Enhancing Leaf Disease Classification with Feature Fusion and Ensemble Learning

Afizou Yayé Kaka[1,*], *Boutaina* Hdioud[2,**], and *Rachid Oulad* Haj Thami[2,***]

[1]IRDA Team, IRDA Laboratory, Rabat IT Center, ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

[2]Department of Computer Science, Rabat IT Center, ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

Abstract. Accurate and efficient detection of plant diseases is vital for enhancing agricultural production and safeguarding food safety. In this insightful study, we dive into a cutting-edge learning technique that integrates various features extracted from leaf images. By employing support vector machines (SVM) alongside the innovative Bootstrap aggregating (bagging) technique, we meticulously assess the contribution of each feature and explore the power of probability-weighted fusion.

Our findings reveal that embracing probability fusion significantly elevates the model's overall performance, surpassing the traditional method of merely equalizing features. This approach underscores the profound impact that multi-feature weighted fusion can have, paving the way for a more robust and effective classification model.

Keywords: Plant disease classification, weighted Probability Fusion, Feature Fusion, Support Vector Machine, Bagging.

1 Introduction

Precision agriculture has experienced significant growth in recent years due to the use of innovative and advanced technologies that enhance plant disease detection, crop management, and irrigation [1, 2]. One of the biggest challenges in this field is the early detection of plant diseases, traditionally performed by agricultural experts. However, advancements in computer vision and machine learning techniques now allow for the development of automated systems that can replace these traditional methods [2]. These systems are capable of processing large volumes of data and autonomously identifying signs of disease.

Plant health classification often relies on various visual characteristics such as texture, color, and shape [3]. However, not all characteristics hold the same importance in the classification process, and a poor combination can lead to suboptimal model performance, as demonstrated for SVM-based models and morphology-based descriptors [4], which shows that non-optimized descriptor selection can significantly degrade the performance of the models especially in the agricultural domain. The most common approach involves concatenating

*e-mail: afizouyayek@gmail.com
**e-mail: boutaina.hdioud@um5.ac.ma
***e-mail: rachid.ouladhajthami@gmail.com

these characteristics to train a model; however, this method does not consider the relative significance of each characteristic [5].

In this study, we propose a weighted fusion of probabilities from distinct models trained on individual features to enhance the accuracy of disease classification from tomato leaf images. We examine the impact of assigning different weights to features and demonstrate that adaptive probability fusion yields superior performance compared to naive concatenation. Additionally, we apply bagging techniques to improve the robustness of individual models.

The objective of this work is to present a more robust, efficient, and interpretable method for plant disease classification by optimally combining visual features. This research aims to provide valuable information for enhancing AI-based detection systems in agriculture. The remainder of this paper will be organized as follows: we will begin with a section dedicated to the state of the art, discussing recent advances in this area. Next, we will outline our methodology, followed by the analysis and interpretation of the results, and conclusion section.

2 Related Works

Plant disease detection through image processing has largely relied on the extraction of visual features such as color, texture, and shape. These techniques have demonstrated their effectiveness in different contexts, with performances varying depending on the feature combinations. Thus, for the detection of rice diseases such as Bacterial leaf blight, Rice blast, and Sheath rot, the contour-texture combination achieved an accuracy of 97.72% compared to 94.4% for color-contours [6], thus highlighting the importance of choosing features according to the problem.

Given this diversity of approaches, there is still no consensus on the optimal choice of features. Some even use only a combination of two features, while others prefer to exploit all three, thus obtaining equally effective results achieving accuracies ranging between 89% and 97 % depending on dataset [5]. These observations suggest that no feature should be discarded a priori, but that their relative weighting would merit specific optimization.

Since the advent of deep learning, convolutional neural networks (CNNs) have demonstrated superior performance by directly extracting complex features. These techniques seamlessly achieve performance above 97 % [7], often outperforming traditional techniques. More recent and specialized techniques for detecting plant diseases include vision transformers (ViT) and EfficientNet, which can achieve impressive accuracy levels of up to 99% in plant disease classification [8–12]. However, their black-box nature poses interpretability challenges, particularly critical in agriculture where understanding the model's decisions is important for domain experts.

To reconcile performance and interpretability, hybrid approaches are emerging, combining the strengths of both traditional techniques and Deep Learning.Hybrid approaches combining handcrafted features and deep learning have also been proposed to improve performance while maintaining a certain level of interpretability. However, most of the work is limited to a simple concatenation of features , thus neglecting the relative contribution of each feature. Despite recent advances in the field, there are still challenges in optimizing feature fusion and developing more transparent models [13]. Our work falls within this theme by proposing a weighted analysis of the probabilities provided by each feature as well as the contribution of techniques such as bagging in order to improve both the robustness and interpretability of diagnoses.

3 Methodology

"Fig. 1" summarizes our methodology's key steps, from data collection to final classification. It includes extraction, training, bagging, and evaluation phases.

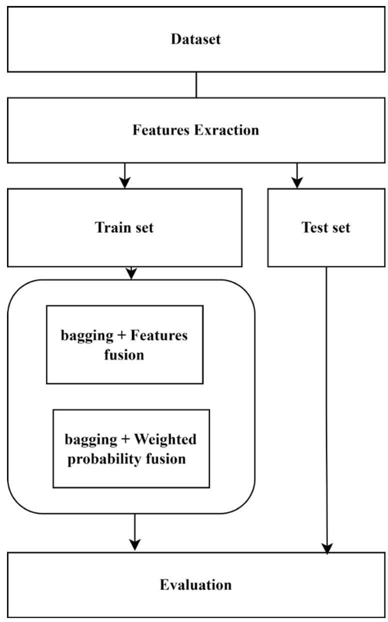

Figure 1. Overview of the proposed workflow for tomato leaf disease classification.

3.1 Dataset

To conduct this work, we utilized images of tomato leaves from the Plant Village dataset, which is one of the most widely used databases in agricultural research [14]. The "Tab. 1" provides important information regarding the distribution of the data.The data is classified into two categories: healthy leaves and those affected by septoria leaf spot (see "Fig. 2"). A total of 3362 images will be used to train our model, with a training/testing ratio of 80-20.

Table 1. Distribution of training and validation data

Class	Train	Validation	Total
Healthy	1272	319	1591
Septoria	1416	355	1771
Total	2688	674	3362

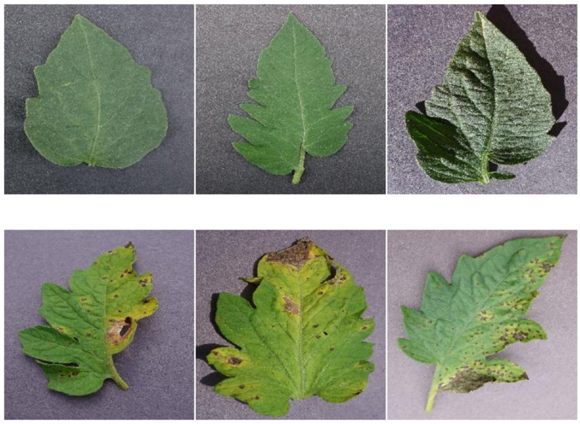

Figure 2. Images of healthy leaves (top row) and images of leaves affected by septoria leaf spot (bottom row).

3.2 Features Extraction

For the extraction of characteristics, we decided to extract three types of visual characteristics most used in the detection of plant diseases namely color, texture and shape[3]. For each of these characteristics we extracted various types of information. Color features are extracted via Red-Green-Blue (RGB) histograms and statistics like means and standard deviation, which capture chromatic variations often indicative of lesions or anomalies. Texture, on the other hand, is analyzed through a co-occurrence matrix, calculating six statistical properties such as contrast, energy, dissimilarity, homogeneity, correlation and Angular Second Moment (ASM) over various angles and distances [5]. To these, we added differential operators like the Laplacian to quantify roughness. Finally, for shape, we calculated Hu and Zernike invariant moments to describe the global geometry of the regions of interest, as well as edge detection with Canny to identify irregular borders. These characteristics will then be used for training the different models.

3.3 SVM training and model bagging

To train our model, we decided to use SVM, a classification algorithm. The objective is to find the optimal hyperplane to accurately separate the different classes, while maximizing the margins between the different classes [4]. It is often formulated as an optimization problem with an objective function "(1)" to be maximized under certain constraints "(2)":

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{n}\xi_i \tag{1}$$

where

$$\begin{cases} y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, & \forall i \in \{1,\ldots,n\} \\ \xi_i \geq 0, & \forall i \in \{1,\ldots,n\} \end{cases} \tag{2}$$

The vector $\mathbf{w}$ is associated with the features $\mathbf{x}_i$ and is used to maximize the margin between different classes. The bias b shifts the hyperplane to improve the separation of the classes.

The relaxation variables ξ_i (or e_i) allow for some tolerance of classification errors, particularly for datasets that are not perfectly linearly separable. The parameter C is a hyperparameter that regulates the balance between maximizing the margin and minimizing classification errors. A large C prioritizes reducing errors, while a small C promotes a wider margin with greater tolerance for errors [4].

The SVM is well suited for medium-sized data and has yielded interesting results on tomato leaf images. First, we used each feature to train an SVM. The Radial Basis Function (RBF) kernel was used due to its strong generalization ability . Then, we combined all the features to train a single classifier using the same hyperparameters as for the previous three classifiers. Finally, we evaluated and compared them.

For the rest of our work, we applied bagging, an ensemble learning method, to each classifier. This method improves the robustness and stability of the model . The hyperparameters used to train all models are defined in Table "Tab. 2".

Table 2. Detailed hyperparameters configuration

Hyperparameters	Value
C (SVM)	1
Kernel (SVM)	rbf
Probability (SVM)	True
n_estimators (bagging)	20
max_samples (bagging)	0.8
max_features (bagging)	0.8

3.4 Weighted probability fusion

Once bagging was applied to our four classifiers, we decided to combine our three features in a weighted manner. Unlike traditional feature concatenation approaches,our approach is inspired by probability fusion methods used in multi-sensor data fusion. However, we use a linear fusion "(3)" rather than an SVM as a meta-merger.

$$p_{\text{fusion}} = \alpha p_c + \beta p_t + \gamma p_s \tag{3}$$

where α, β γ are the weighting coefficients and p_c the probability predicted by the color-based model, p_t the probability predicted by the texture-based model and p_s the probability predicted by the shape-based model ("Fig. 3").

This weighting will allow us to determine the importance of each feature in the final decision. To do this, we tested different combinations ranging from 0.05 to 0.95 with a step size of 0.5, subject to a constraint to ensure that the final probability calculated by fusion remains within a range comparable to a true probability, but also to reduce the search space "(4)".

$$\begin{aligned} \alpha + \beta + \gamma &= 1 \\ \alpha, \beta, \gamma &\geq 0 \end{aligned} \tag{4}$$

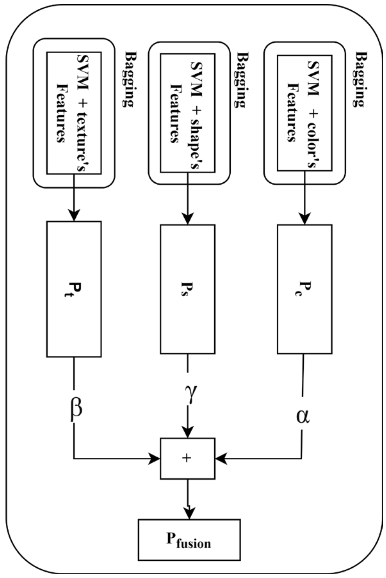

Figure 3. Architecture of the proposed weighted probability fusion framework.

4 Results

The results presented in this section are obtained after testing our different models on the test set. The first phase of our work consisted of training our features separately which gave us interesting results. Even though all the features provide results above 80%, we notice that the performance varies widely from one feature to another, confirming the results [6]. The color obtained the highest performance going up to 95.99% for the accuracy, followed by the texture with 91.99 %. This could be explained by the fact that the septoria leaf spot (as observed in "Fig. 2") causes yellowing and change in leaf texture much more than the change in shape . For more details the results obtained are recorded in "Tab. 3".

Table 3. Performance results (in %) obtained from the various models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Color	95.99	97.95	94.36	96.12
Texture	91.99	90.13	95.21	92.60
Shape	84.57	91.41	78.02	84.19
Combined features	94.07	95.65	92.95	94.28

As expected the equal fusion of the three features improves the performance of the model compared to the model including only texture and shape with an accuracy of 94.07% but remains slightly lower than the color. This could indicate that the addition of texture and shape does not provide sufficient discriminating information compared to color [15]. We can therefore deduce that not all features have the same importance (but this does not mean they are useless) in decision-making and can even introduce noise if they are not well weighted (see "Fig. 4").

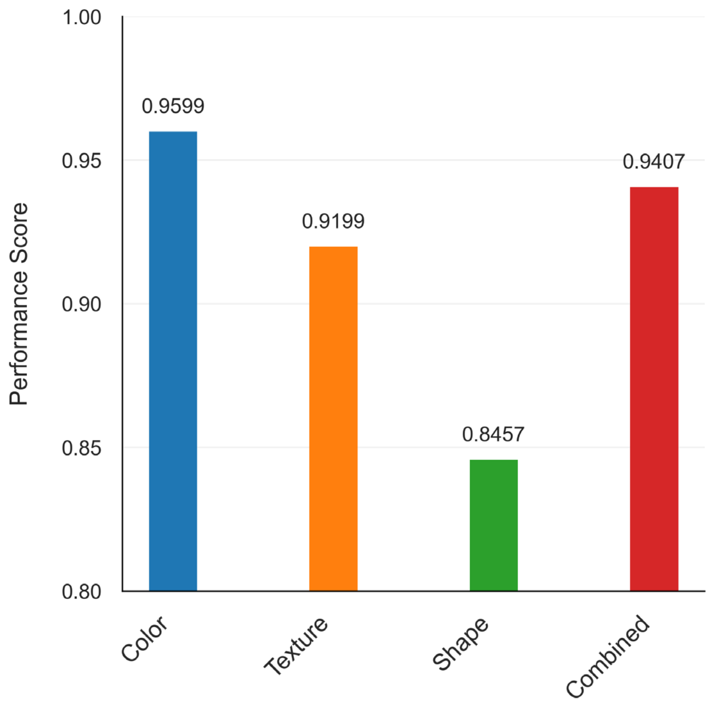

Figure 4. Comparison of the classification accuracy obtained using different feature configurations and fusion strategies.

To improve the performance of our models, we applied bagging to each results obtained above and observed, in the majority of cases, an increase in the overall performance of the models, except for the shape feature, where most performance decreased slightly. Generally, bagging improves model robustness by reducing overfitting and variance. This decrease in performance can be explained by the fact that the shape is less reliable or that bagging generates too much diversity, which alters its contribution. Despite the application of bagging and the performance improvement, we note that the model combining features still lags behind compared to color (see "Tab. 4"). This therefore justifies the use of bagging to strengthen models but also the need for appropriate feature weighting.

Table 4. Accuracy (in %) obtained before and after bagging

Model	**Before Bagging**	**After Bagging**
Color	95.99	96.74
Texture	91.99	93.03
Shape	84.57	84.12
Combined features	94.07	94.21

So we decided to use a weighted fusion of our different characteristics, the observed results show that the weighted fusion of the different characteristics allowed to obtain higher performances going up to 97.63% in accuracy, against 94.21% for the model with bagging combining the results equally. It should be noted that even if this weighting of the probabilities of the different characteristics allows to improve the model, it can also make it less efficient if we do not find the ideal combination. For this reason we decided to test various combinations of α , β and γ in order to find the optimal combination of probabilities provided by each characteristic (see "Fig. 5").

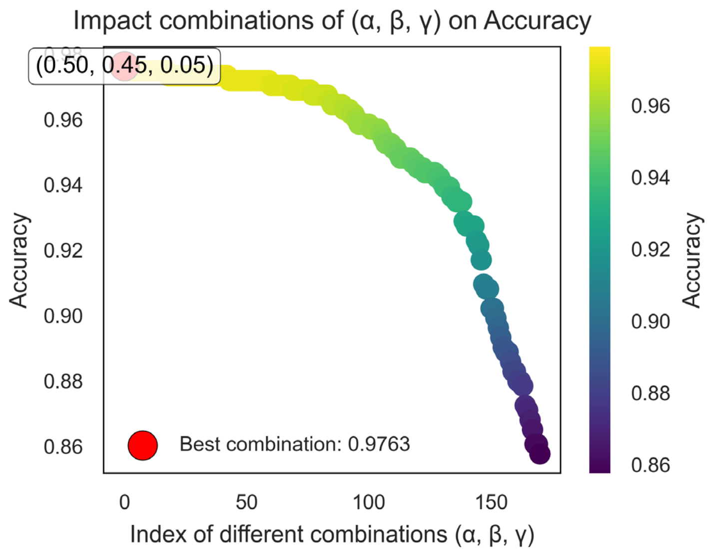

Figure 5. Evolution of accuracy as a function of the values of α, β and γ.

We therefore obtained our best model with α=0.5, β=0.45 and γ=0.05 respectively for color, texture and shape ,thus confirming the importance of color and texture for the detection of certain diseases [15]. These values obtained show that color and texture contribute more to the final decision and that shape makes a marginal contribution, which confirms the previous results. "Fig. 6" represents the ROC curve of our two models (probability-weighted and unweighted feature fusion), we note that they have a very strong ability to distinguish between our different classes with a slight advantage for the probability-weighted fusion model.

The results obtained in this work show that weighted probability fusion increases the model's performance by providing to each feature a weight based on its importance in detecting septoria leaf spot. The final weights obtained demonstrate the importance of color as a key characteristic in disease detection, thus reinforcing its role in identifying plant diseases .

However, shape plays a less important role, with a very low weight, demonstrating its limited contribution to our model. This result could be explained by the fact that this disease does not cause major changes in leaf shape. Nevertheless, this characteristic remains important for detecting other plant diseases, particularly those that cause leaf deformation [6].

This analysis allows us to understand the relative importance of each characteristic in detecting Septoria leaf spot, while highlighting the interpretability of our approach. Furthermore, the use of methods such as bagging allowed our model to be more robust and to

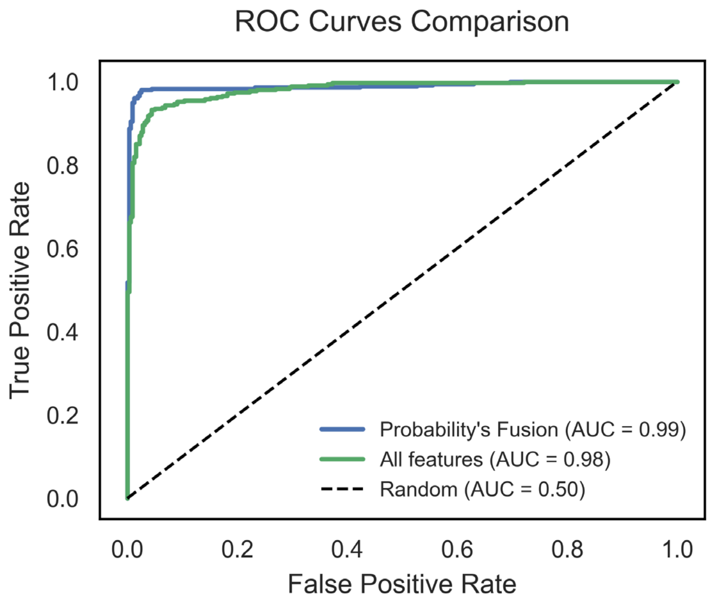

Figure 6. Comparison of the ROC curves obtained using the probability-weighted fusion model and the traditional feature concatenation model.

achieve competitive performance compared with several recent deep learning techniques [10–12]. While our method shows strong performance, some recent works [8, 9] have achieved slightly higher scores, surpassing our results by 2% and reaching performances above 99%. Nonetheless, our approach has a distinct advantage by providing an optimal balance between:

- Classification performance
- Interpretability of results

These characteristics are particularly important in the agricultural sector, where decision transparency is often as vital as accuracy. "Tab. 5" provides a detailed comparative analysis of recent techniques for detecting foliar diseases.

Table 5. Comparison of model performance (in %) with other works- Accuracy (Acc) ,Precision (P) ,Recall (R) and F1-score (f1)

Ref	Algorithms	Acc	P	R	F1
[8]	ViT with L1-norm and Data Augmentation	99.74	99.66	99.39	99.53
[9]	EfficientNet	99.01	98.88	98.95	99.01
[10]	CNN + 3 Channel Attention	97.56	98.00	98.00	97.00
[11]	ViT with inception module in place of MLP	96.93	96.98	96.83	96.90
[12]	CNN + ViT	-	96.59	96.80	96.69
Our	**Bagging + Weighted probability fusion**	**97.63**	**98.01**	**97.46**	**97.74**

However, despite these advantages, our technique has a significant limitation related to weighting. As shown in "Fig. 5", incorrect weighting can lead to a significant decrease in performance. It is also worth noting that these results could be improved by exploring a larger search space with a finer step size than the one we set.

5 Conclusion

Through our work, we were able to demonstrate that the various features commonly used in the field do not have the same importance, and this difference can significantly affect a model's performance. This highlights the importance of appropriately weighting the information extracted from our data to obtain more efficient and robust models.

Our model, combining both bagging and weighted probability fusion for each feature, enabled us to develop a system that is both robust and efficient, achieving an accuracy of 97.63% on test data. Despite these encouraging results, our approach presents a risk related to the weighting scheme, as poor tuning can significantly reduce the model's capabilities.

This limitation, however, opens up new possibilities, such as the integration of more advanced machine learning mechanisms (e.g., meta-classifiers) for adaptive probability fusion, or extending it to more diverse data to improve the model's robustness. Finally, this technique could be integrated into embedded systems to assist farmers and experts in their decision-making, by providing additional information to confirm their diagnoses.

References

[1] Farhan S.M., Yin J., Chen Z., and Memon M.S., "A Comprehensive Review of LiDAR Applications in Crop Management for Precision Agriculture." Sensors **24**, 5409 (2024). https://doi.org/10.3390/s24165409

[2] Chin R., Catal C., and Kassahun A., "Plant disease detection using drones in precision agriculture." Precision Agriculture **24**, 1663–1682 (2023).

[3] Vishnoi V.K., Kumar K., and Kumar B., "A comprehensive study of feature extraction techniques for plant leaf disease detection." Multimedia Tools and Applications **81**, 367–419 (2022). https://doi.org/10.1007/s11042-021-11375-0

[4] Lubis I., Siregar R., et al., "Recognition of Medicinal Plant Leaf Patterns Using Morphology-Based and GLCM Feature Extraction." Journal of Artificial Intelligence and Engineering Applications **3**, 834–839 (2024).

[5] Sharif M., Khan M.A., Iqbal Z., Azam M.F., Lali M.I.U., and Javed M.Y., "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection." Computers and Electronics in Agriculture **150**, 220–234 (2018). https://doi.org/10.1016/j.compag.2018.04.023

[6] Yao Q., Guan Z., Zhou Y., Tang J., Hu Y., and Yang B., "Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Color Texture Features." In: 2009 International Conference on Engineering Computation, Hong Kong, China, pp. 79–83 (2009). https://doi.org/10.1109/ICEC.2009.73

[7] Iftikhar M., Kandhro I.A., Kausar N., et al., "Plant disease management: A fine-tuned enhanced CNN approach with mobile app integration for early detection and classification." Artificial Intelligence Review **57**, 167 (2024).

[8] Tiwari M., Kumar H., Prakash N., et al., "Tomato Disease Detection Using Vision Transformer with Residual L1-Norm Attention and Deep Neural Networks." International Journal of Intelligent Engineering and Systems **17** (2024).

[9] Tanim S.A., Anik Z.H., Hossain M.I., et al., "Precise detection of tomato leaf diseases using deep learning approach with EfficientNet." In: 2023 26th International Conference on Computer and Information Technology (ICCIT), pp. 1–6 (2023).

[10] Chen H., Wang Y., Jiang P., et al., "LBFNet: A Tomato Leaf Disease Identification Model Based on Three-Channel Attention Mechanism and Quantitative Pruning." Applied Sciences **13**, 5589 (2023).

[11] Gole P., Bedi P., Marwaha S., et al., "TrIncNet: a lightweight vision transformer network for identification of plant diseases." Frontiers in Plant Science **14**, 1221557 (2023).

[12] Wang Y., Chen Y., and Wang D., "Convolution network enlightened transformer for regional crop disease classification." Electronics **11**, 3174 (2022).

[13] Nagachandrika B., Prasath R., and Joe I.P., "An automatic classification framework for identifying type of plant leaf diseases using multi-scale feature fusion-based adaptive deep network." Biomedical Signal Processing and Control **95**, 106316 (2024).

[14] Li L., Zhang S., and Wang B., "Plant disease detection and classification by deep learning—a review." IEEE Access **9**, 56683–56698 (2021).

[15] Ahmad N., Asif H.M.S., Saleem G., et al., "Leaf image-based plant disease identification using color and texture features." Wireless Personal Communications **121**, 1139–1168 (2021). https://doi.org/10.1007/s11277-021-09054-2