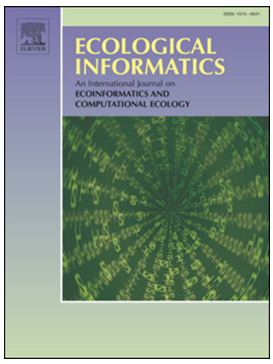

SALMA: A machine learning tool for precise leaf morphology measurements

Ilya Shabanov, Julie Deslippe, Andrew Lensen

PII: S1574-9541(25)00601-6

DOI: https://doi.org/10.1016/j.ecoinf.2025.103592

Reference: ECOINF 103592

To appear in: *Ecological Informatics*

Received date: 10 August 2025

Revised date: 29 December 2025

Accepted date: 29 December 2025

Please cite this article as: I. Shabanov, J. Deslippe and A. Lensen, SALMA: A machine learning tool for precise leaf morphology measurements, *Ecological Informatics* (2024), https://doi.org/10.1016/j.ecoinf.2025.103592

SALMA: A Machine Learning Tool for Precise Leaf Morphology Measurements

Ilya Shabanov[1], Julie Deslippe[1], Andrew Lensen[2],

1. Victoria University of Wellington, School of Biological Sciences, Wellington, New Zealand
2. Victoria University of Wellington, School of Engineering and Computer Science, Wellington, New Zealand

Abstract

Leaf area is a critical plant functional trait, widely used for understanding plant responses to climate change, ecosystem productivity, and species' adaptive strategies. Inaccurate leaf area measurements compromise the accuracy of other traits normalised by area, such as foliar chemical traits, respiration, and photosynthesis. However, existing measurement methods are ineffective for small-leaved plants and often necessitate manual processing, which limits sample sizes and potentially obscures subtle trait-environment relationships. We developed SALMA (Semi-Automated Leaf Morphological Analysis), which employs logistic regression trained on one to four human-

generated examples per species to delineate leaf boundaries for that species accurately. SALMA's training step adapts to species-specific features by integrating multiple characteristics, such as colour variations and edge details. The approach is validated on an extensive dataset (64 species, 3332 images) that covers 91.4% of the worldwide leaf area variation, as well as two smaller datasets comprising low-quality photographs of morphologically complex or damaged leaves. SALMA consistently achieved leaf area errors 2 to 15 times lower than existing algorithms and a theoretical upper bound of any grayscale intensity-based method. Critically, we identify a previously overlooked power-law relationship between leaf area and measurement error, demonstrating that existing methods may overestimate leaf area by at least 5% for 43% of global species, whereas SALMA achieves comparable errors for only 2.1% of species. SALMA is a standalone software with an intuitive interface that supports parallel processing, making it accessible for large-scale ecological studies globally. It can potentially enhance the quality of trait datasets and facilitate large-scale sampling, thereby advancing our understanding of plant-environment interactions. Our published dataset contains manually created binary segmentations of leaves and background, providing a baseline for future leaf measurement algorithms.

1. Introduction

Plant functional traits are central to predicting community assembly and ecosystem function (Dubuis et al., 2013; Funk et al., 2017; Lavorel & Garnier, 2002); among them, leaf area metrics consistently emerge as important traits. Of particular importance is specific leaf area (SLA, i.e. leaf area divided by weight) as the most studied leaf trait relevant to predicting responses to climate change (Green et al., 2022; Kühn et al., 2021). SLA describes how a plant invests resources for growth and survival, making it a key trait in balancing resource use, longevity and resilience (Díaz et

al., 2016; Wright et al., 2004). As a measure of carbon investment per unit of light capture, SLA is linked to key ecological processes, including demographic rates (Swenson et al., 2020; Wright et al., 2004), above-ground productivity (Violle et al., 2007) and leaf life span (Osnas et al., 2013). Leaf area is preferable to mass for normalising nitrogen, phosphorus, dark respiration, and photosynthesis rates measurements because these traits are linked to a leaf's ability to capture light and CO_2 , while mass-based normalisation distorts these relationships due to differences in leaf structure (Osnas et al., 2013). Errors occurring in leaf area measurements therefore compromise the values of many important traits, biasing and limiting our ability to uncover trait–environment relationships and ecological trade-offs.

Precise measurements are particularly challenging for small-leaved species typical of alpine, heathland, and arid ecosystems, where strong abiotic filtering favours plants with small, thick leaves adapted to high radiation, low temperatures, and drought (Parkhurst & Loucks, 1972). These species are highly sensitive to climatic changes and often serve as important indicators of environmental responses (Fagúndez, 2013; Grabherr et al., 2010; Prager et al., 2022). Their prevalence across resource-poor environments makes small leaves a key functional adaptation within the global leaf-economics spectrum (Wright et al., 2004). However, global databases (e.g. TRY (Kattge et al., 2020)) are biased towards plants with higher SLA (and thus larger leaf area) (Sandel et al., 2015), likely due to the size and morphological complexity of small-leaved plants. This leads to a bias in our understanding of global trait distributions and trade-offs towards larger-leaved temperate and tropical species. Accurate leaf area data for small-leaved plants are therefore crucial to better represent resource-poor ecosystems and promote understanding of global trait–environment relationships.

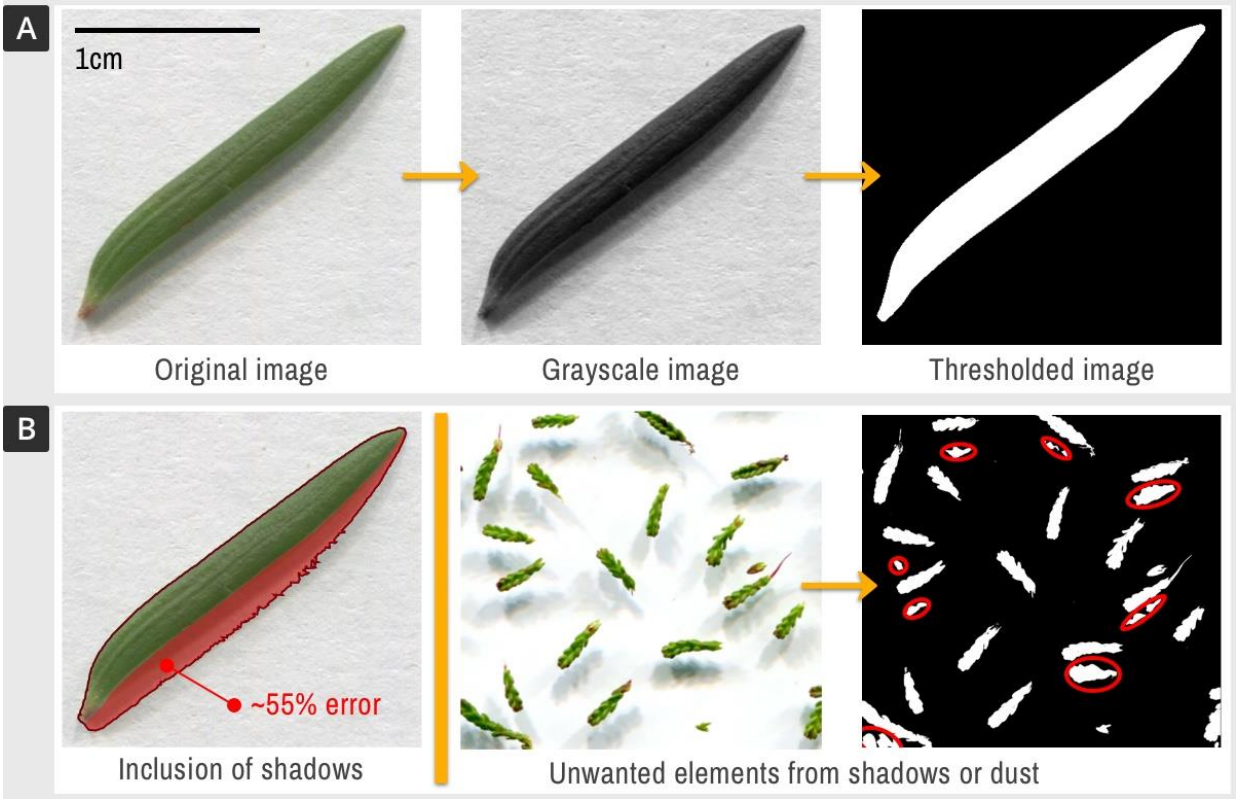

Figure 1: *Leaf area detection and some of its problems. A: Typically, the process involves converting an image to a grayscale and thresholding at a specific (often automatically determined) value to separate the background from the leaf. The resulting outline or segmentation can then be automatically measured. B: The most common error is the inclusion of shadows in the foreground, which leads to an overestimation of leaf area, especially for smaller leaves (left). When dealing with many fragments, this method necessitates laborious manual screening for unwanted particles resulting from shadows or dust if they can't be automatically distinguished by size (right).*

Most commonly, leaf area is measured by capturing digital images on a flatbed scanner or camera, converting them to a grey-scale intensity image and then setting a threshold value that separates the background from leaf pixels, resulting in a binary segmentation image or mask (Pérez-Harguindeguy et al., 2013) (Fig. 1A). Small leaves, in particular, pose a challenge as they produce

large shadows relative to the leaf area (Fig. 1 B left) that are visually similar to the leaf and are indistinguishable in shape from plant matter after thresholding (Fig. 1 B right). This process can be done manually using image processing software (e.g. ImageJ (Schneider et al., 2012)) or through automated software such as LeafByte (Getman-Pickering et al., 2020), LeafArea R Package (Katabuchi, 2015), USPLeaf (Meira et al., 2020) or FAMeLeS, an ImageJ macro, (Montès et al., 2024) (See supplementary materials in (Montès et al., 2024) for a comprehensive list of methods). However, most algorithms use simple automated thresholding methods like "Otsu's threshold" (Otsu, 1979) (USPLeaf and LeafByte) or "Minimal threshold" (Prewitt & Mendelsohn, 1966) (LeafArea), or set some fixed threshold after preprocessing (FAMeLeS), making them functionally very similar to one another and neglecting the diversity of shape and colour among species of the global flora. Exceptions include Leaf-IT (Schrader et al., 2017), which traces one continuous line around leaves using image gradients and logical rules, as well as Easy Leaf Area (Easlon & Bloom, 2014), which sets thresholds on the ratio of red/green channels. Rigorous tests and precision comparisons across methods are largely absent (e.g., 25 leaves from 18 species were used to validate Leaf-IT, limiting assessment of intraspecific variation). Ultimately, this leads to unexpected errors when ecologists use automated methods (e.g. Fig. 1 B left generated with Otsu's threshold), which erodes trust in the methods and may lead users to revert to manual leaf area measurements, negating efficiency gains.

Here, we present SALMA, a semi-automated algorithm that detects leaf segmentations and extracts morphological features (leaf area, axes length, perimeter, solidity, eccentricity) from standard flatbed scanner images or photographs. SALMA learns from human-generated segmentations of as little as one leaf per species (few-shot learning), allowing it to identify the most relevant features specific to each species, rather than using a uniform approach across all species. Firstly, we aim to

make SALMA reliable for tiny leaves commonly found in alpine floras. Secondly, we investigate the relationship between leaf area and error rate to assess which fraction of the global flora can be reliably evaluated using existing methods. Finally, we publish the dataset used in this study, including human-generated segmentations for 3332 leaves in 64 species, to encourage further development and rigorous comparison of related algorithms. SALMA is freely available, interactive, requires no technical expertise, can batch-process images, incorporate manual corrections, and export the results to a spreadsheet.

2 Methods

Our method, SALMA (Semi-Automated Leaf Morphological Analysis), uses logistic regression inside a multi-dimensional colour and colour-gradient space to learn the boundary between background and foreground (i.e. leaf) pixels in an image (Fig. 2). For training, one or more examples of leaves and their binarised background/foreground segmentation need to be provided by a human. To deal with colour variation, training can be improved iteratively, whereby the worst segmented leaves (presumably those where the colour variation is most substantial) are added to the training set to retrain and improve the model in every step. If images are provided as photographs at varying lighting conditions, the background can be adjusted using an automated white-balance correction (see section 2.2 for an example). This is generally unnecessary when working with images generated on a flatbed scanner.

After training, the model automatically processes any number of images of the same species taken under similar light conditions. The resulting binary segmentations are used for morphological measurements such as leaf area, perimeter length, eccentricity, length or width. The segmentations

can also be analysed outside of the SALMA pipeline to extract other morphological information of interest.

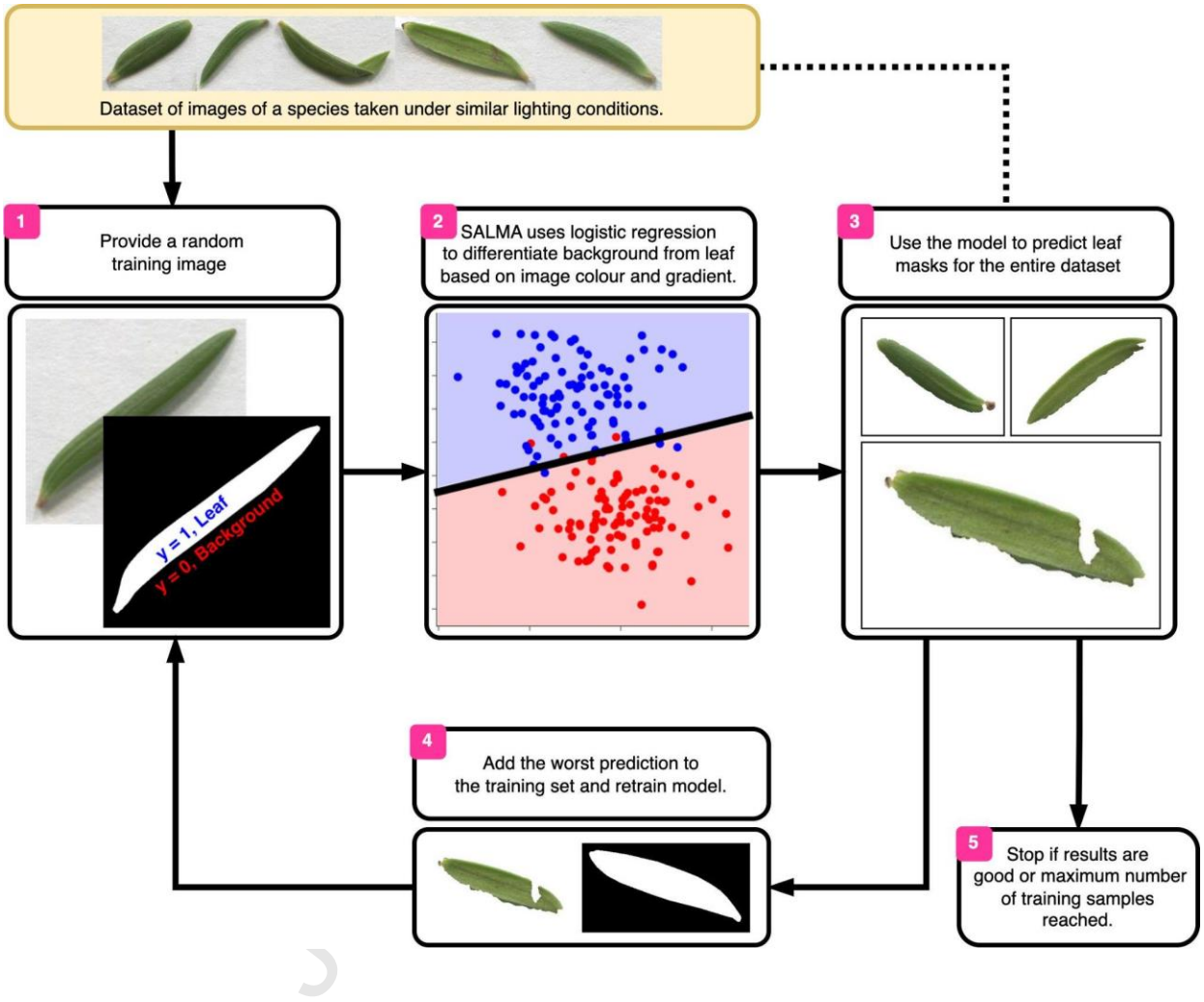

Figure 2: *Schematic overview of SALMAs function. Step 1: A random leaf and a human-generated segmentation separating the leaf from the background is provided for training. Step 2: The SALMA model is trained to separate the leaf and background using the colours of the pixels and other information derived from them. Step 3: The SALMA model predicts all leaves in the dataset. Step 4: A human picks a leaf with a large error and adds it to the training set to repeat the training cycle. Step 5: If no large errors are present, the process is complete, and the binary leaf-background segmentations can be used to calculate leaf area. (While the illustration refers to single leaves, in practice, an image can contain multiple leaves or fragments.)*

2.1 Main Dataset

Our dataset consisted of 3332 healthy, intact leaf scans (at 600 dpi) from 64 species native to New Zealand. These species' leaves spanned 4.6 to 14202 mm^2 in area, covering nearly the full global spectrum, from the 0.86^{th} percentile (smallest leaves) to the 92.3^{rd} percentile (largest leaves) (Díaz et al., 2022). Detailed information on species, counts, raw images and human-generated outlines are available at (Shabanov et al., 2025)**.**

2.2 Low-Contrast Dataset

To explore how the quality of images affects SALMA's performance, we evaluated an additional dataset (not included in the main dataset) of 130 leaves from four alpine species obtained during fieldwork (**Fig. S3**). This dataset was technically motivated rather than ecologically motivated, and we used it to quantify the effect of image quality on SALMA's accuracy. The dataset consisted of digital photographs under various adverse lighting conditions (e.g., blurriness, varying contrast, blue or red colouration). The species spanned 31-147 mm^2 in area, situating them below the 20th percentile of

the full global spectrum (Díaz et al., 2022). The white balance of all images was manually corrected by selecting a piece of background and using an ImageJ Macro (Mascalchi, 2016), restoring quality loss primarily due to low contrast or colour tints (**Fig. S3**). The training and evaluation procedures were otherwise the same as described above.

2.3 Diverse Morphology and Damage Dataset

We assembled a third dataset to validate SALMA on 212 leaves from six species containing damage from dessication and herbivory, compound leaves, narrow-leaved grasses and leaf fragments (Examples and species list in Fig. 6C and Fig. 7). Similarly to the low-contrast image dataset, the goal of this dataset was technical, as we aimed to explore how leaf shape and condition influenced segmentation accuracy. The dataset was assembled from fieldwork in New Zealand, Argentina and an openly available dataset of old-growth tropical forest in Borneo (Both et al., 2020). The latter classified scans in “easy” and “tough” depending on the difficulty of leaf area analysis. Our subset contained only scans in the “tough” category. The training and evaluation procedures were otherwise identical to those described above.

2.4 Model Training

Models for each species are trained by providing RGB images and corresponding binary segmentations separating the background and foreground (i.e. leaf area) pixels. We first translate the RGB image into two additional colourspaces: HSV (Hue–Saturation–Value), separating chromatic content from brightness, and LAB (Lightness–a–b) representing colour in a perceptually more uniform way through a lightness and two chromatic channels. This leads to more robust segmentation. For example, LAB’s lightness channel is well-suited for distinguishing pale leaves,

from darker backgrounds, while RGB channels better capture the strong green saturation of tropical evergreens.

1. SALMA uses HSV and LAB colour spaces alongside RGB to adapt to species-specific variations in leaf colours. The hue channel (in HSV) is excluded due to its circular structure (0° - 360°), which distorts distance measurements. This produces eight channels (2 from HSV, 2x3 from LAB and RGB). Each channel is normalised to 0-1 by dividing by the maximal value (e.g. 255 for the red channel). (see **Fig. S2** B for examples)
2. The Laplace filter (Jain et al., 1995) is applied to each of the eight channels to detect areas where the value changes (i.e. colour gradients). We used the fastest Laplacian kernel (3x3), creating eight additional channels; see **Fig. S2** B for examples.
3. The full training set consists of N pixels from one or more training images $X = \{x_i \in \mathbb{R}^{16} | i = 1 \dots N\}$ and training labels $Y = \{y_i \in \{0,1\} | i = 1 \dots N\}$. Because N can be very large (e.g. N >> 10^6 for a single image) and negatively impact performance, we randomly subsample it to $N' = 1000$, whereby an equal number of leaf and background pixels is included. This step mitigates computational overhead by limiting the training set size without impacting training performance (**Fig. S1**). When multiple images are used as inputs $K > 1$, we subsample each image individually, increasing the training set size to $N' \cdot K$ to ensure equal contribution of each training image.
4. Finally, we fit a generalised linear model (binomial family with a logit link, no random effects) without interaction terms. A regularisation constant parameter C was used to prevent overfitting. The best value for C is found by training models for the values 1, 10^2, 10^3, 10^4, 5 * 10^4, 10^5, 5 * 10^5,10^6 as recommended by the employed scikit-learn Python library (Pedregosa

et al., 2011) through 5-fold cross-validation. The best hyperparameter values are then used to train the final model on the entire dataset.

5. As the leaves in our dataset are all healthy, small holes below 10% of the detected area were removed after prediction. This parameter is adjustable in the interactive version and can be iteratively tuned to suit the dataset (Fig. 7). For example, it allows the user to assess the area of leaf damage by herbivory or pathogen infection.

2.5 Comparison with existing methods

We compared our method with three representative algorithms (Table 1), some implemented in software designed to make standard segmentation methods easier to use. For instance, LeafByte (Getman-Pickering et al., 2020), is a user-friendly mobile phone application that applies Otsu's threshold method (Otsu, 1979), which determines its segmentation performance. The optimal threshold mimics the semi-automated ImageJ workflow (Schneider et al., 2012), in which users adjust threshold values to capture the leaf outline rather than tracing it manually. Unlike the other methods, the optimal threshold is not automated, as it requires human participation. Its performance, therefore, acts as an upper bound for any grayscale threshold-based method (e.g. USPLeaf, LeafByte, and LeafArea).

Table 1: The three algorithms SALMA is compared to, the rationale behind including them and their working principle.

Method	Reason for Inclusion	Working Principle

Otsu's Threshold	Widely used fully automated method for segmentation; Internally used by USPLeaf and LeafByte.	Automatically determines an optimal global threshold by minimising intra-class variance between leaf and background pixel intensities in a grayscale image.
FAMeLeS	A recently published method, reported to outperform existing approaches for small leaves.	Uses a multi-step combination of automated thresholding on colour channels and bandpass filters to segment background from leaf pixels.
Optimal Manual Threshold	Serves as an upper benchmark for all threshold-based methods (USPLeaf, LeafByte, and LeafArea) by simulating manual threshold setting and providing a proxy for a human using grayscale thresholds.	Converts images to grayscale and finds the threshold that maximises the similarity between the resulting segmentation and human-generated masks (applying the Jaccard index). This simulates a human manually setting the threshold in software like ImageJ rather than being a true "model".

To determine quantitative error measures, we compared the binary segmentations produced by the algorithms with the human-made ones and computed:

1. Precision: The proportion of algorithm-predicted leaf pixels correctly matching human-labelled leaf pixels.
2. Recall: The proportion of human-labelled positive pixels correctly identified by the algorithm.

3. Errors of perimeter length, area, minor and major axes of algorithm-predicted to human segmentations. The errors are defined as $E = |100 - \frac{100 \cdot v_{alg}}{v_{human}}|$ for v_{alg} and v_{human} being the values determined by the algorithm or human, respectively. The error is a percentage starting at 0% and growing with deviation from the human assessment. Minor and major axes are effectively the length and width of the leaf, computed from the object's pixel covariance, representing its shortest and longest spans.

As SALMA requires training data, while other methods are fully automated, we used the following training schedule:

1. The SALMA-1 model is trained on a random leaf image for the species. (The numeral refers to the number of leaves used for training).
2. SALMA-1 predicts leaf segmentations for all images of a species, and the leaf with the highest area error is added to the training set, which can be considered a form of boosting (Freund & Schapire, 1997), known to increase the learning ability of classifiers with a performance above random.
3. SALMA-2 is now trained on the extended dataset containing two leaves. This simulates the strategy SALMA users would likely pursue, e.g., train the simplest model (using only one leaf) and add more training data if the model makes mistakes.
4. Up to two more leaves are added to produce more refined models (SALMA-3 and SALMA-4).
5. The best-performing model of the four is selected as the final SALMA model for this species.

2.6 Global Dataset Comparison

We used trait data from (Díaz et al., 2022) that included mean leaf area values for 12164 species to assess which percentage of species globally have leaves below a specific area A, yielding a monotonously rising function $C(A) \in [0,100]$ (Fig. 4A, grey area). The relationship between the leaf area A and the area error R for each of the four methods was approximated, fitting a log-linear model $R = \alpha \cdot log_{10}(A) + \beta$ to our dataset's results from the 64 species. To estimate the applicability of each algorithm m, we combine these results to estimate what percentage of global species (based on their leaf area A) are expected to be assessed below a given area error R by a specific algorithm. This yields a monotonously rising function $D_m(R) = 100 - C(A) = 100 - C(10^{\frac{R-\beta}{\alpha}}) \in [0,100]$. For example, $D_m(10\%) = 45\%$ means that the algorithm m will classify approximately 45% of species globally below an area error of 10%.

3 Results

3.1 Method performance

We evaluated by how much the area predicted by the methods differed from the human-delineated area using the area error metric (Fig. 3 A). SALMA showed superior performance for all species, with the worst median area error of 9.6% for the small-leaved angiosperm species (*Prumnopitys taxifolia*). In contrast, the error rates of the other algorithms were 38.1% (FAMeLeS), 29.1% (Otsu), and 23.3% (Optimal).

Across all species, SALMA needed an average of 3.2 leaves for training (Average of bars in Fig. 3 B top). More training images decreased the error across species from 6.5% to 2.4% and the variance of the error from σ=9.2% to 2.4% (Fig. 3 B bottom). Lastly, we compared the different error measurements across all algorithms (Fig. 3C). While recall was comparable for all algorithms, SALMA's precision was greater than other algorithms by ~0.1. The largest errors were generated by estimates of leaf perimeter and area. The perimeter error was higher than the area error for all algorithms but FAMeLeS. Minor axis errors were higher than major axis errors for all algorithms.

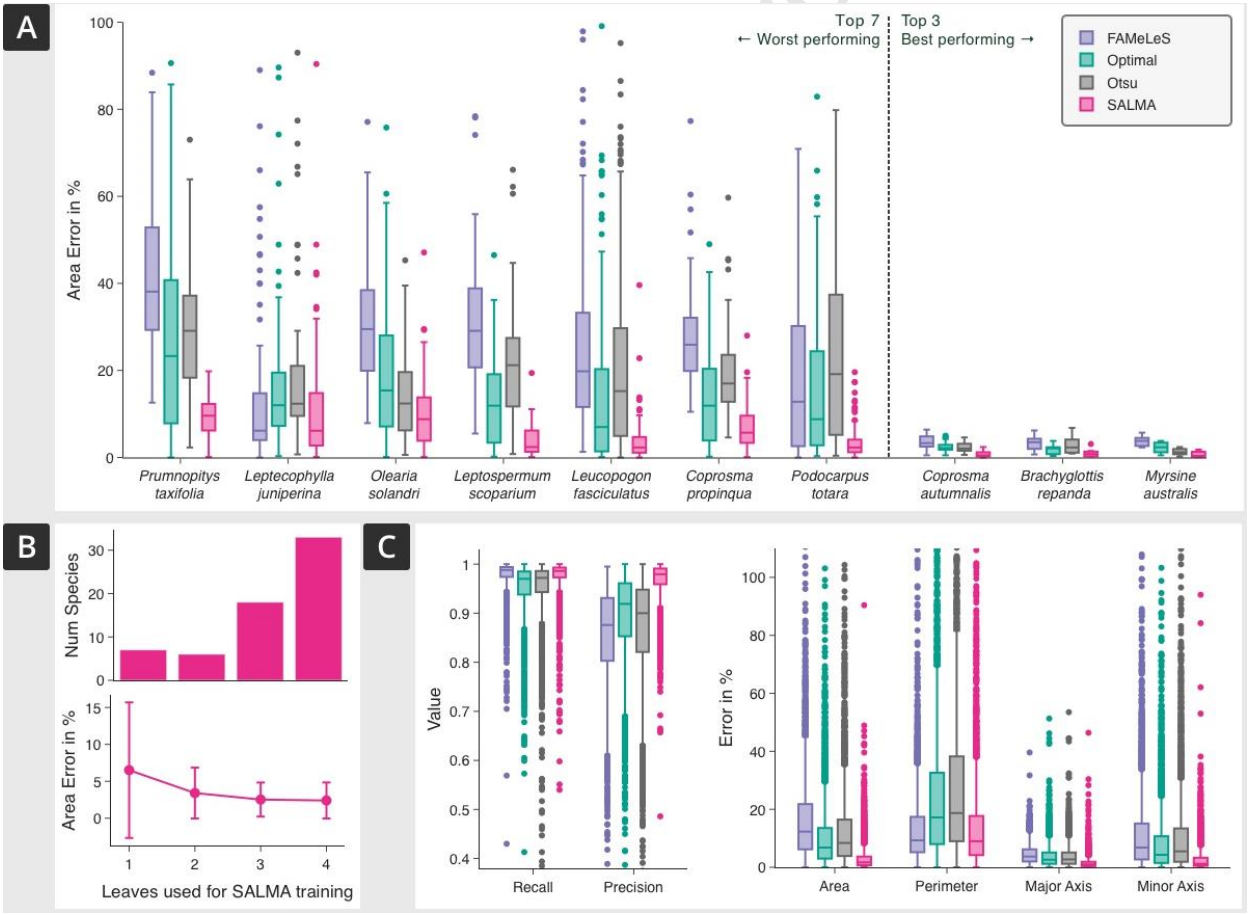

Figure 3: *Performance comparison of different methods. A: Area error across species sorted by the mean. Only the worst 7 and best 3 species are shown for simplicity. B: Number of leaves used to train the SALMA models used in panel A and the average area error (± one standard deviation) of models trained with the same number of leaves (bottom). C: Comparison of different error metrics for the four methods.*

3.2 Leaf area and error relationship

We assessed the relationship between a species' average leaf area and the area error generated by the four methods for each species (markers in Fig. 4A). For comparison, this relationship was approximated with a log-linear function with the goodness of fit (R^2 value) between 0.45 and 0.73 (Fig. 4A, lines), indicating a power-law relationship between area error and leaf area regardless of the measurement method employed. We plotted the fraction of species globally below a certain leaf size for 12164 species in (Díaz et al., 2022) (Fig. 4A, shaded area, secondary y-axis). Using the global leaf area distribution and the log-linear area-error relationship for the different methods, we assessed what portion of species globally (based on their leaf area) can be assessed given a maximum acceptable error rate (Fig. 4 B). For example, at an acceptable area error threshold of 5%, SALMA is expected to produce reliable results for 97.9% of global species, while the other methods would produce reliable results only for 47-56% of species. To ensure reliable measurements of the other methods for 95% of species, the acceptable area error threshold needs to be increased to 13%-21%.

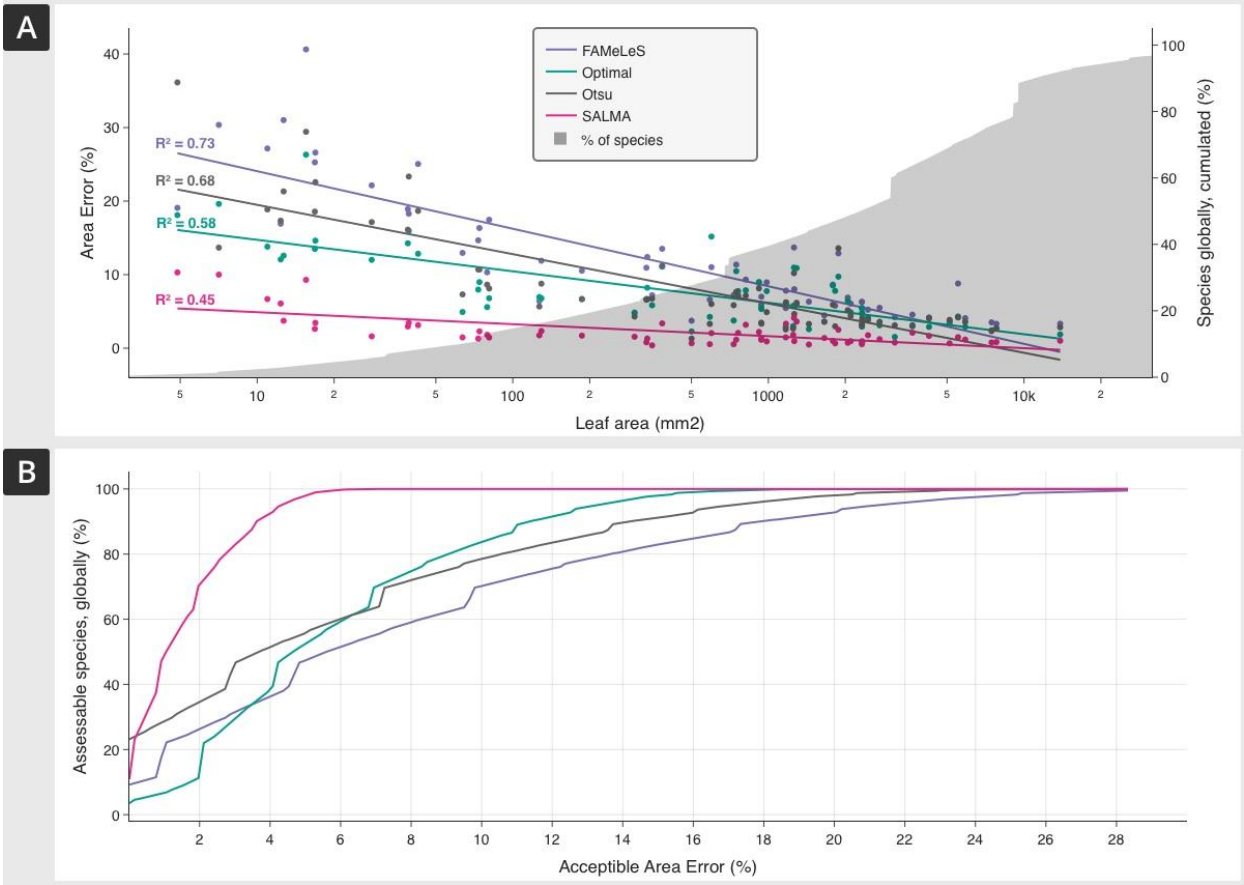
A
FAMeLeS
Optimal
Otsu
SALMA
% of species
R² = 0.73
R² = 0.68
R² = 0.58
R² = 0.45
Area Error (%)
Species globally, cumulated (%)
Leaf area (mm2)
B
Assessable species, globally (%)
Acceptible Area Error (%)

Figure 4: *Relationship between leaf area and area error. A: Area error for species in our dataset plotted against their average leaf area (log scale) for the different methods. Lines are log-linear fits, and goodness of fit (R^2 values) appear on the left. The shaded region illustrates the percentage of global species below a given leaf size. Steep increases in the number of species (right-hand side y-axis) likely reflect the logarithmic x-axis and uneven sampling of species across leaf area ranges. B: Estimated percentage of species globally that can be assessed below an acceptable area error. The estimate is generated using the log-linear approximation (lines in panel A) and the global distribution of leaf area (shaded region in panel A).*

3.3 Performance for low-contrast images

To explore the impact of image quality on the performance of the algorithms, we evaluated them on a separate low-quality dataset captured with a mobile phone in the field (example images in **Fig. S3**). The best average area error was achieved by the "Optimal" algorithm at 5.3%, followed by SALMA 8.1% (Fig. 5A). FAMeLeS and Otsu had very high errors of 116.7% and 32.4% due to the large number of leaves where the algorithm failed to identify the background correctly, leading it to categorise the entire image as background or leaf, and inflating the area error (Fig. 5C, top right). The number of training samples required for SALMA was 3 (Fig. 5B). Lastly, we compared the predicted errors for optimal lighting (from the log-linear model in Fig. 4A) with the measured errors for the low-contrast images. All algorithms, except the manual optimal threshold, showed higher area error rates: SALMA by 4.67%, Otsu by 21.59% and FAMeLeS by 98.24% (Fig. 5D).

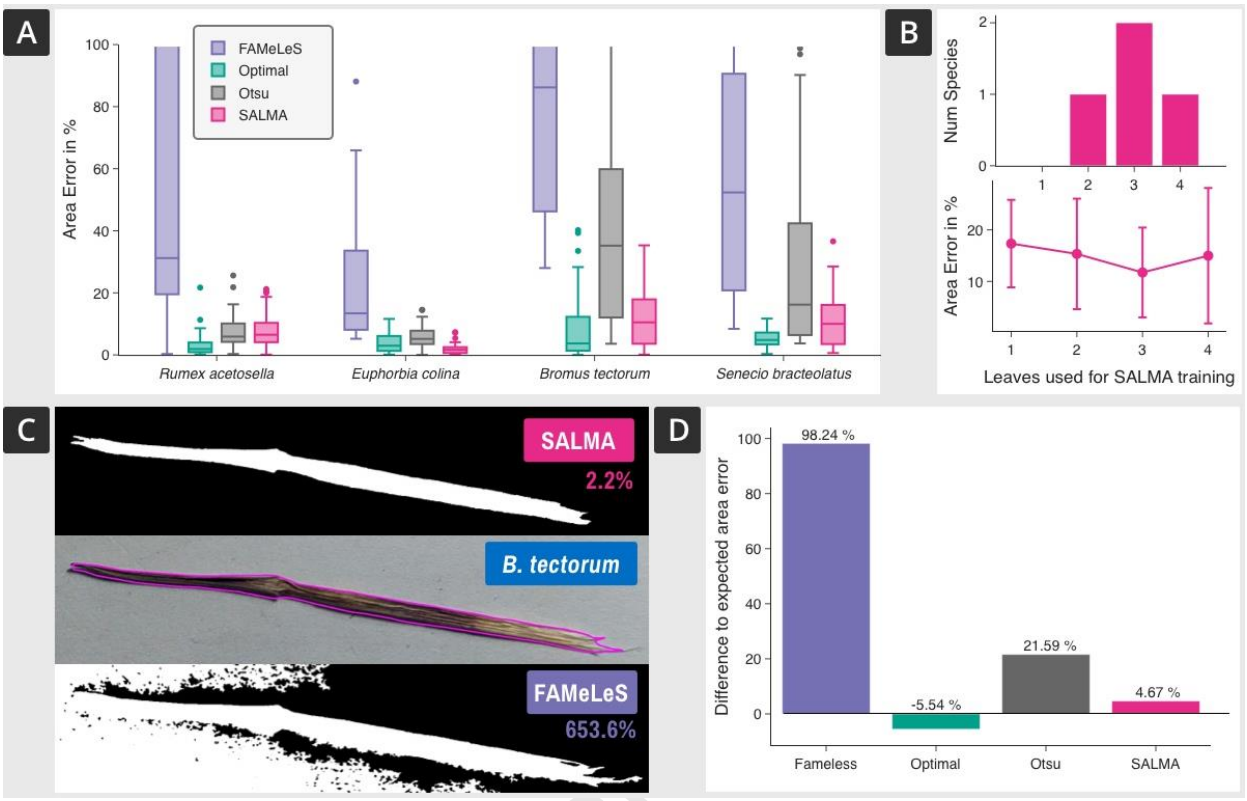

Figure 5: *Performance comparison of different methods on the low-contrast images. A: Area error across species. The error axis is limited to 100% for the legibility of the smaller values. B: Number of leaves used to train the SALMA models used in panel A and the average area error (± one standard deviation) of models trained on that number of leaves (bottom). C: Example leaf with human-delineated outline in pink and predictions made by the algorithms. The number refers to the area error. FAMeLeS has an inflated error, since most of the background is included in the prediction. D: Mean difference of the area error and expected error under optimal light conditions (see previous section). Values above 0% indicate a worse-than-expected performance.*

3.4 Performance for alternative leaf morphologies

Across four species in the alternative morphologies data set that included compound, damaged and fragmented leaves, all algorithms achieved excellent performance at a median area error below 3% (Fig. 6A, left). SALMA used two to three images for training (Fig. 6B) and outperformed the other algorithms. Two species led to diminished performance: The low contrast images of *Acaena splendes* and the thin grass *Rytidosperma setifolium,* with errors averaging 14.6% and 26%, respectively (Fig. 6A, right). *R. setifolium* was the only species where SALMA performed worse than other algorithms. However, the blades of *R. setifolium* were only a few pixels thick, and the human-generated outlines were imprecise (Fig. 6C, orange box).

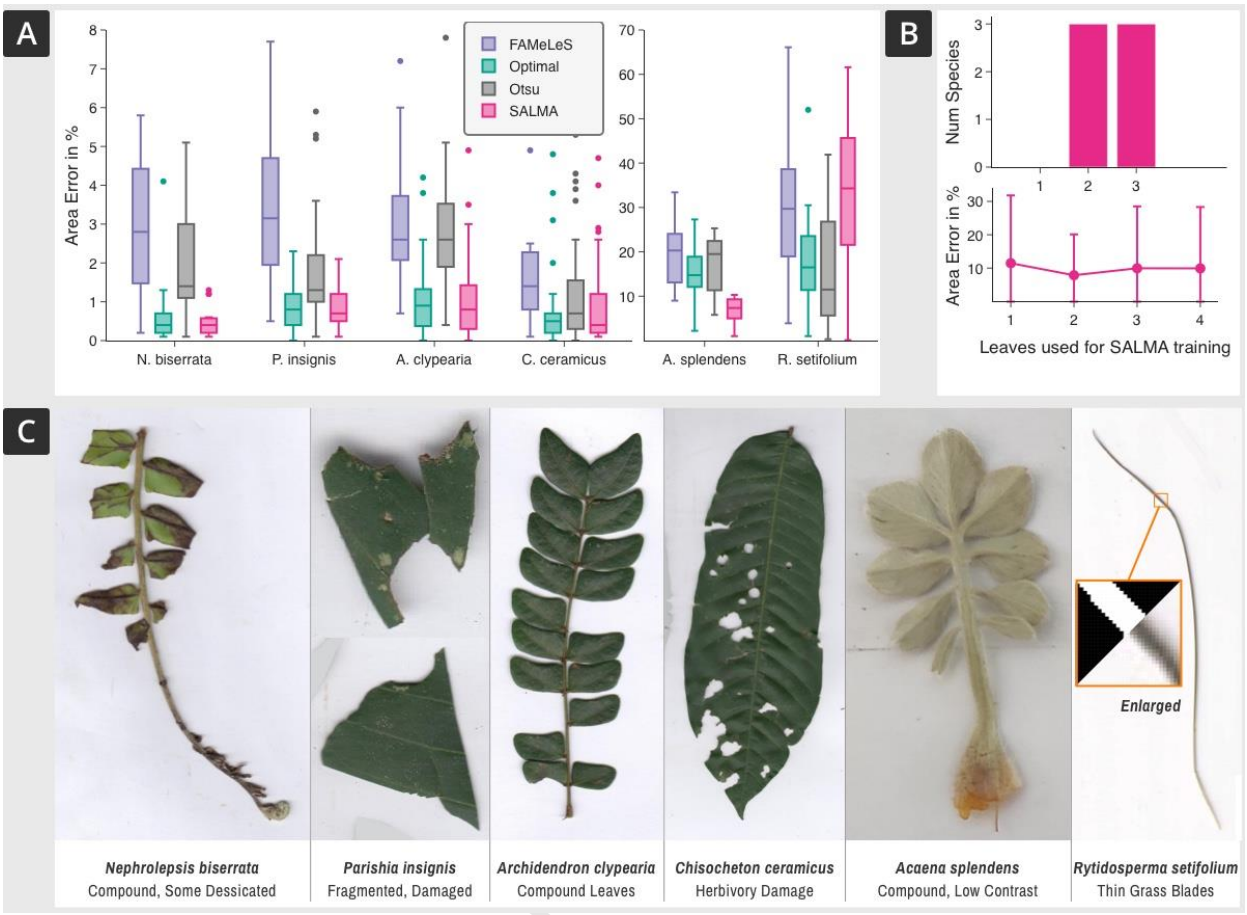

Figure 6: *Performance comparison of different methods on the alternative morphology images. A: Area error across species. B: Number of leaves used to train the SALMA models used in panel A and the average area error (± one standard deviation) of models trained on that number of leaves (bottom). C: Examples of leaves in the alternate morphology dataset. Species names are given at the top in bold, and the distinguishing morphological feature at the bottom. For the grass (R. setifolium) a small section of the image (lower corner) and manual segmentation (upper corner) are depicted to visualise a possible ambiguity of segmenting the blade, which is only a few pixels thick.*

Discussion

We developed SALMA, a semi-automated method for measuring morphological leaf traits, which is robust to poor image quality (Fig. 5A), diverse leaf morphologies (Fig. 6A) , but requires the manual input of a one to four leaves per species as training examples (Fig. 3B). We show that SALMA outperforms existing general methods like FAMeLeS (Montès et al., 2024) and intensity threshold-based methods (subsuming the methods USPLeaf (Meira et al., 2020), LeafByte (Getman-Pickering et al., 2020), LeafArea (Katabuchi, 2015)) (Fig. 3A). The performance gain is achieved through training a logistic regression model on one to four data samples (Fig. 3B), utilising colour and gradient information. The training step enables the model to identify and utilise the most relevant species-specific features, unlike current algorithms, which apply a uniform approach across all species. Additionally, instead of isolating a single dimension (e.g., intensity) to differentiate the leaf from the background, the regression model integrates multiple characteristics simultaneously, including colour variation and edge morphology (**Fig. S2**), leading to a more accurate binarisation of the image than even an optimal threshold (Fig. 3A). Since SALMA requires training for each species, we packaged the algorithm into a stand-alone, easy-to-use graphical user software, capable of parallel batch processing, applicable to dissected, healthy or damaged (e.g. having holes) leaves of any morphology. The graphical user interface has additional features not illustrated in this manuscript (e.g., shape smoothing, see point 4 in Fig. 7), allows for manual corrections by the users, and exports results to a spreadsheet of a specifiable format, making SALMA a compelling choice for leaf area measurements (Fig. 7).

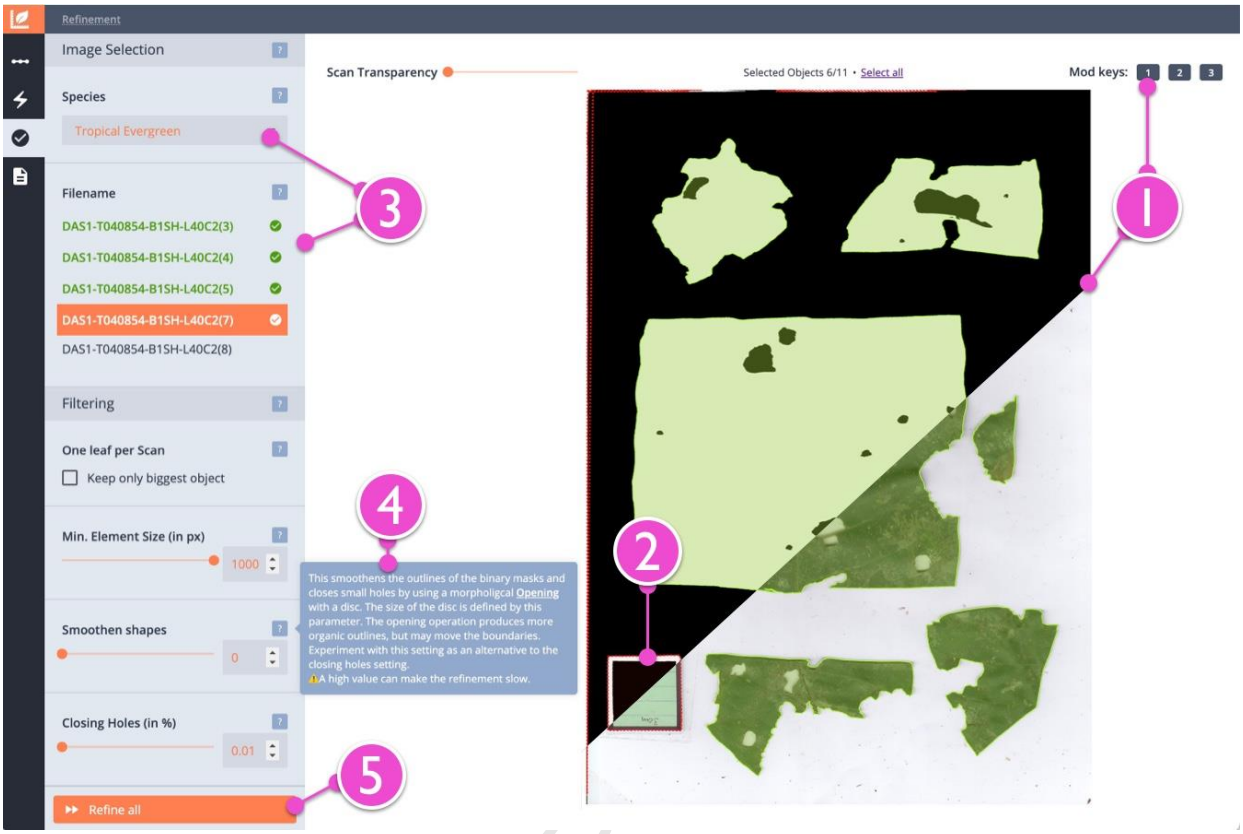

Figure 7: *Screenshot of the "refinement step" in the SALMA standalone software with some features highlighted. The step depicted occurs after a model has been trained and applied to a single image. 1: Users can switch between viewing the segmented image, the scan or the detected leaves/fragments. 2: Erroneously detected objects can be manually excluded by clicking on them. If the scan consists of dissected leaf fragments, their area can later be exported as a sum. 3: SALMA uses a working folder structure organised by species and filenames. Filenames can contain any information, e.g., sites, IDs, replicates etc., that is exported into separate columns of the final CSV file. 4. Each parameter in SALMA is explained through tooltips. Any changes to the settings are reflected in real-time on the screen. 5. A batch processing mode enables the simultaneous processing of all images using parallel processing and allows for the interactive exploration of the results.*

Most algorithms use thresholds on pre-processed grayscale versions of the original image to differentiate the background from the leaf. However, even an optimal threshold set individually for each image does not capture the leaf precisely and results in errors (Fig. 3A and C, example in Fig. 1B) that grow exponentially as leaf size decreases (Fig. 4A). This fundamental problem makes threshold-based methods unreliable for small-leaved plants (e.g. if only errors up to 5% are permitted, they are unreliable for 40% of global flora (Fig. 4B)). Even FAMeLeS (Montès et al., 2024), designed with small leaves in mind, on average, performed worse than Otsu's threshold (Fig. 3A and C) on our dataset. Alternatives to threshold-based methods have severe limitations: Leaf-IT (Schrader et al., 2017) lacks batch processing capabilities and can't handle complex leaf shapes, and Easy Leaf Area (Easlon & Bloom, 2014) is limited to predominantly green leaves due to its reliance on the green channel. Moreover, these methods were validated on fewer than 50 images (e.g., Easy Leaf Area used only a single species), limiting objective assessment of their performance. Collectively, these limitations compromise high-throughput leaf area measurements for large multispecies ecological datasets.

While deep learning methods show promise for solving image-based tasks, to our knowledge, no dedicated leaf area measurement tool exists. Popular segmentation models like YOLO (Redmon et al., 2015) or SAM (Kirillov et al., 2023) utilise 3 million (Yolov8n) to over 600 million (SAM) parameters, requiring thorough technical expertise and millions of images to fit the model. While, deep-learning approaches have been successfully used for plant segmentation, object recognition and counting (Bhagat et al., 2022), their pixel area measurement is not precise. For instance, a recent deep learning leaf outline model trained on relatively large Eucalyptus spp. leaves reported a precision of 0.77 (Guo et al., 2024). While this method solved the harder task of identifying overlapping leaves in herbarium scans, its precision was notably lower than SALMA's precision of >0.95 (Fig. 3C). This

might be due to the model architecture, which limits the inputs to a fixed size (e.g. 640x640 pixels for Yolo), requiring scaling or tiling images to work with large leaf scans (e.g. ~5000x7000 pixels for a scan at 600dpi), which possibly limits the precision gain from high-resolution scans. By contrast, SALMA is minimal, utilising only 17 parameters, which can be fitted on 1-4 leaves in seconds on a standard laptop without any technical expertise. Because single pixels are used as inputs, it applies to images of any size. Owing to their generic nature, current deep-learning models provide advantages in segmentation, object recognition and counting, but are not precise for area measurements.

Large datasets, assembled by automated measurements, have the potential to unlock fundamentally new insights, even if their precision is lower. For example, (Guo et al., 2024), gathered 136,599 leaf area measurements (from 38,867 herbarium sheets) of the Eucalyptus genus, which, despite the lower reported precision (0.77), enabled them to discover a link between leaf area and precipitation at the sub-genus but not species level, demonstrating the potential of large-scale trait measurements, aided by automation. In contrast, the average number of datasets containing leaf area measurements in the globally largest trait database (TRY) is 9.98 (Díaz et al., 2022), indicating only a handful of different sampling locations. The need for larger datasets also stems from phenotypic plasticity, variation in traits within a species in response to biotic and abiotic factors, which plays an important role in plant growth (da Silveira Pontes et al., 2010) and community assembly (Lepš et al., 2011). Larger sample sizes across environmental gradients and genotypes can significantly increase the statistical power of conducted studies (Gianoli & Valladares, 2012). This problem is exacerbated for alpine species where phenotypic plasticity can influence elevational distribution (Rixen et al., 2022). Because intraspecific trait variation is generally lower than variation among species (Albert et al., 2011), detecting it in small alpine plants requires highly precise

measurements. Automated and accurate leaf area measurements, therefore, have the potential to advance our understanding of trait-environment relationships.

SALMA's main limitation is the need for one or more human-generated examples for training. However, we found that the average error, even with only one example leaf, is very low (ca. 4%, Fig. 3B) and decreases insignificantly with more examples. This will likely depend on the species, with more complex leaves (e.g. ones with strong colour variation) requiring more examples. After training, models can be shared and reused, with additional examples added as needed to reduce the data requirements for future model training. SALMA's precision is worse than recall, which indicates that it errs on the side of overestimating leaf area through the inclusion of shadows (Fig. 3C). However, this effect is much less pronounced than for existing methods. Exploring species-invariant models, the impact of post- and pre-processing of images and adding further channels to improve performance will be part of future research.

Under suboptimal lighting conditions, SALMA demonstrated high resilience, achieving a mean area error of 8.1%—only slightly above the optimal threshold method (5.3%) but substantially outperforming Otsu (32.4%) and FAMeLeS (116.7%) (Fig. 5A). FAMeLeS' high area error results from erroneously identifying the background as leaf (Fig. 5B), leading to an error proportional to the ratio of leaf to background. Alternative error metrics do not rank the algorithms differently (**Fig. S4**). The adverse light conditions increased SALMA's area error by 4.67% (Fig. 5D), making it perform worse than the manual optimal threshold setting. This indicates that under optimal light conditions, colour and gradient information, exploited by SALMA, contain more information than the grayscale intensity used by other methods, leading to a better performance and indicating that an intensity

threshold is insufficient for accurate leaf segmentation. Poor lighting and contrast, affecting the colour and gradient information respectively, erode this additional information, making SALMA perform slightly worse than a human optimally setting the threshold. However, while the error increases, SALMA's segmentation is automated, and in cases where human involvement is infeasible due to the size of the dataset, it remains a viable choice.

Diverse leaf morphologies and leaf damage had an insubstantial effect on SALMA's performance and data requirements (Fig. 6A left and B). However, low scanning resolution (e.g. for *R. setifolium,* Fig. 6C) strongly increased the area error to 32%. This is likely caused by the thickness of the scanned grass blades being only 4 pixels thick, making a precise and consistent human segmentation difficult and resulting in training the model on ambiguous data (Fig. 6C orange box). In other words, if a pixel can't be uniquely (and consistently) identified as background or leaf, due to the combination of morphology and resolution, performance will worsen. However, SALMA is fundamentally invariant to leaf morphology, provided all data have the same resolution, because it is pixel-based. Using flatbed scanners with a minimum resolution of 400 dpi will improve performance, preventing errors caused by low contrast, insufficient resolution or a higher shadow-to-leaf-area ratio of morphologically complex leaves.

SALMA is a fundamentally novel approach for ecologists interested in precise, automated leaf area measurement in the context of plant functional traits or herbivory. We validated SALMA's superior performance and broad applicability on three diverse datasets, including temperate to tropical leaves of global flora, including woody and grassy species of various morphologies, damage levels

and under poor lighting conditions. This increase in accuracy has the potential to enhance our ability to identify novel trait-environment relationships and elucidate the ecological role of phenotypic plasticity. Importantly, we demonstrated that measurement errors decrease exponentially with leaf size, making the increased precision necessary when working with small-leaved plants (e.g., alpine regions). The method is provided as easy-to-use browser-based software, accompanied by video tutorials at https://github.com/lnilya/salma. Our dataset can serve as a benchmark for future software development and is available at (Shabanov et al., 2025).

Data Availability & Code

The full code, build pipeline, walkthrough video, and documentation for SALMA are available at https://github.com/lnilya/salma. This repository also contains executables that can be downloaded and used without any technical knowledge. All three datasets, their original flatbed scanner images/photographs and the segmentations resulting from the presented algorithms and used here are publicly available (Shabanov et al., 2025).

AUTHOR CONTRIBUTIONS

IS and AL conceived the initial ideas and designed the methodology; IS compiled, analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The full code, build pipeline, walkthrough video, and documentation for SALMA are available at https://github.com/lnilya/salma. This repository also contains executables that can be downloaded and used without any technical knowledge. All three datasets, their original flatbed scanner images/photographs and the segmentations resulting from the presented algorithms and used here are publicly available at https://doi.org/10.5281/zenodo.18090524.

ACKNOWLEDGEMENTS

We acknowledge Rachael Lockhart, Olivia Bird, Indira Leon Garcia and Janelle Veenendaal for contributing data. We are grateful to Moses Njau, Vaustine Obura, and Job Munene for segmentation of the images. We acknowledge the use of ChatGPT 4o, o1-preview and o1 models (chatgpt.com), as well as Grammarly (Grammarly.com) to generate information for background research, suggestions for improvements in writing style, outline structure, proofreading and high-

level manuscript feedback. After use, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G., & Violle, C. (2011). When and how should intraspecific variability be considered in trait-based plant ecology? *Perspectives in Plant Ecology, Evolution and Systematics*, *13*(3), 217–225. https://doi.org/10.1016/j.ppees.2011.04.003

Bhagat, S., Kokare, M., Haswani, V., Hambarde, P., & Kamble, R. (2022). Eff-UNet++: A novel architecture for plant leaf segmentation and counting. *Ecological Informatics*, *68*(101583), 101583. https://doi.org/10.1016/j.ecoinf.2022.101583

Both, S., Paine, C. E. T., & Bongalov, B. (2020). *Technical method comparison of leaf area assessments from leaves scanned with a flatbed scanner* [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.3780391

da Silveira Pontes, L., Louault, F., Carrère, P., Maire, V., Andueza, D., & Soussana, J.-F. (2010). The role of plant traits and their plasticity in the response of pasture grasses to nutrients and cutting frequency. *Annals of Botany*, *105*(6), 957–965. https://doi.org/10.1093/aob/mcq066

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I. C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P. B., Moles, A. T., Dickie, J., Gillison, A. N., Zanne, A. E., … Gorné, L. D. (2016). The global spectrum of plant form and function. *Nature*, *529*(7585), 167–171. https://doi.org/10.1038/nature16489

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I. C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P. B., Moles, A. T., Dickie,

J., Zanne, A. E., Chave, J., … Zotz, G. (2022). The global spectrum of plant form and function: enhanced species-level trait dataset. *Scientific Data*, *9*(1), 755. https://doi.org/10.1038/s41597-022-01774-9

Dubuis, A., Rossier, L., Pottier, J., Pellissier, L., Vittoz, P., & Guisan, A. (2013). Predicting current and future spatial community patterns of plant functional traits. *Ecography*, *36*(11), 1158–1168. https://doi.org/10.1111/j.1600-0587.2013.00237.x

Easlon, H. M., & Bloom, A. J. (2014). Easy Leaf Area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in Plant Sciences*, *2*(7), 1400033. https://doi.org/10.3732/apps.1400033

Fagúndez, J. (2013). Heathlands confronting global change: drivers of biodiversity loss from past to future scenarios. *Annals of Botany*, *111*(2), 151–172. https://doi.org/10.1093/aob/mcs257

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Funk, J. L., Larson, J. E., Ames, G. M., Butterfield, B. J., Cavender-Bares, J., Firn, J., Laughlin, D. C., Sutton-Grier, A. E., Williams, L., & Wright, J. (2017). Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biological Reviews of the Cambridge Philosophical Society*, *92*(2), 1156–1173. https://doi.org/10.1111/brv.12275

Getman-Pickering, Z. L., Campbell, A., Aflitto, N., Grele, A., Davis, J. K., & Ugine, T. A. (2020). LeafByte: A mobile application that measures leaf area and herbivory quickly and accurately. *Methods in Ecology and Evolution*, *11*(2), 215–221. https://doi.org/10.1111/2041-210x.13340

Gianoli, E., & Valladares, F. (2012). Studying phenotypic plasticity: the advantages of a broad approach: STUDYING PHENOTYPIC PLASTICITY. *Biological Journal of the Linnean Society. Linnean Society of London*, *105*(1), 1–7. https://doi.org/10.1111/j.1095-8312.2011.01793.x

Grabherr, G., Gottfried, M., & Pauli, H. (2010). Climate change impacts in alpine environments: Climate change impacts in alpine environments. *Geography Compass*, *4*(8), 1133–1153. https://doi.org/10.1111/j.1749-8198.2010.00356.x

Green, S. J., Brookson, C. B., Hardy, N. A., & Crowder, L. B. (2022). Trait-based approaches to global change ecology: moving from description to prediction. *Proceedings. Biological Sciences / The Royal Society*, *289*(1971), 20220071. https://doi.org/10.1098/rspb.2022.0071

Guo, K., Cornwell, W. K., & Bragg, J. G. (2024). Using machine learning to link climate, phylogeny and leaf area in eucalypts through a 50- fold expansion of leaf trait datasets. *The Journal of Ecology*. https://doi.org/10.1111/1365-2745.14354

Jain, R. C., Kasturi, R., & Schunck, B. G. (1995). *Machine Vision*. McGraw Hill Higher Education.

Katabuchi, M. (2015). LeafArea: an R package for rapid digital image analysis of leaf area. *Ecological Research*, *30*(6), 1073–1077. https://doi.org/10.1007/s11284-015-1307-x

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar C, C., Aleixo, I., Ali, H., … Wirth, C. (2020). TRY plant trait database - enhanced coverage and open access. *Global Change Biology*, *26*(1), 119–188. https://doi.org/10.1111/gcb.14904

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2304.02643

Kühn, N., Tovar, C., Carretero, J., Vandvik, V., Enquist, B. J., & Willis, K. J. (2021). Globally important plant functional traits for coping with climate change. *Frontiers of Biogeography*, *13*(4). https://doi.org/10.21425/F5FBG53774

Lavorel, S., & Garnier, E. (2002). Predicting changes in community composition and ecosystem

functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, *16*(5), 545–556. https://doi.org/10.1046/j.1365-2435.2002.00664.x

Lepš, J., de Bello, F., Šmilauer, P., & Doležal, J. (2011). Community trait response to environment: disentangling species turnover vs intraspecific trait variability effects. *Ecography*, *34*(5), 856–863. https://doi.org/10.1111/j.1600-0587.2010.06904.x

Mascalchi, P. (2016). *ImageJ_Auto-white-balance-correction: ImageJ / Fiji macro to automatically correct white balance in RGB images*. Github. https://github.com/pmascalchi/ImageJ_Auto-white-balance-correction

Meira, L. A., Pereira, L. E. T., Santos, M. E. R., & Tech, A. R. B. (2020). USPLeaf: Automatic leaf area determination using a computer vision system. *Ciencia Agronomica*, *51*(4). https://doi.org/10.5935/1806-6690.20200073

Montès, N., Tosini, L., Laffont-Schwob, I., Le Bagousse-Pinguet, Y., & Folzer, H. (2024). FAMeLeS: A multispecies and fully automated method to measure morphological leaf traits. *Methods in Ecology and Evolution*, *15*(3), 484–492. https://doi.org/10.1111/2041-210x.14287

Osnas, J. L. D., Lichstein, J. W., Reich, P. B., & Pacala, S. W. (2013). Global leaf trait relationships: mass, area, and the leaf economics spectrum. *Science (New York, N.Y.)*, *340*(6133), 741–744. https://doi.org/10.1126/science.1231574

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/tsmc.1979.4310076

Parkhurst, D. F., & Loucks, O. L. (1972). Optimal leaf size in relation to environment. *The Journal of Ecology*, *60*(2), 505. https://doi.org/10.2307/2258359

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine*

Learning Research, *12*, 2825–2830.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M. S., Cornwell, W. K., Craine, J. M., Gurvich, D. E., Urcelay, C., Veneklaas, E. J., Reich, P. B., Poorter, L., Wright, I. J., Ray, P., Enrico, L., Pausas, J. G., de Vos, A. C., … Cornelissen, J. H. C. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, *61*(3), 167. https://doi.org/10.1071/bt12225

Prager, C. M., Classen, A. T., Sundqvist, M. K., Barrios-Garcia, M. N., Cameron, E. K., Chen, L., Chisholm, C., Crowther, T. W., Deslippe, J. R., Grigulis, K., He, J.-S., Henning, J. A., Hovenden, M., Høye, T. T. T., Jing, X., Lavorel, S., McLaren, J. R., Metcalfe, D. B., Newman, G. S., … Sanders, N. J. (2022). Integrating natural gradients, experiments, and statistical modeling in a distributed network experiment: An example from the WaRM Network. *Ecology and Evolution*, *12*(10), e9396. https://doi.org/10.1002/ece3.9396

Prewitt, J. M., & Mendelsohn, M. L. (1966). The analysis of cell images. *Annals of the New York Academy of Sciences*, *128*(3), 1035–1053. https://doi.org/10.1111/j.1749-6632.1965.tb11715.x

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You only look once: Unified, real-time object detection. In *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1506.02640

Rixen, C., Wipf, S., Rumpf, S. B., Giejsztowt, J., Millen, J., Morgan, J. W., Nicotra, A. B., Venn, S., Zong, S., Dickinson, K. J. M., Freschet, G. T., Kurzböck, C., Li, J., Pan, H., Pfund, B., Quaglia, E., Su, X., Wang, W., Wang, X., … Deslippe, J. R. (2022). Intraspecific trait variation in alpine plants relates to their elevational distribution. *The Journal of Ecology*, *110*(4), 860–875. https://doi.org/10.1111/1365-2745.13848

Sandel, B., Gutiérrez, A. G., Reich, P. B., Schrodt, F., Dickie, J., & Kattge, J. (2015). Estimating themissing species bias in plant trait measurements. *Journal of Vegetation Science: Official Organ of the International Association for Vegetation Science*, *26*(5), 828–838.

https://doi.org/10.1111/jvs.12292

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Schrader, J., Pillar, G., & Kreft, H. (2017). Leaf-IT: An Android application for measuring leaf area. *Ecology and Evolution*, *7*(22), 9731–9738. https://doi.org/10.1002/ece3.3485

Shabanov, I., Deslippe, J., & Lensen, A. (2025). *Datasets for SALMA: A Machine Learning Tool for Precise Leaf Morphology Measurements of Small Leaves* [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.18090525

Swenson, N. G., Worthy, S. J., Eubanks, D., Iida, Y., Monks, L., Petprakob, K., Rubio, V. E., Staiger, K., & Zambrano, J. (2020). A Reframing of Trait–Demographic Rate Analyses for Ecology and Evolutionary Biology. *International Journal of Plant Sciences*, *181*(1), 33–43. https://doi.org/10.1086/706189

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos (Copenhagen, Denmark)*, *116*(5), 882–892. https://doi.org/10.1111/j.0030-1299.2007.15559.x

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., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Highlights

- Introducing SALMA: A machine learning tool for precise leaf morphology measurements.
- Features a graphical user interface for users without technical expertise
- Achieved 2-15x lower leaf area errors than existing grayscale algorithms.
- Leverages few-shot learning (1-4 examples) and multi-color/gradient analysis.
- Identifies a power-law relationship between leaf area and measurement error.