/abs/1808.01974, arXiv:1808.01974 [cs].

Teng, D., Li, F., Zhang, W., 2021. Using comprehensive machine-learning models to classify complex morphological characters. Ecol. Evol. 11, 10421–10431.

Van Kleunen, M., Weber, E., Fischer, M., 2010. A meta-analysis of trait differences between invasive and non-invasive plant species. Ecol. Lett. 13, 235–245. http://dx.doi.org/10.1111/j.1461-0248.2009.01418.x, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2009.01418.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2009.01418.x.

Vieira, M., Zetter, R., Coiro, M., Grímsson, F., 2022. Pliocene *lythrum* (loosestrife, lythraceae) pollen from Portugal and the Neogene establishment of European lineages. Rev. Palaeobot. Palynol. 296, 104548. http://dx.doi.org/10.1016/j.revpalbo.2021.104548, URL: https://www.sciencedirect.com/science/article/pii/S003466672100172X.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M.L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J., Villar, R., 2004. The worldwide leaf economics spectrum. Nature 428, 821–827. http://dx.doi.org/10.1038/nature02403, URL: https://www.nature.com/articles/nature02403.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?. http://dx.doi.org/10.48550/arXiv.1411.1792, URL: http://arxiv.org/abs/1411.1792, arXiv:1411.1792 [cs].