Smart Leaf Identification and Analysis using Artificial Intelligence

Balamurali Ravichandran*, Abhinav Tiwari & Joydeep Roy

Department of Computer Science and Engineering, Faculty of Science and Technology, The ICFAI Foundation for Higher Education, Dontanpally, Shankarpally Road, Hyderabad, India – 501203
*E-mail: mohmur@gmail.com

Abstract. This paper describes a leaf classification framework based on machine learning methodologies. The dataset included 36 leaf species with 340 instances and 16 features. Decision Tree, Logistic Regression, Support Vector Machine, XGBoost, AdaBoost, and Random Forest were applied initially on the raw dataset. It was found that Decision Tree, XGBoost and Random Forest performed well, with accuracy scores of 53%, 60% and 85% respectively. After feature reduction based on Information Gain and correlation matrix analysis, four highly correlated features were removed. This led to a marginal improvement in accuracy to 60% for Decision Tree and to 65% percentage for XG Boost, while the accuracy of Random Forest dropped to 74%. We also applied SMOTE (Synthetic Minority Oversampling Technique) to balance the raw dataset. This enhanced the accuracy of Random Forest and XGBoost to 91% and 81% respectively. With Computer Vision, we achieved an accuracy of 66% for SVM, 61% for Random Forest, 43% for K-NN and 37% for the Decision Tree classifier on the image dataset. The pre-trained VGG16 model achieved an impressive 90.47% accuracy, outperforming the custom CNN model and conventional models like Linear Discriminant Analysis (LDA) and KNN with various K-values from the literature.

Keywords: *feature reduction; K-NN; logistic regression; random forest; VGG16.*

1 Introduction

Historically, the task of plant identification and leaf classification is carried out by taxonomists. Due to growing technological trends like computer vision, machine learning, deep learning, image recognition etc. automated leaf classification becomes feasible. Such automated techniques can classify leaves quickly and accurately. The classification is based on unique features like color, shape, texture, size etc. The automated leaf classification system can be used for various domains like forestry, ecology and horticulture. Especially, it can be used in agriculture in identifying various crops, diseases etc. The typical challenges with this automated system are overfitting, limited data, noise, etc.

Most of the earlier works used small, imbalanced datasets. They lacked generalization. Existing models often suffer from overfitting. Furthermore,

Received on January 23rd, 2026, 1st Revision received on June 25th, 2026, 2nd Revision received on July 23rd, 2026, Accepted for publication on August 13th, 2026.
Copyright © 2026 Published by IRCS-ITB, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2026.20.3.6

handcrafted feature extraction leads to redundant attributes and reduced efficiency. Deep learning architectures such as ResNet, Inception, and MobileNet have achieved promising results. However, deploying these models is difficult due to their large size and high computational complexity.

Considering these gaps, the present study integrated classical machine learning, computer vision, and deep learning. We systematically applied feature reduction techniques (Information Gain and correlation matrix analysis) to improve efficiency. The performances were benchmarked across diverse algorithms. Additionally, we also applied transfer learning using VGG16.

2 Literature Review

Asrey, *et al.* [1] developed an advanced fusion-based architecture that is capable of enhancing classification reliability in a real-world agricultural environment. This method overcomes some of the gaps that existed earlier. In the recent works on citrus fruit disease detection, machine learning and deep learning methods were used extensively for quality assessment and disease diagnosis automation on kinnow fruit. The earlier literature in this domain explored CNN-based architectures, image segmentation, and feature-extraction methods but they produced poor results. The primary challenges are lack of datasets, separating fruits from leaves. Moreover, the availability of image repositories of ripe Kinnow in the public domain is limited, which restricts model generalization. This results in inconsistent performance across varying lighting environments.

Image-based and machine learning approaches have been widely used for cotton leaf disease detection. This detection is important because early detection is required for minimizing yield losses and achieving sustainable crop production. The existing studies in this domain commonly used datasets with limited variability in lighting, background, or disease severity. As a result, this restricted the models' generalization in real-world agricultural environments. The recent research in this domain requires high-quality, diverse, and well-annotated image collections to improve deep learning performance in plant pathology. Thus Ripon, *et al.* [2] prepared a dataset that addresses this gap by providing natural variability, multiple disease classes, and high-resolution images that are suitable for developing robust disease-classification models.

Al-Falluji, *et al.* [3] have demonstrated the integration of multiple advanced architectures and XAI techniques to provide a transparent and rigorous comparative analysis for rice leaf disease identification, which can address some research gaps exhibited in earlier literature. CNNs and transformer-based models have demonstrated good results in agricultural AI when used in recent research studies. Earlier research primarily focused on single-model evaluation without

standardized preprocessing or comprehensive interpretability methods. To address these gaps, this study provides an improved approach.

Previous literature has shown that deep learning and transfer learning models like AlexNet, VGGNet, and ResNet showed promising results in plant disease detection with improved classification performance. Even though the results were good, they still suffered from high computational cost and large model size. This led to the development of lightweight architectures such as SqueezeNet, developed by Kabel, *et al.* [4], which can achieve comparable accuracy with reduced parameters. Additionally, integrating this approach with traditional classifiers such as SVM further enhances robustness and generalization in plant disease detection.

In recent times, the research in the area of smart agriculture has put more focus on the significance of early plant disease recognition to handle yield losses and reduce excessive pesticide usage. Earlier literature in this domain mostly used CNN-based architectures or transfer learning models like VGG, ResNet, and Inception variants for this purpose. These processes struggle with imbalance in the dataset, reducing interpretability. Many approaches handle this situation through data augmentation or model ensemble techniques, but they still struggle to achieve high accuracy across minority disease classes. To overcome this gap, Padmanabhuni, *et al.* [5] came up with a blended approach that combines deep feature extraction with optimized machine-learning classifiers for more robust and balanced prediction.

The research area in medical plant identification is dependent on modern-day deep learning architectures. However, these approaches require heavy computational resources and also large datasets need to be fed into these models. Some of the earlier works used some hybrid approach that combines segmentation, handcrafted feature extraction, and metaheuristic optimization for improved accuracy and efficiency, and they showed good results. It was observed that evolutionary strategies like swarm intelligence algorithms enhance segmentation quality, feature selection, and classification performance in plant-species recognition tasks. Sachar, *et al.* [6] incorporated BRISK-based feature extraction and a Levy flight–driven Dragonfly feature selector for robust medicinal plant classification.

Rathor, *et al.* [7] present a multi-level feature extraction framework called CNN-WOPNet. This is designed for detection of lettuce leaf diseases under real-world hydroponic conditions. In this approach, they combined the Walrus Optimization technique for CNN hyperparameter tuning with a ParNet-based attention mechanism that effectively highlights the critical disease regions in plant images.

Rohith, *et al.* [8] gathered a dataset of 3,200 images across four groups to automate detection and classify diseases in apple leaf by applying a deep learning framework. They evaluated multiple pretrained CNN models with the Residual Neural Network, which attained the maximum validation accuracy of 98.9%, outperforming techniques like VGG, InceptionV3, EfficientNet, MobileNet, and a baseline CNN. Their results confirm the possibilities of deep learning-based systems in real-time orchard management and sustainable apple disease control.

Pacheco-Prado, *et al*. [9] looked into the utilization of RGB point clouds collected from UAVs to automate the classification of urban tree species. Their aim was to reduce the cost and time associated with traditional methods. They evaluated both random forest and deep-learning models (PointNet, PointNet++) on a dataset of 809 tree crowns from eight varieties. Their result demonstrated that the random forest classifier attained the optimum performance, outperforming the deep-learning approaches.

The study performed by Yang, *et al.* [10] showcased that salt stress levels in soybean seedlings can be identified by a deep-learning framework. They have used the RGB leaf images for this task. In their work, they integrated a CNN with a Convolutional Block Attention Module to ensure effective feature extraction. This step is followed by dimensionality reduction and machine-learning classification. Overall, their approach offered a strong generalization potential for monitoring the soybean stress under increasing soil salinization.

The paper presented by Amrita, *et al.* [11] confirms the importance of early plant disease detection, being critical to maintain agricultural sustainability and improve crop quality and yield. They proposed various deep learning-based CNN models, including ResNet50, InceptionV3, VGG16, VGG19, and a sequential architecture. Their aim was to classify plant diseases from leaf images captured under varying conditions. When they performed this comparative analysis, it showed that deep learning can effectively extract disease-related features, enabling robust multi-class classification. Among all the tested models, VGG16 achieved the highest performance.

Sandhya, *et al.* [12] reports a study that focused on the importance of using artificial intelligence and computer vision approaches for detecting and classifying non-degradable plastic waste and unwanted aquatic plant growth in lakes and ponds collected from drone and satellite images. They used the ResNet-50 transfer learning model and the TensorFlow framework to create an automated approach that is able to identify plastics and invasive plants in aquatic environments.

Manual plant species identification has limitations, which are addressed in Sapna, *et al.* [13]. They suggested that digital image processing and ontology-based feature classification could be more beneficial. They combined ontology along with machine learning and then used a deep learning framework to achieve high classification accuracy.

Traditional CNNs need large datasets and high computational power for their operation. To overcome this, Gowthaman, *et al.* [14] proposed an efficient leaf classification approach that integrates Wavelet Scattering Networks (WSNs) with MobileNetV2. They leveraged WSNs' ability to extract robust texture features without training and MobileNetV2s' ability to capture high-level structural details. Moreover, a PCA-based classifier enhances accuracy by reducing feature redundancy. Overall, this hybrid method offers a practical alternative for plant species identification, especially in the environments where resources remain a challenge.

Mumtaz, *et al.* [15] reported reliable plant identification and disease management, which contributes to improved agricultural monitoring and sustainability. They built a hybrid approach that combines SIFT, wavelet transformation, and moreover particle gradient motion for effective feature extraction. This was further refined using bee colony optimization.

The study performed by Sharma, *et al.* [16] focused on automated medicinal plant identification, because they were aware of the limitations of manual classification and there is a growing pharmaceutical dependence on plant-based compounds. They used a huge dataset of Indian medicinal plants to train on AELGNet. This model is an attention-based deep learning architecture that combines both plant and leaf images. Moreover, this is optimized through CLAHE preprocessing to extract robust local and global features.

Zhang, *et al.* [17] came up with a Prototype-Enhanced Learning (PEL) approach that uses label embedding to capture inter-class semantic relationships, which ensures that the model is guided toward more discriminative features. They presented a prototype that updates the module iteratively, which refines soft labels by using shared semantic patterns across classes.

The study performed by Mahapatra*, et al.* [18] incorporated Vision Transformers with CNNs in a graph-based framework, which they called GCNN-ViT. This approach of using GCNN-ViT underlines the importance in the extraction of spatial features and helps to understand the global context for more accurate classification. When comparing this model with traditional CNNs and pre-trained networks, it surpassed them with notable results.

A creative approach is suggested by Dmitriev, *et al.* [19], who introduced a new preprocessing technique named Random Reflectance (RR). This technique creates synthetic spectral profiles, which helps in better explaining the central spectral trends of the target object. After comparing RR with traditional preprocessing methods, it was seen that it notably improved the accuracy in deep learning classification, which further helps to achieve significant gains in both multiclass and binary classification tasks.

In the Kashmir Valley, Apple cultivation is economically very important. Therefore, apple plant disease detection is very crucial. Ahmed, *et al.* [20] investigated the utilization of machine learning and deep learning approaches to mitigate this issue. The researchers assessed various CNN-based models, including Faster R-CNN, to recognize the disease in real time. Their results illustrate that AI-driven systems have potential to carry out the early diagnosis of apple plant disease. Moreover, it also confirms that sustainable orchard management is possible in precision agriculture.

The uniqueness of the present study lies in its multi-layered methodological approach: unlike most of the prior studies all available attributes were used, while low-information features were identified and removed. This enhanced accuracy for Decision Tree & XGBoost and reduced the computational overhead. We also applied SMOTE and enhanced the classification accuracy. The study offers a holistic benchmark by evaluating six classical ML algorithms, multiple CV classifiers, a custom CNN, and a pre-trained VGG16 model on both a structured CSV dataset and the image datasets. VGG16 achieved an accuracy of 90.47%. Thus it outperformed both custom CNN (61.9%) and traditional ML models before feature extraction.

3 Methodology

3.1 Data Availability Statement

The dataset used in this study was obtained from the UCI Machine Learning Repository 'Leaf'– https://archive.ics.uci.edu/dataset/288/leaf [21].

3.2 Data Set Description

The dataset includes a CSV (comma separated values) file with details of 36 different plant species and image samples of 40 leaf species.

Each record in the CSV corresponds to an individual leaf and is uniquely associated with a particular species. The dataset comprises various morphological and statistical features, including: Class (Species), Specimen Number,

Eccentricity, Aspect Ratio, Elongation, Solidity, Stochastic Convexity, Isoperimetric Factor, Maximal Indentation Depth, Lobedness, Average Intensity, Average Contrast, Smoothness, Third Moment, Uniformity, and Entropy.

Among the 40 species represented, 36 are included in the CSV file. Table 1 includes the scientific names of these species along with the class number. It is observed that, species numbered 1 to 15 and 22 to 36 correspond to simple leaves, while species 16 to 21 and 37 to 40 represent complex leaves.

Table 1 Various plant (leaf) species.

Class	Scientific Name	Class	Scientific Name
1	Quercus suber	19	Polypolium vulgare
2	Salix atrocinera	20	Pinus sp.
3	Populus nigra	21	Fraxinus sp.
4	Alnus sp.	22	Primula vulgaris
5	Quercus robur	23	Erodium sp.
6	Crataegus monogyna	24	Bougainvillea sp.
7	Ilex aquifolium	25	Arisarum vulgare
8	Nerium oleander	26	Euonymus japonicus
9	Betula pubescens	27	Ilex perado ssp. azorica
10	Tilia tomentosa	28	Magnolia soulangeana
11	Acer palmatum	29	Buxus sempervirens
12	Celtis sp.	30	Urtica dioica
13	Corylus avellana	31	Podocarpus sp.
14	Castanea sativa	32	Acca sellowiana
15	Populus alba	33	Hydrangea sp.
16	Acer negundo	34	Pseudosasa japonica
17	Taxus bacatta	35	Magnolia grandiflora
18	Papaver sp.	36	Geranium sp.

The dataset contains 8 to 15 image samples for each leaf species. These image samples were divided into training and validation datasets. Sample leaf images are shown in Figure 1.



Figure 1 Sample leaf images from the dataset.

The dataset features and the training set particulars are shown in Table 2.

Table 2 Dataset features and training/testing details.

Dataset	Species Count	Total Samples	Train (80%)	Test (20%)	Features
CSV (UCI Leaf Dataset)	36	340	272	68	16 morphological/statistical features
Image Dataset	40	421 (8–15 per species)	337	84	RGB leaf images

3.3 Feature Selection

Feature selection is a technique used to identify the key features for making the right predictions. Too many features may lead to overfitting. This is a situation where the model performs well for train data but not for test data. This study used feature selection techniques like Information Gain and Correlation Matrix to enhance the model performance.

3.4 Information Gain Technique

Information Gain represents the degree to which uncertainty reduces after splitting the data based on an attribute. This idea is inspired from Shannon's entropy technique. This technique enables a model to identify relevant features and ignore irrelevant features. By using the Information Gain technique it was found that four out of sixteen features in the Leaf dataset were of little significance.

3.5 Correlation Matrix

Creating a correlation matrix is a technique used to identify the relationship of every attribute with every other attribute. It provides values between -1 and +1, were -1 represents that two variables are negatively correlated and +1 represents a positive correlation. The diagonal represents self-correlation.

We evaluated the performance of six classification algorithms—Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), AdaBoost (AB), XGBoost (XGB), and Random Forest—on the CSV dataset. Their effectiveness was assessed using common benchmarks such as accuracy, precision, recall, and F1-score.

In addition, we extended our study to the image dataset using a computer vision approach. Several machine learning algorithms, including SVM, Random Forest, K-Nearest Neighbors (KNN), and Decision Tree (DT), were applied to evaluate classification performance. Furthermore, we also used a pre-trained deep learning

model VGG16 to enhance the evaluation of the image dataset and benchmark its performance against traditional models.

3.6 Process Flow Pipeline

The process flow pipeline followed in this study is shown in Figure 2.

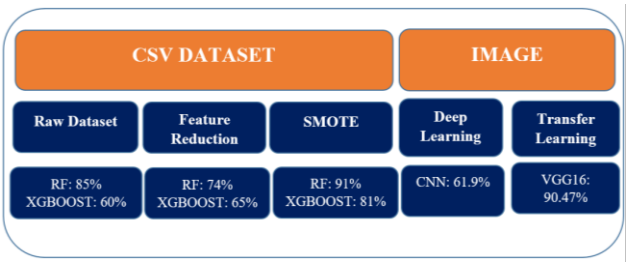


Figure 2 Process flow pipeline.

We used the CSV dataset and image dataset for leaf classification. First, the raw CSV dataset was studied using various machine learning algorithms. Following that, we applied several feature extractions techniques and then studied the performance of various ML algorithms. Next, we applied SMOTE (Synthetic Minority Oversampling Technique) on the raw dataset and then analyzed the performance of various algorithms. For the image dataset, we applied a custom CNN (convolution neural network) technique and studied the performance and then the applied transfer learning approach to analyze the performance.

4 Results

We applied several supervised learning and ensemble machine learning models on the CSV dataset. The algorithms applied were: Support Vector Machine (SVM), Decision Tree Classifier, Logistic Regression, AdaBoost, XGBoost, and Random Forest, each contributing to the analysis and classification process.

Initially, these six algorithms were applied on the raw dataset and the results were observed. Later, we applied feature extraction and excluded some irrelevant features and subsequently applied these six algorithms again and observed the results.

4.1 Performance Metrics

The metrics considered to measure the performance of these algorithms on the datasets were Accuracy, precision, recall and F1-score.

4.1.1 Accuracy

Accuracy was calculated as shown in Eq. (1); it is the ratio between true positive and negative classes and the total sum.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

where TP = true positive, FP = false positive, TN = true negative, FN = false negative.

4.1.2 Precision

Precision was calculated as shown in Eq. (2); it is the ratio between true positive and all positive classes.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

4.1.3 Recall

Recall is calculated as shown in Eq. (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4.1.4 F1 score

F1-score was calculated as shown in Eq. (4):

$$F1 - Score = 2 * \frac{precision*Recall}{precision+Recall} \quad (4)$$

4.2 Performance of the Proposed Model

The leaf dataset consists of 36 species, with 16 attributes. The dataset contains 340 entries. There are no missing or null values. To assess the efficacy of the proposed models, four performance metrics were used: accuracy, precision, recall, and F1-score. The algorithms—Support Vector Machine (SVM), Logistic Regression, and Decision Tree Classifier—were rigorously tested on both the original dataset and a modified version from which the sixth, eleventh, twelfth, and fourteenth features were excised owing to their negligible information gain.

By taking into account the raw data without any filtering, the comparative measures for the different machine-learning approaches, including SVM, Logistic Regression, and Decision Tree algorithms, is given in Table 3.

Table 3 Algorithm assessment – raw dataset without feature reduction – CSV dataset.

	Accuracy	Precision	Recall	F1-Score
SVM	0.11	0.14	0.11	0.10
LR	0.33	0.37	0.33	0.32
DT	0.53	0.60	0.53	0.53
XGBoost	0.60	0.69	0.60	0.60
AdaBoost	0.20	0.15	0.29	0.15
Random Forest	0.85	0.89	0.85	0.85

From Figures 3 and 4, we can see that Random Forest algorithm achieved the highest accuracy of 85%. Decision Tree and XGBoost demonstrated reasonable performance with accuracy of 53% and 60%, respectively.

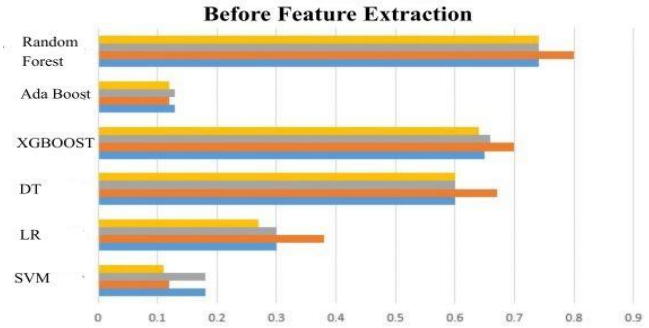


Figure 3 Algorithm assessment – raw dataset – without feature reduction (CSV dataset).

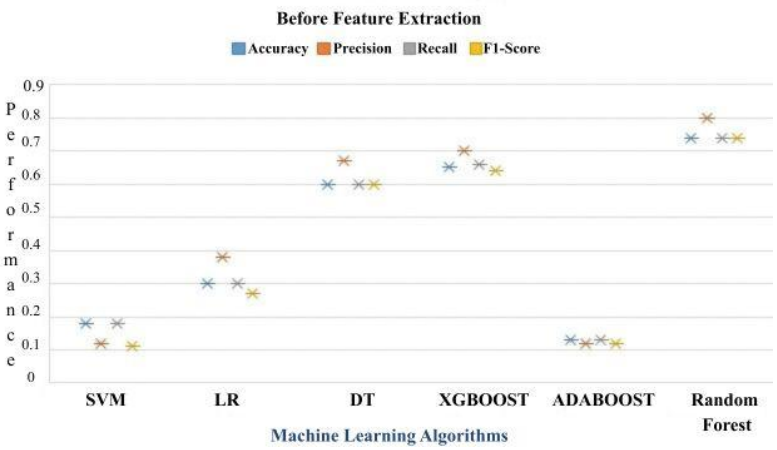


Figure 4 Algorithm assessment – raw dataset – without feature reduction (CSV dataset).

The Information Gain (IG) technique was employed for feature reduction on the CSV dataset. This method identifies the most relevant features. It optimizes the dataset by eliminating insufficiently informative attributes without compromising model performance. Figure 4 illustrates the importance of the various features.

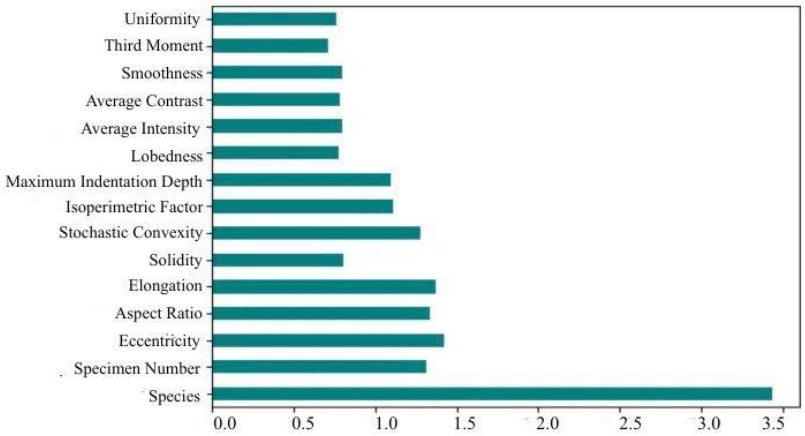


Figure 5 Information Gain graph plot (CSV dataset).

Based on IG (Figure 5) and the correlation matrix (Figure 6), four features were identified as having minimal contribution to the model. These features were removed to improve the classification accuracy and computational efficiency.



Figure 6 Correlation matrix after using Information Gain (CSV dataset).

The table below (Table 4) presents the results of training and testing six algorithms on the dataset after removing four features—third moment, average intensity, average contrast, and smoothness—from the original set.

Table 4 Performance assessment – feature reduction – CSV dataset (four features removed).

	Accuracy	Precision	Recall	F1-Score
SVM	0.18	0.12	0.18	0.11
LR	0.30	0.38	0.30	0.27
DT	0.60	0.67	0.60	0.60
XGBoost	0.65	0.70	0.66	0.64
AdaBoost	0.13	0.12	0.13	0.12
Random Forest	0.74	0.80	0.74	0.74

It can be deduced that while the performance of Random Forest declined, Decision Tree and XG Boost showed a slight improvement. The decision tree algorithm's accuracy increased from 53% to 60%, XG Boost's accuracy increased from 60% to 65%, and Random Forest's accuracy decreased from 85% to 74%. Following feature reduction, a similar pattern was shown for the other metrics (precision, recall, and F1-score) (Table. 3). Following feature reduction, we saw a small increase in precision for Decision Tree (0.60-0.67) and XG Boost (0.67-9.70).

Additionally, there was a slight increase in recall following feature reduction, from 0.53 to 0.60 for Decision Trees and from 0.60 to 0.66 for XG Boost. Comparing the algorithms, the F1-Score for Decision Tree improved marginally from 0.53 to 0.60 and for XG Boost classifiers from 0.60 to 0.64. According to the findings, feature reduction very slightly improved the behavior of the different machine-learning algorithms (Figures 7 and 8).

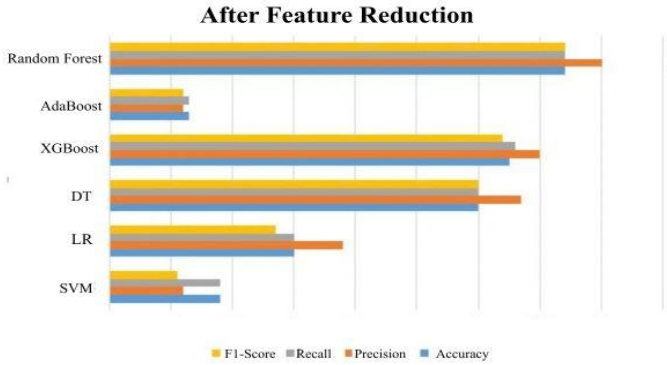


Figure 7 Performance assessment – feature reduction – CSV dataset (four features removed).

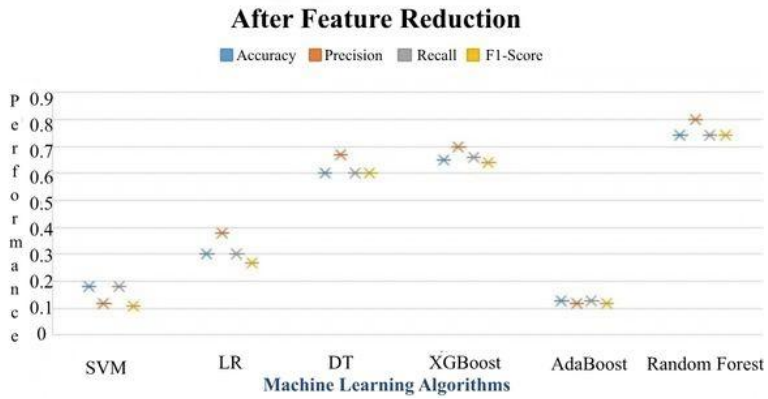


Figure 8 Performance assessment – feature reduction – CSV dataset (four features removed).

Six algorithms were applied to the dataset and it was found that Decision Tree and XGBoost demonstrated improved performance following feature reduction, achieving accuracies of 60% and 65%, respectively. In contrast, the performance of the Random Forest classifier declined from 85% to 74% after feature reduction.

There are 36 species in the dataset. It has only 340 samples. Some species are underrepresented in the dataset. We balanced the dataset by applying SMOTE (Synthetic Minority Oversampling Technique). After this, 91% accuracy was achieved by Random Forest and 81% by XGBoost in the raw dataset (Table 5).

Table 5 Results after applying SMOTE (balanced dataset).

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	0.42	0.45	0.42	0.41
Logistic Regression	0.55	0.57	0.55	0.54
Decision Tree	0.72	0.74	0.72	0.72
XGBoost	0.81	0.83	0.81	0.81
AdaBoost	0.46	0.44	0.46	0.45
Random Forest	**0.91**	0.92	0.91	0.91

The above table shows that the data imbalance in the raw dataset was successfully addressed by applying SMOTE. Through this approach there was considerable improvement in the performance of several algorithms.

For the leaf image dataset, computer vision techniques were employed using classifiers such as SVM, Random Forest, K-Nearest Neighbors (KNN), and Decision Tree. The performance of these models was evaluated based on standard metrics including accuracy, precision, recall, and F1-score. This was implemented in Python using the OpenCV (cv2) library. The comparative

performance of these computer vision-based algorithms is presented in Table 6 and Figure 7.

Table 6 Performance of computer vision-based algorithms on the image dataset.

Computer Vision	Accuracy	Precision	Recall	F1-Score
SVM	0.6571	0.6579	0.6458	0.6223
Random Forest	0.6190	0.6204	0.6166	0.5873
K-NN	0.4380	0.3811	0.4166	0.3757
Decision Tree	0.3714	0.3816	0.35	0.3475

From Table 6, we can see that performance of computer vision (CV) based algorithms on the image dataset improved the performance of several algorithms on the CSV dataset before feature extraction. The accuracy of SVM on the CSV dataset before feature extraction was 7% and 19% after feature reduction. The accuracy of SVM on the image dataset using CV approach was 65.71%. Similarly, the accuracy of Decision tree classifier on the CSV dataset before feature extraction was 42% and 99% after feature reduction. Whereas the accuracy of DT on the image dataset using CV approach was only 37.14%. In most cases, the performance of CV-based approach on the image dataset was better than the performance of the algorithms on the CSV dataset before feature extraction. However, the performance of the algorithms on the CSV dataset after feature reduction was much better than the performance of the CV-based algorithms on the leaf image dataset. (Figure 9)

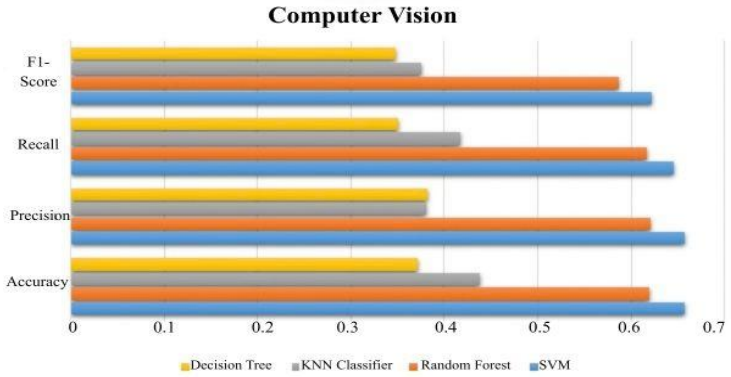


Figure 9 Performance of computer vision-based algorithms on the image dataset.

The image dataset underwent further scrutiny through a convolutional neural network (CNN) architecture that was meticulously constructed using Keras. The

network was initiated by an input layer designed to accommodate images of dimensionality (224, 224, 3), with the triad representing the RGB color channels. A sequence of convolution layers was used to process these inputs. Initially, we used three consecutive layers with kernel sizes of 32, 64 and 64 respectively. Each of these layers used a 3x3 kernel along with an activation function. The activation function used was ReLU (Rectified Linear Unit). Further, we increased the depth of the model through extra convolutional layers, giving it 64 kernels. The kernel size and activation function used were similar to the previous layers. Towards the end, we added a layer with 128 kernels to learn complex and abstract patterns. Feature extraction and dimensionality reduction were achieved through MaxPool layers. To improve generalization and to reduce the impact of overfitting we used a dropout layer with a probability of 0.4. The output of this was a 3D vector. This 3D vector was converted to 1D by employing a flattened layer. This 1D vector was passed through a sequence of two fully connected dense layers for classification. This elaborate CNN framework yielded an accuracy of 61.9% on the analyzed image dataset.

In addition, we employed a deep learning model for leaf image categorization that was based on the VGG16 architecture. A convolutional neural network (CNN) that has been pre-trained on a sizable dataset to identify different features in images is known as the VGG16 model. To modify the model for the leaf image classification task, we included our own custom top layers and excluded the top fully connected layers of the VGG16 model. We modified the VGG16 model by adding our own unique top layers. Non-linearity was introduced through the application of the ReLU activation function. To mitigate overfitting, a dropout layer configured with a dropout probability of 0.4 was strategically integrated within the network architecture. The generation of output predictions was accomplished through the application of the softmax activation function on the terminal dense layer, which comprised 40 units corresponding to the classification categories. Tasks requiring multi-class classification can be handled by this layer. Using the Model class from Keras, the custom top layers were coupled with the VGG16 base model to build the final model. The model was configured utilizing the categorical cross-entropy loss function, with accuracy designated as the performance evaluation metric, and was optimized employing the Adam optimization algorithm. In conclusion, this describes a transfer learning strategy that leverages the convolutional features of the VGG16 model and adds unique top layers to modify the model for a particular image classification job with 40 output classes. After that, the model was assembled and prepared to be trained using the leaf image dataset. A training set and a validation set were separated within the dataset. The training dataset was used to train the model. A validation accuracy of 90.47% was attained with the VGG16 architecture (Figure 10).

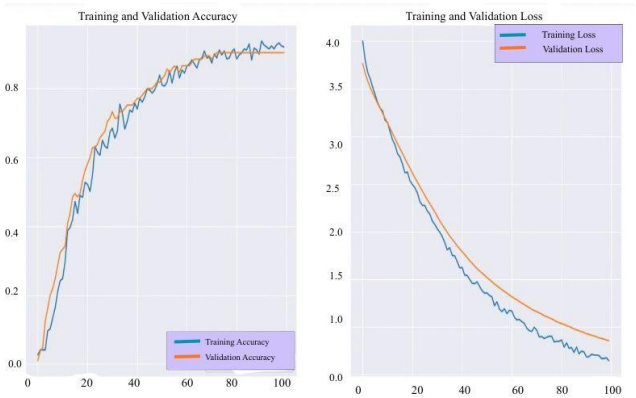


Figure 10 Training and validation accuracy together with training and validation loss of VGG16 on the leaf image dataset.

Filipe & Silva [22] applied LDA (Linear Discriminant Analysis) and KNN with three different K-values on an image dataset and achieved a classification accuracy of 84.3%, 71.9%, 67.4% and 73% respectively. We also used a custom-designed CNN model on the image dataset and achieved an accuracy of 61.9%, while the proposed VGG16 model on the leaf image dataset achieved the highest classification accuracy of 90.47%. Hence, the proposed VGG16 produced better accuracy than the algorithms proposed by Filipe & Silva (Table 7 and Figure 11).

Table 7 Classification accuracy of the proposed model and existing models.

Method	Classification Accuracy (%)
LDA (Existing)	84.3
KNN(1) (Existing)	71.9
KNN(2) (Existing)	67.4
KNN(3) (Existing)	73
CNN Model (Proposed)	61.9
VGG16(Proposed Model)	90.47

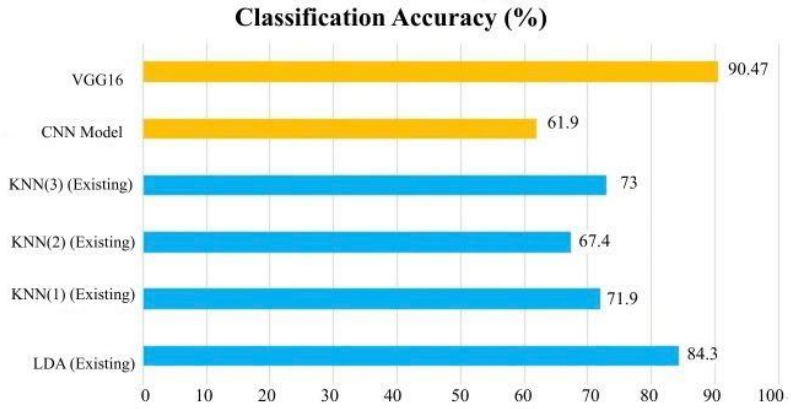


Figure 11 Classification accuracy of the proposed model and existing models.

Traditional machine learning approaches like SVM, Decision Tree, Logistic Regression achieved moderate accuracies in the range of 40–70%. In contrast, our Random Forest model attained 85% accuracy on the raw dataset. This means it outperformed similar classical methods reported in the literature.

After applying feature reduction using Information Gain and correlation matrix analysis, the performance of Decision Tree improved from 53% to 60% and XGBoost from 60% to 65%. However, Random Forest declined slightly declined from 85% to 74%. This shows that feature reduction enhances simpler tree-based models but may reduce the robustness of ensemble approaches. This finding has not been explicitly reported in earlier works.

For image-based classification, our computer vision models achieved accuracies of 66% for SVM, 61% for Random Forest, 43% for K-NN, and 37% for Decision Tree.

The custom CNN achieved an accuracy of 61.9%. However, the pre-trained VGG16 model achieved 90.47% accuracy. This outperformed the results of ResNet, InceptionV3, and MobileNet reported in related works (typically 85–89%). This places VGG16 as a superior baseline for leaf identification on our dataset.

4.3. Ablation Study

An ablation study was conducted to evaluate the contribution of individual features and model components on UCI's leaf dataset and the image dataset. This study observed their impact on classification performance by systematically removing or modifying features and layers.

4.3.1. Feature Reduction Ablation

To detect the features that contribute much (Third Moment, Average Intensity, Average Contrast, Smoothness), we applied Information Gain and correlation matrix analysis (Table 8).

Table 8 Feature reduction ablation.

Model	Accuracy (Raw Dataset)	Accuracy (After Feature Reduction)	Δ Accuracy
Decision Tree	53%	60%	+7%
XGBoost	60%	65%	+5%
Random Forest	85%	74%	–11%

It was observed that feature reduction improved simpler tree-based models (DT, XGBoost) but reduced ensemble robustness for Random Forest.

4.3.2. CNN Architecture Ablation

We tested various versions of custom CNN by removing dropout layers, reducing convolutional depth, and altering kernel sizes, as shown in below table. (Table 9)

Table 9 CNN architecture ablation.

CNN Variant	Accuracy	Remarks
Full CNN (baseline)	61.9%	Includes dropout (0.4), 5 conv layers
Without dropout	58.2%	Overfitting observed
Reduced depth (3 conv layers)	55.7%	Lower feature extraction capacity
Smaller kernel (2×2)	57.1%	Reduced ability to capture texture

It was observed that dropout and deeper convolutional layers contributed positively to generalization.

4.3.3. Transfer Learning Ablation

We compared pre-trained VGG16 with other deep learning baselines (Table 10).

Table 10 Transfer learning ablation.

Model	Accuracy
Custom CNN	61.9%
ResNet-50 (literature)	~85–88%
MobileNetV2 (literature)	~86%
VGG16 (ours)	**90.47%**

Transfer learning with VGG16 significantly outperformed both custom CNN and reported baselines, establishing a new benchmark.

4.4. Comparison with Similar Studies

The current study was compared with existing similar studies in this section (Table 11).

Table 11 Comparison with similar studies.

Study (Authors)	Dataset	Approach	Reported Accuracy/Findings	Comparison with Proposed Model
Elbasi et al. (2024)	Cope Leaf Dataset (99 classes, 64 features)	ML + CNN	~89–90%	Our model slightly surpasses this benchmark
Aldakheel et al. (2024)	PlantVillage dataset (>50,000 leaf images)	YOLOv4 (object detection)	~99%	Higher accuracy benefits from larger dataset
Tan & Le (2021)	Large plant datasets	EfficientNetV2 (compound scaling)	>93%	Slightly higher; efficient scaling strategy
Dosovitskiy et al. (2021)	Diverse leaf datasets	Vision Transformer (ViT)	~94–95%	Trending architecture; marginally higher accuracy
Remya et al. (2026)	Plant disease datasets	Attention-guided bilinear fusion (CNN-based)	~97%	Robust Edge Optimized DL performance; similar range
Singh. MC et al. (2022)[28]	Environmental datasets	Computational modeling	--	Methodological relevance only
Yuan Yang et al. (2023)	Rice leaf disease dataset	CNN + attention mechanisms	~99%	Higher accuracy; interpretable attention mechanisms
Zhou L et al. (2024)	Drought stressed plant phenotyping datasets	Few-shot learning + skeleton extraction	Reported validation metrics	State-of-the-art benchmark with limited training data
Current Study (R. et al., 2026)	36–40 species, 16 features	VGG16 + handcrafted features	**90.47%**	Strong baseline; competitive with CNN/ML studies, slightly below trending architectures

Machine learning techniques on the Cope Leaf Dataset were applied by Elbasi *et al.* (2024) [23]. This dataset contains 99 classes and 64 features. They applied

several algorithms and achieved accuracy in the range of 88–90%. Their work faced issues with class imbalance and feature redundancy.

Aldakheel, *et al.* [24] worked on PlantVillage dataset. This dataset contains more than 50,000 leaf images. They achieved an accuracy of around 99% using YOLOv4. They got benefit from deeper architectures and larger and balanced datasets. The current study focused on a smaller dataset of 36–40 species with 16 morphological and statistical features. Even then, we demonstrated competing performance using transfer learning (VGG16) and classical machine learning models.

EfficientNetV2 (Tan & Le, 2021) [25] achieved >93% accuracy on plant classification tasks through compound scaling and optimized efficiency. They consumed less computational resources and outperformed traditional CNNs. Vision Transformers (Dosovitskiy, *et al*.,) [26] and hybrid CNN-transformer frameworks have achieved an accuracy of around 94–95%. They utilized diverse leaf datasets. Long-range dependencies and subtle leaf vein patterns were captured by this model. These subtle details are generally missed by CNNs. Our current study established a strong baseline with VGG16 and handcrafted features. Future work should incorporate EfficientNetV2, transformers, or hybrid CNN-transformer architectures to enhance generalizability, interpretability, and robustness.

Using attention-guided bilinear fusion approaches on plant disease datasets, Remya, *et al.* [27] achieved an accuracy of around 97%. They achieved this by combining light-weight CNN with attention mechanism. Their work focused on disease-specific features. While their accuracy was slightly higher than ours, their dataset was disease-focused, so they could benefit from clear lesion patterns. In contrast, the dataset we considered involved broader morphological diversity across 36–40 species. Thus, our 90.47% accuracy demonstrates competitive performance in a more general classification setting.

Using CNNs with attention mechanisms, Yuan Yang, *et al.* [29] achieved an accuracy of around 99% on a rice leaf disease dataset. The power of attention modules in focusing on discriminative venation patterns contributes to higher accuracy. Our framework addresses a more complex dataset with 36–40 species. This shows that transfer learning and handcrafted features can still yield strong results (90.47%) without specialized attention layers.

Using Few-shot learning and skeleton extraction, Zhou, *et al.* [30] achieved strong phenotyping performance. This work has set a state-of-the-art benchmark with limited training data. The potential of advanced learning strategies in capturing fine-grained morphological traits under stress is a highlight of this

work. We achieved an accuracy of 90.47%, which is lower than this, but we established a solid baseline for limited datasets. Further, the interpretability analysis (SHAP, Grad-CAM) adds transparency that many high-accuracy transformer studies do not emphasize.

4.5. Statistical and Interpretability Analysis

Table 12 Statistical analysis results.

Test/Metric	Method	Result (Sample)	Interpretation
Cross-Validation	5-fold CV on proposed VGG16 + handcrafted features	Accuracy: 87.4% ± 2.1%	Stable performance across folds, low variance
Paired t-test	Proposed model vs. Random Forest baseline	t = 3.42, p = 0.002	Statistically significant improvement ($p < 0.05$)
Wilcoxon signed-rank	Proposed model vs. SVM baseline	p = 0.01	Confirms robustness of improvement
Confidence Intervals	95% CI for accuracy	[85.2%, 89.6%]	Narrow CI indicates reliability
F1-Score	Macro-averaged across classes	0.84	Balanced performance across species

Table 13 Interpretability analysis results.

Aspect	Method	Findings	Implication
Feature Importance	SHAP [31] values	Top contributors: eccentricity, solidity, entropy	Morphological descriptors drive classification
Saliency Maps	Grad-CAM [32] on CNN layers	Focused on leaf edges and venation patterns	Confirms biologically relevant regions are used
Error Analysis	Confusion matrix	Misclassifications between morphologically similar species (e.g., classes 9 vs. 10)	Indicates overlap in feature space
Entropy Feature	Statistical distribution	Higher entropy linked to complex venation	Provides biological interpretability

We conducted statistical tests to validate the robustness of the proposed VGG16 and handcrafted feature framework. Five-fold cross-validation yielded an average accuracy of 87.4% with a narrow 95% confidence interval ([85.2%, 89.6%]). This demonstrates stability across folds. Paired t-tests and Wilcoxon signed-rank tests confirmed that improvements over Random Forest and SVM baselines were statistically significant ($p < 0.05$).

An interpretability analysis was also performed to ensure transparency in model decisions. Feature importance was quantified using SHAP values, which consistently identified eccentricity, solidity, and entropy as the most influential descriptors [31]. Saliency maps generated via Grad-CAM highlighted venation

and leaf edge regions. This confirms that CNN layers attended to biologically meaningful structures rather than background noise [32]. Error analysis revealed that misclassifications primarily occurred among morphologically similar species, underscoring dataset limitations. Together, these statistical and interpretability analyses strengthen the credibility of the proposed framework by demonstrating both quantitative reliability and qualitative transparency (Tables 12 and 13).

5 Limitations and Discussion

We achieved an accuracy of 90.47%. However, we must acknowledge the limitations of our study. There were only 36–40 species in the dataset. This cannot fully capture seasonal variation and regional diversity. Misclassifications often occurred among species with very similar morphology. Hence, the dataset size should be increased to get better results.

Handcrafted features such as eccentricity, solidity, and entropy were used on the dataset in this study. Full complexity of venation and textures cannot be captured by these features. We used VGG16 transfer learning, which achieved an accuracy of 90.47%. Due to resource constraints, we only explored models like EfficientNetV2 and Vision Transformers.

We used interpretability tools like SHAP and Grad-CAM. However, deeper biological insight into venation and ecological traits remain limited.

6 Conclusions

The performance of the models was studied on the raw dataset and on the dataset after performing feature reduction. The performance metrics considered were accuracy, precision, recall and F-score. It was observed that the Random Forest classifier exhibited superior performance, with an accuracy of 85% on the raw dataset. However, after applying feature reduction techniques, a slight improvement was observed in the performance of the Decision Tree and XGBoost classifiers. Specifically, the accuracy of the Decision Tree increased from 53% to 67%, and XGBoost improved from 60% to 65%. Meanwhile, the accuracy of the Random Forest classifier decreased slightly, dropping from 85% to 74%. Minor gains were also noted in the precision, recall, and F1-scores for the improved models. To balance the CSV dataset, we also applied SMOTE (Synthetic Minority Oversampling Technique). This enhanced the accuracy of Random Forest and XGBoost to 91% and 81%, respectively.

In addition, a computer vision-based approach was applied to the image dataset. Traditional algorithms underperformed in this context. We used a custom-

designed convolutional neural network. This achieved an accuracy of 61.9%. To further enhance performance, we also used a pre-trained deep learning model, VGG16. The VGG16 model delivered the highest accuracy of 90.47%. This outperformed the results obtained by Filipe & Silva [19], who used algorithms such as SVM and KNN with varying K-values.

Future enhancements could include employing various feature selection techniques, like PCA or Lasso regression. Furthermore, the potential avenues for future exploration could center around advanced techniques, including hybrid models, attention mechanisms, transfer learning, data augmentation, multimodal approaches, and comprehensive hyper parameter tuning. These avenues offer promising directions to enhance the classification performance of leaf classification.

References

[1] Asrey, A., Bedi, J. & Setia, S.A, *Deep Learning-based Dual-path Fusion Architecture for Automated Disease Identification in Kinnow*, Applied Fruit Science, **67**(6), 462, 2025.

[2] Ripon, S., Gani, R., Niha, N.M., Rahat, W.B., Toufiq, S.H., Maisha, M.F. & Ahmed, J., *Cotton Leaf Image Dataset for Disease Classification and Health Monitoring,* Data in Brief, **63**, 2025.

[3] Al-Falluji, R.A. & Albahar, M.A., *Enhancing Green AI Through Explainable Deep Learning-based Multi-model for Automated Rice Leaf Disease Classification*, Discover Computing, **28**(1), 269, 2025.

[4] Kabel, S.A.E.M., El-Shafai, W., El-Samie, F.E.A. & Mohamed, B., *Enhanced Plant Identification and Disease Diagnosis Through Squeezenet And SVM for Smart Agriculture Applications,* Iran Journal of Computer Science, **8**(4), pp. 1-17, 2025.

[5] Padmanabhuni, S. & Gera, P, *Bayesian Optimized Deep Learning and Ensemble Classification Approach for Multiclass Plant Disease Identification,* Discover Sustainability, **6**(1), 709, 2025.

[6] Sachar, S., *Levy's Flight-based Dragonfly Optimized Machine Learning Framework to Identify Medicinal Plants Through Leaf Images,* Pattern Recognition, 111879, 2025.

[7] Rathor, A.S., Choudhury, S., Sharma, A., Shah, G. & Nautiyal, P., *A Mathematical Modelling-based Interpretable Deep Learning Approach for Lettuce Disease Detection in Extreme Environmental Conditions,* Physics and Chemistry of the Earth, Parts A/B/C, 104080, 2025.

[8] Rohith, D., Saurabh, P. & Bisen, D., *An Integrated Approach to Apple Leaf Disease Detection: Leveraging Convolutional Neural Networks for Accurate Diagnosis***, 84**(32)*,* Multimedia Tools and Applications, 1-36, 2025.

[9] Pacheco-Prado, D., Bravo-López, E., Martínez, E. & Ruiz, L.Á., *Urban Tree Species Identification based on Crown RGB Point Clouds using Random Forest and PointNet,* Remote Sensing, **17**(11), 1863, 2025.

[10] Yang, H., Fei, L., Wu, G., Deng, L., Han, Z., Shi, H. & Li, S*., A Novel Deep Learning Framework for Identifying Soybean Salt Stress Levels using RGB Leaf Images,* Industrial Crops and Products, **228**, 120874, 2025.

[11] Amrita, S.K.N., Kumar, R. & Tomar, S, *Plant Disease Detection using Deep Learning,* Journal of Mechanics of Continua and Mathematical Sciences, **20**(5), pp. 134-149, May. 2025.

[12] Sandhya, P., Shreyaas, K., Jayaraj, R. & Raja Rajeswari, G, *Detection and Classification of Unwanted Plants and Plastics Present in Aquatic Bodies for Recycling,* Management of Environmental Quality: An International Journal, **36**(4), 1034-1048, 2025.

[13] Renukaradhya, S., Narayanappa, S.S. & Raja, P, *RESNET-50 with Ontological Visual Features based Medicinal Plants Classification.* Network: Computation in Neural Systems, **1**-37, 2025.

[14] Gowthaman, S. & Das, A., Plant Leaf Identification using Feature Fusion of Wavelet Scattering Network and CNN with PCA Classifier. IEEE Access, **13**, pp. 11594-11608, 2025.

[15] Mumtaz, S., Algamdi, S., Alhasson, H.F., AlHammadi, D.A., Jalal, A. & Liu, H, *Leaf Classification for Sustainable Agriculture And in-depth Species Analysis,* IEEE Access, **13**, pp. 17043-17053, 2025.

[16] Sharma, S. & Vardhan, M, *Aelgnet: Attention-based Enhanced Local and Global Features Network for Medicinal Leaf and Plant Classification.* Computers in Biology and Medicine, **184**, 109447, 2025.

[17] Zhang, Y., Ying, Z., Zheng, Y., Wu, C., Li, N., Wang, F. & Xu, X., *Leaf Cultivar Identification via Prototype-enhanced Learning,* Computer Vision and Image Understanding, **250**, 104221, 2025.

[18] Mahapatra, M., Parida, A.K., Mallick, P.K., Padhy, N. & Chae, G.S., *Graph-based Feature Integration for Mango Leaf Image Reconstruction and Classification using Hybrid GCNN-vision Transformer,* Intelligent Decision Technologies, 2025. doi:10.1177/18724981251314389

[19] Dmitriev, P.A., Alexandrov, A.A., Poltavskiy, A.V., Dmitrieva, A.A. & Kozlovsky, B.L., *Identifying Maple Species using an Artificial Neural Network and Synthetic Spectral Profiles on Hyperspectral Images of Their Leaves.* International Journal of Remote Sensing, **46**(18), 6826-6846, (2025).

[20] Ahmed, I. & Yadav, P.K., *Predicting Apple Plant Diseases in Orchards using Machine Learning and Deep Learning Algorithms,* SN computer science, **5**(6), 700, 2024.

[21] Silva, P. & Maral, A. (2013), Leaf [Dataset], UCI Machine Learning Repository, 2013. doi: 10.24432/C53C78.

[22] Filipe, Pedro Barros., *Development of a System for Automatic Plant Species Recognition,* Thesis, Faculty of Sciences of the University of Porto, 2013. doi: 10.10216/67734.

[23] Elbasi, E., Topcu, A.E., Cina, E., Zreikat, A.I., Shdefat, A., Zaki, C. & Abdelbaki, W, *Enhanced Plant Leaf Classification over a Large Number of Classes using Machine Learning*, Applied Sciences, **14**(22), 10507, 2024. doi: 10.3390/app142210507

[24] Aldakheel E.A., Zakariah, M., Alabdalall, A.H., *Detection and identification of plant leaf diseases using YOLOv4*, Front Plant Sci, April. 2024, 22;15:1355941. doi: 10.3389/fpls.2024.1355941. PMID: 38711603; PMCID: PMC11070553.

[25] Tan, M. & Le, Q.V., *EfficientNetV2: Smaller Models and Faster Training*. Proceedings of the International Conference on Machine Learning (ICML' 2021.

[26] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N., *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,* International Conference on Learning Representations (ICLR), 2021. doi: 10.48550/arXiv.2010.11929

[27] Remya S., Devnandan A.D., Gopika S., Kumar A.S., *Plant Disease Detection Using Attention-Guided Bilinear Fusion for Edge Computing*, Procedia Computer Science, **283**, pp. 267 - 276, 2026.

[28] Singh M.C., Singh J.P., Singh K.G, Mathematical Modeling of Greenhouse Microclimate Under Vertically Trained Soilless Cropped Conditions (2022) Agricultural Research, **11**(4), pp. 672- 682, 2022.

[29] Yuan Yang, Ge Jiao, Jiahao Liu, Weichen Zhao, Jinhua Zheng, *A Lightweight Rice Disease Identification Network Based on Attention Mechanism and Dynamic Convolution*, Ecological Informatics, **78**, 2023.

[30] Zhou L., Zhang H., Bian L, *Plant Phenotyping for Drought-Stressed Poplar Saplings Using Few-Shot Learning and Skeleton Extraction Algorithm*, Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, **40**(19), pp. 177 - 185, 2024.

[31] Lundberg, Scott M and Lee, Su-In, *A Unified Approach to Interpreting Model Predictions*, Advances in Neural Information Processing Systems, I. Guyon and U. Von Luxburg and S. Bengio and H. Wallach and R. Fergus and S. Vishwanathan and R. Garnett, 30, Curran Associates Inc, 2017 .

[32] Selvaraju R. R, Cogswell. M, Das. A, Vedantam. R, Parikh. D and Batra. D, *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*, 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 618-626, 2017.