



## Comparisons of Machine Learning Versus Deep Learning Architectures for Plant Disease Classification

Erdem Uysal<sup>1\*</sup>, Osman Doğuş Gülgün<sup>1</sup>, Hamza Erol<sup>2</sup>

<sup>1</sup> Department of Computer Engineering, Institute of Science, Mersin University, Mersin 33110, Turkey

<sup>2</sup> Department of Computer Engineering, Mersin University, Mersin 33110, Turkey

Corresponding Author Email: [erdemuysal@mersin.edu.tr](mailto:erdemuysal@mersin.edu.tr)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430303>

### ABSTRACT

**Received:** 23 March 2026

**Revised:** 8 June 2026

**Accepted:** 21 June 2026

**Available online:** 30 June 2026

#### Keywords:

*plant disease detection, hybrid feature extraction, Support Vector Machine, XGBoost Classification, DINOv2, machine learning, deep learning*

Automated diagnosis of plant diseases plays a critical role in enhancing agricultural productivity. In this study, a high-performance hybrid classification framework is proposed to accurately distinguish among 38 distinct plant leaf disease categories. The architecture fuses deep semantic representations extracted from a pre-trained ResNet-50 network with handcrafted textural and morphological features. To optimize computational efficiency and prevent overfitting, the fused feature space was reduced to 256 dimensions using Principal Component Analysis (PCA). The proposed framework was extensively benchmarked against advanced state-of-the-art architectures. While baseline models using only handcrafted features exhibited limitations, deep learning models such as DINOv2, ResNet-152+CBAM and YOLO11 demonstrated strong classification capabilities. Standing as a highly competitive alternative to these advanced architectures, the proposed hybrid model achieved a robust accuracy of 96.16%. The findings demonstrate that integrating deep and handcrafted features provides a competitive, robust, and computationally efficient solution for complex agricultural image classification tasks.

## 1. INTRODUCTION

The identification and classification of plant leaf diseases is an important aspect of agricultural management. It significantly affects crop yield and food safety. As global agricultural practices evolve, the importance of timely and accurate identification of diseases has increased [1]. The development and use of machine learning (ML) and deep learning (DL) methods in technology have become widespread. It has opened new avenues for automating this process. It has offered the potential to increase the accuracy of disease detection [2]. This comparative study examines various ML and DL approaches for identifying and classifying plant leaf diseases using the New Plant Disease dataset (<https://www.kaggle.com/datasets/vip00000l/new-plant-diseases-dataset>). By examining this dataset, the strengths and weaknesses of different methodologies have been highlighted. To this end, insights into the effectiveness of each approach have been provided. In addition to the ML and DL methods applied in this study, data preprocessing and cluster number determination processes were performed on the dataset. The results of this study were evaluated through performance metrics and comparative analysis of the datasets. A comprehensive understanding of plant disease recognition and classification is presented. Important insights were identified in this research, challenges encountered were highlighted, and recommendations for future research were provided.

The importance of automated systems for the detection and classification of plant diseases is emphasized in order to

increase agricultural productivity and ensure food security in the face of growing global demands [3].

In the study conducted by Gupta et al. [4], ML is applied to the Pear and Watermelon datasets, and 2,994 features were used for classification. These features were generated by adding the mean and standard deviation of each channel RGB, as well as Gray-Level Co-occurrence Matrix (GLCM) and Histogram of Oriented Gradients (HOG) features. The study achieved a success rate of 93.4%. Additionally, class imbalance is addressed using the Synthetic Minority Over-sampling Technique (SMOTE).

Another study [5] presents a new plant leaf identification approach that combines shape and texture features. The proposed Multiscale Triangle Descriptor (MTD) is used to characterize the leaf's shape information, while the Local Binary Pattern Histogram Fourier (LBP-HF) is used as the texture feature. Shape and texture features are combined with weighted distance measurement; the distance is used for shape, and the Chi-square distance is used for texture. The method is tested on the Flavia, Swedish, and MEW2012 datasets. It achieved classification accuracies of 99.16%, 98.48%, and 95.64%, respectively. The results are better than or comparable to existing modern methods.

While recent studies have successfully applied architectures such as EfficientNet, SimCLR, and ConvNeXt to plant disease datasets, these works primarily focus on maximizing raw accuracy, often overlooking the deployment constraints of agricultural edge devices. Furthermore, existing literature rarely quantifies the inherent statistical complexity of the

datasets used for evaluation. To address these gaps, the specific contributions of this study are threefold: (1) proposing a hybrid feature-fusion framework that combines deep ResNet-50 representations with handcrafted descriptors for accurate yet lightweight classification; (2) providing a statistical characterization of dataset complexity using GMM with the AIC and BIC criteria; and (3) establishing a comparative benchmark that explicitly measures the trade-off between classification accuracy and model storage size. This study provides a practical engineering perspective on selecting suitable architectures for resource-constrained agricultural systems by evaluating both ML workflows that combine features derived from deep feature extraction with manually crafted features, as well as modern DL models.

Gil et al. [6] tested various DL architectures on two different datasets. The proposed ConvNeXt architecture achieved accuracy rates of 99.47% and 93.83% on the Plant Village and Plant Pathology datasets, respectively, following hyperparameter optimization using the grid search method. Galabuzi et al. [7] achieved a 99.65% accuracy using the EfficientNet architecture with the New Plant Disease image dataset. Another study using the New Plant Disease dataset is the FL-EfficientNet model. The proposed FL-EfficientNet model solved the sample imbalance problem in plant disease

classification using Focal Loss and reduced overfitting using DropConnect. It demonstrated better performance than other comparison networks with an average accuracy rate of 99.72% and fast convergence time. These features provide a strong foundation for deploying the model on mobile devices [8, 9]. Apleni et al. [10] proposed an ensemble model for the dataset. A success rate of 97.00% is achieved by utilizing the features of three different CNN architectures. Another study using this dataset examines the effectiveness of self-supervised learning (SSL) using contrastive learning of visual representations. Unsupervised pre-training scenarios on unlabeled plant images were investigated. The results showed that the performance of SSL in plant disease detection is comparable to that of traditional supervised learning approaches [11]. While DL models typically require massive datasets, integrating targeted segmentation techniques can mitigate this dependency. For instance, a recent study demonstrated that employing FCM-based segmentation to isolate diseased regions allowed a custom CNN architecture to achieve a 98% classification accuracy on apple leaf diseases without relying on a large-scale training set [12].

Relative studies and their characteristics are summarized in Table 1.

**Table 1.** Studies using the New Plant Disease dataset and its features

| Reference | Primary Model / Transfer or Pretrained                                                            | Notable Modifications or Modules                           | Reported Top Accuracy                   |
|-----------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------|-----------------------------------------|
| [6]       | ConvNeXt<br>Fine-tuned pretrained model                                                           | Grid Search to optimize hyperparameters                    | 99.47% and 93.83%                       |
| [7]       | EfficientNetB0<br>Fine-tuned with Neural Architecture Search (NAS)                                | ENet [8]                                                   | 99.65% on the New Plant disease Dataset |
| [9]       | FL-EfficientNet (EfficientNet family variant)<br>NAS-informed design (custom)                     | Moving flip bottleneck conv, attention modules, Focal loss | 99.72% on the New Plant disease Dataset |
| [10]      | Ensemble-based Feature Fusion<br>Pretrained CNNs                                                  | VGG16, ResNet-50, InceptionV3                              | 97.00% on the New Plant disease Dataset |
| [11]      | SimCLR pretraining + downstream classifier<br>Self-supervised pretrain then supervised classifier | Contrastive pretraining pipeline (SimCLR)                  | Comparable to supervised baseline       |

## 2. PRELIMINARY INFORMATION

### 2.1 Traditional machine learning in agriculture

Traditional ML in agriculture relies on methods like Support Vector Machine (SVM), Random Forest and K-Nearest Neighbors (k-NN) for tasks such as disease detection, yield forecasting and soil analysis, using imagery, spectral, sensor and historical data. Random Forest and ensemble methods generally demonstrate superior performance. SVM widely used for image and spectral-based classification and binary disease detection due to robust decision boundaries for moderate feature dimensionality. Random Forest is used extensively for regression and classification (yield, crop recommendation, disease presence) because of its built-in feature selection and resistance to overfitting. Decision tree is used for interpretable rules and quick baseline models in crop and soil decision tasks. k-NN is applied in similarity-based classification and small-sample problems such as seed and crop type recognition [13]. Naive Bayes and Linear Discriminant Analysis (LDA) / Quadratic Discriminant Analysis (QDA) are used for lightweight classifiers on engineered features, especially where independence assumptions hold roughly [14]. XGBoost classifier is increasingly used for high-accuracy tabular predictions and

crop recommendation and selection problems [15]. Clustering and dimensionality reduction (k-means, Principal Component Analysis (PCA), Self-Organizing Map (SOM)) are used for unsupervised land-cover separation, anomaly detection and feature reduction prior to modelling [13, 16].

Future trends and research directions are Automated Machine Learning (AutoML) and pipeline optimization, transfer learning and hybrid pipelines that combine handcrafted features with deep feature extractors to work under limited-label regimes and wider use of ensembles and gradient boosting [15, 17, 18].

### 2.2 Convolutional neural networks

Deep convolutional neural networks (CNNs) have become highly prominent in agricultural image classification due to their superior feature extraction capabilities. The reported accuracies are high on controlled datasets. Common datasets are PlantVillage [19], Digipathos [20], Plant-DOC [21] and other public or institutional datasets. Recent advancements in smart farming have demonstrated the high efficacy of CNNs in multi-class plant pathology; for instance, the Krishi-Konect framework achieved an exceptional accuracy of 99.2% across 38 distinct crop and disease classes utilizing the PlantVillage dataset [22]. InceptionV3 outperformed VGG19 in the six-

class leaf disease task (InceptionV3  $\approx 0.97$  accuracy, VGG19  $\approx 0.83$ ) in a multi-model comparison [23]. To mitigate the high parameter count typical of deep architectures, Hassan and Maji [24] introduced a novel CNN utilizing depthwise separable convolutions and inception layers, achieving a remarkable 99.39% classification accuracy on the PlantVillage dataset while remaining computationally efficient. Also, a custom nine-layer Deep CNN model significantly outperformed traditional ML approaches, achieving a 96.46% accuracy rate on a comprehensive dataset [25].

### 2.3 The shift to Vision Transformers and self-supervision

Vision Transformers (ViTs) adapt Transformer encoders from natural language to images by splitting an image into patch tokens and applying multi-head self-attention to model global relations; positional embeddings and a classification token produce downstream predictions. This patch-token attention gives ViTs large representational capacity and less built-in locality bias than CNNs, which can help capture long-range agricultural patterns but increases reliance on suitable pretraining data [26, 27].

SSL trains models on proxy tasks using unlabeled images, producing task-agnostic feature extractors that reduce dependency on costly annotations. In agriculture, where labeled datasets are scarce, noisy, or costly, SSL enables pretraining on abundant unlabeled field or satellite images and improves downstream performance with far fewer labels [28, 29].

## 3. MATERIALS AND METHODS

The New Plant Disease image dataset (<https://www.kaggle.com/datasets/vipooool/new-plant-diseases-dataset>) consists of 38 classes. This dataset is recreated using offline augmentation from the original dataset. It consists of about 87,000 RGB images of healthy and diseased plant leaves, divided into 38 different classes. To strictly prevent data leakage and ensure robust evaluation, the dataset split (70% training, 20% validation, and 10% test) was conducted at the original image level prior to the application of offline augmentation techniques. This protocol guarantees that augmented variants or near-duplicate samples of the same source leaf do not cross-contaminate the validation and test sets, thereby confirming that the high accuracy rates are a true reflection of the models' generalization capabilities rather than dataset memorization. The training set contains 61,502 images, the validation set contains 17,574 images, and the test set contains 8,791 images.

The dataset contains a new directory with 33 test images for prediction purposes. Additionally, all images are  $256 \times 256 \times 3$  in size. Healthy and diseased leaf images from different plant species in the New Plant Disease dataset are shown in Figure 1.

In this study, dataset complexity is approached through the statistical density and structural distribution of the data rather than simple visual background clutter. To statistically evaluate this distribution, Gaussian Mixture Models (GMMs) were utilized as an unsupervised density estimation tool. Instead of direct class separability, we determined the optimal number of components using the Akaike Information Criterion (AIC) [30] and Bayesian Information Criterion (BIC) [31], which serve as established model selection criteria. The criteria are

defined as:

$$AIC = 2k - 2\ln(L) \quad (1)$$

$$BIC = \ln(n)k - 2\ln(L) \quad (2)$$

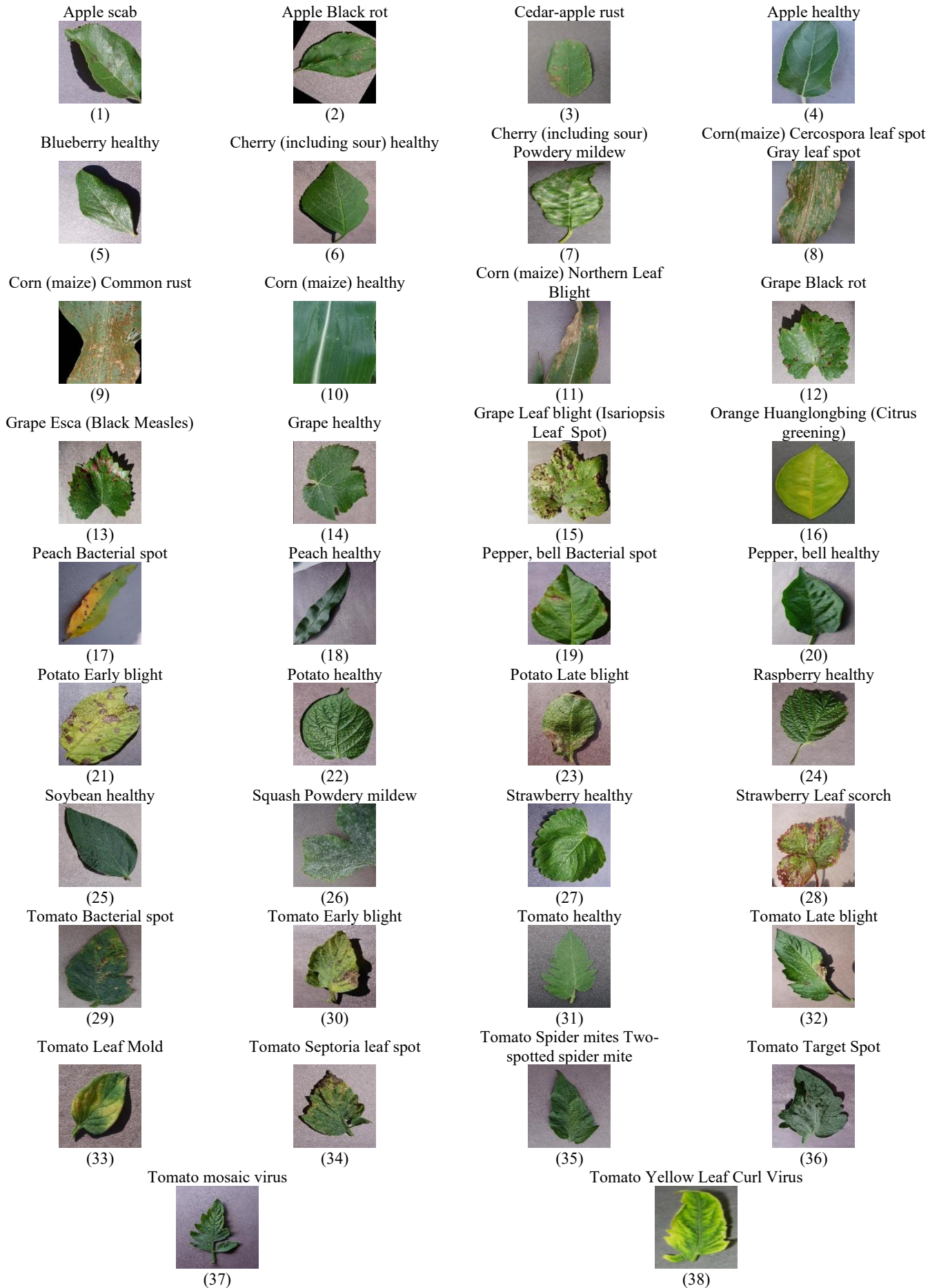
where,  $k$  is the number of parameters,  $n$  is the sample size, and  $L$  is the maximum value of the likelihood function. A steady decrease in these scores indicates an improved model fit to the underlying feature space without excessive penalization for complexity. As shown in Figure 2, the clear inflection point converging at 38 clusters aligns with the structural variance of the dataset, confirming that the latent feature space requires robust model architectures capable of handling high-dimensional representations.

For this purpose, one image is taken from each class and resized to  $224 \times 224$ . Subsequently, 38 images were added together (50176,3 each) to obtain a (1906688,3) dimensional matrix. To evaluate the structural distribution of the dataset, Gaussian Mixture Models (GMM) are employed as an unsupervised density estimation technique. The optimal number of Gaussian components is determined using the AIC and BIC scores, which strictly function as model-selection tools balancing goodness-of-fit against parameter complexity. The GMM is initialized using the k-means algorithm, with the number of components systematically tested across a range from 2 to 40. To ensure reproducibility and convergence stability, the model is configured with a full covariance matrix, a random seed of 123, and 10 random initializations ( $n\_init = 10$ ). The Expectation-Maximization (EM) optimization parameters included a convergence tolerance of 0.001, a regularization covariance of  $1e-06$ , and a maximum of 100 iterations. It is important to note that AIC and BIC do not directly measure spatial class separability or simplicity; rather, a relative decline and subsequent stabilization of these scores indicate a suitable probabilistic representation of the underlying feature space density without excessive overfitting. It should be noted that this analysis is intentionally performed at the raw pixel-level as a lightweight, model-agnostic screening of the dataset's global color-intensity distribution, rather than as a measure of disease-class separability. A more discriminative feature-space analysis — for example, applying the same GMM/AIC/BIC procedure to deep ResNet-50 embeddings — is left as a direction for future work. The breaks in the GMM created using AIC and BIC scores for the New Plant Disease image dataset are shown in Figure 2.

The New Plant Disease dataset has image complexity and computational complexity due to its big data characteristics and the large number of classes it contains. For these reasons, both ML and DL methods have been applied for object recognition in the image data.

The ML methods (i) XGBoost classifier [32], (ii) Decision Tree [33] and (iii) SVM [34] are utilized for the classification of the New Plant Disease image dataset.

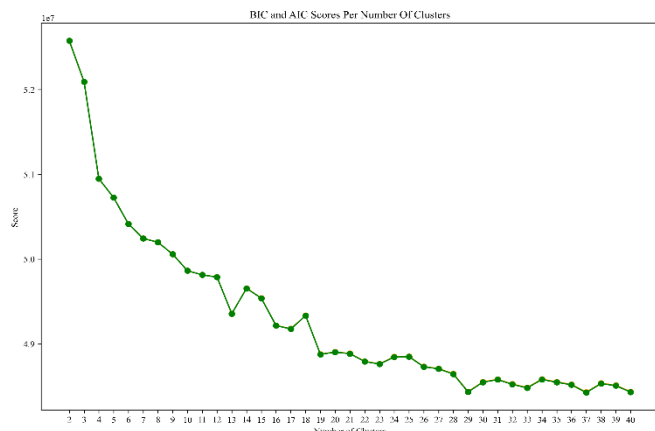
DL models (iv) ResNet-152 [35], (v) YOLO11 [36], and (vi) DINOv2 [37] were trained for the classification of the New Plant Disease image dataset. Unlike traditional supervised learning, DINOv2 operates on a self-supervised distillation principle. It utilizes a student-teacher network architecture where the student network learns to predict the output of the teacher network. The teacher network is updated as an exponential moving average (EMA) of the student weights, ensuring stable training representations without the need for negative pairs used in contrastive learning.



**Figure 1.** Healthy and diseased leaf images from different plant species in the New Plant Disease dataset

The training hyperparameters of the models are described in the implementation details section. Performance metrics,

including accuracy, precision, recall, and F1 score, were used to evaluate the effectiveness of each method across the datasets.



**Figure 2.** Breaks in the GMM created using AIC and BIC scores for the New Plant Disease image dataset  
 Note: GMM = Gaussian Mixture Model; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

## 4. IMPLEMENTATION DETAILS

### 4.1 Machine learning methods

In response to the necessity for a standardized and fair comparative baseline, the traditional ML pipeline is reconstructed utilizing a hybrid feature fusion strategy. Rather than relying exclusively on low-level visual descriptors, a pre-trained ResNet-50 architecture is deployed as a global deep feature extractor. Note that this ResNet-50 backbone is used solely as a fixed feature extractor for the hybrid machine-learning pipeline and is distinct from the ResNet-152+CBAM model evaluated as a standalone deep-learning classifier in Section 4.2. The high-dimensional semantic representations obtained from the deep backbone were subsequently concatenated with handcrafted structural descriptors—specifically, Haralick texture features [38], LBP [39], and basic color statistics. This fusion mechanism effectively bridges the gap between automated deep representation learning and traditional textural analysis. Finally, SVM classifier is trained on this concatenated feature space, ensuring that the comparative analysis between ML and DL architectures is rooted in an equivalent and fair foundation of feature quality.

To establish a baseline and demonstrate the necessity of integrating deep representations, an initial ablation study is conducted using the handcrafted feature set solely. The 123-dimensional vectors, encompassing GLCM Haralick textures, LBP descriptors, and color channel statistics, were standardized and fed into the XGBoost classifier, Decision Tree, and SVM. The Decision Tree and SVM classifiers were implemented using the sklearn library, while the XGBoost classifier was implemented using the XGBoost library.

#### 4.1.1 Feature extraction and fusion

The proposed hybrid classification framework is evaluated on a comprehensive dataset consisting of 38 distinct plant disease categories, partitioned into 61,502 training and 8,791 test samples. To capture both high-level semantic information and low-level textural details, a dual-stream feature extraction

strategy is employed. First, deep feature representations were extracted using a pre-trained ResNet architecture, generating a 2048-dimensional vector for each image. Concurrently, a robust set of handcrafted features, resulting in a 123-dimensional vector, is computed to capture specific morphological and textural characteristics. These two complementary feature sets were then concatenated at the feature level, resulting in a fused, high-dimensional feature vector of 2171 dimensions (2048 deep + 123 handcrafted) per sample.

#### 4.1.2 Dimensionality reduction

Given the high dimensionality of the fused feature space (2171 dimensions), feeding this data directly into a classifier would be computationally expensive and prone to the curse of dimensionality. To mitigate this, PCA is applied to the fused feature vectors. The feature space is successfully reduced to 256 principal components, which collectively retained 92.0% of the original variance. This strategic reduction significantly optimized the computational efficiency while preserving the most discriminative information required for accurate classification.

#### 4.1.3 Hyperparameter optimization

To ensure rigorous evaluation without data leakage, a 5-fold cross-validation Grid Search method is applied to the training set (61,502 samples) for each ML model.

Following hyperparameter optimization, the XGBoost model that demonstrated the best performance is evaluated on a previously unseen test dataset (8,791 samples). The model achieved an overall accuracy rate of 91.59%, demonstrating strong predictive capabilities. The macro-average and weighted-average F1 scores were recorded at 91.50% and 91.53%, respectively, indicating consistent performance even with an imbalanced class distribution. A detailed analysis of the classification metrics reveals exceptional recognition rates for specific categories; for instance, Corn Common Rust and Cherry Healthy achieved F1-scores of 98.27% and 98.11%, respectively. However, the model exhibited relative difficulty in distinguishing visually subtle or overlapping symptoms in certain classes, such as Tomato Septoria Leaf Spot (F1-score: 79.90%) and Pepper Bell Healthy (F1-score: 83.64%). Overall, the optimized XGBoost framework has proven to be a highly competitive classifier for complex multi-class plant pathology tasks.

In the optimization table for the decision tree, various structural parameters were examined, including the splitting criterion, the maximum tree depth, and the minimum number of samples required to enter an internal node or create a leaf node. Grid Search identified the optimal architecture using the ‘entropy’ criterion (maximizing Information Gain), a maximum depth capped at 15, a ‘min\_samples\_split’ set to 10, and a ‘min\_samples\_leaf’ set to 4. This highly structured architecture achieved a cross-validation score of 54.72% during the training phase.

Hyperparameter optimization was performed on the SVM algorithm using 5-fold cross-validation and the Grid Search method. As a result of the search, the most successful model parameters are determined to be  $C = 100$ ,  $\gamma = 0.01$ , and the ‘Radial Basis Function’ (RBF) as the kernel function. The best cross-validation accuracy on the training data is measured at 83.11%.

Hyperparameter tuning for the XGBoost classifier was performed utilizing an exhaustive grid search strategy.

Evaluated via 5-fold cross-validation, the model achieved its highest performance metric (89.60%) using a learning rate of 0.2, a maximum tree depth of 7, and 200 estimators.

4.2 Deep learning methods

ResNet-152+CBAM, YOLO11 and DINOv2, which are frequently used DL architectures, form the basis for classification.

As neural networks become deeper, they often suffer from the 'vanishing gradient' problem, where gradients become infinitesimally small during backpropagation, halting the learning process. ResNet-152 addresses this by utilizing 'skip connections' (or residual blocks) that allow the gradient to bypass layers and flow directly to earlier parts of the network. This enables the training of significantly deeper networks without performance degradation, allowing the model to capture highly complex, hierarchical features of plant diseases.

While YOLO architectures are traditionally recognized for object detection through bounding box regression, this study leverages the YOLO11 architecture specifically in its classification mode (via the Ultralytics framework). In this configuration, the bounding box regression head and associated detection metrics (such as mean average precision (mAP) or intersection over union (IoU)) are omitted. Instead, the YOLO11 backbone acts as a highly efficient feature extractor, passing the learned representations to a global average pooling layer and a fully connected classification head. Designed for direct image-level classification, these models eliminate complex post-processing steps. This structural simplicity drastically reduces computational latency, enabling real-time inference on resource-constrained agricultural embedded systems.

Unlike standard supervised learning that relies heavily on labeled data, DINOv2 utilizes a self-supervised approach with a student-teacher network architecture. It employs a discriminative self-supervised method where the model learns to match the output of a 'student' network to a 'teacher' network. This allows the model to learn robust features (like leaf texture and shape) without explicit labels, making it highly effective for complex agricultural datasets where labeling is expensive.

All DL models were retrained across three independent runs using the fixed random seeds 42, 123 and 1024 to ensure

reproducibility and the reported metrics correspond to the mean  $\pm$  standard deviation across these runs. The models are trained using Cross Entropy loss. MixUp technique and exponential learning rate decay were utilized during training time. To improve the model's generalization ability and prevent overfitting, a series of data augmentation and preprocessing steps were applied during the training phase. The input images were first subjected to random horizontal flipping and then to random rotations sampled at equal intervals within the range  $[-10^\circ, 10^\circ]$ . Following the geometric transformations, the images were converted into tensors and resized to the predefined network input dimensions using bicubic interpolation. Finally, the image tensors were normalized using the dataset's standard mean and standard deviation values. The Python programming language and the PyTorch framework were used as the software development environment.

The DINOv2 model is used in conjunction with the pre-trained dinov2\_vits14 model, and the features extracted from the model were fed into a two-layer feed-forward neural network. From the Ultralytics library, the pretrained YOLO11x-cls model is used for the YOLO11 training. ResNet-152 with Convolutional Block Attention Module (CBAM) model doesn't use pre-trained weights. The model's parameter count, computation, and model size metrics are given in Table 2. All reported model sizes correspond to the on-disk size of the FP32 model weights (parameters only), excluding optimizer states, gradients, and auxiliary checkpoint metadata, thereby ensuring a consistent and architecture-independent basis for comparison.

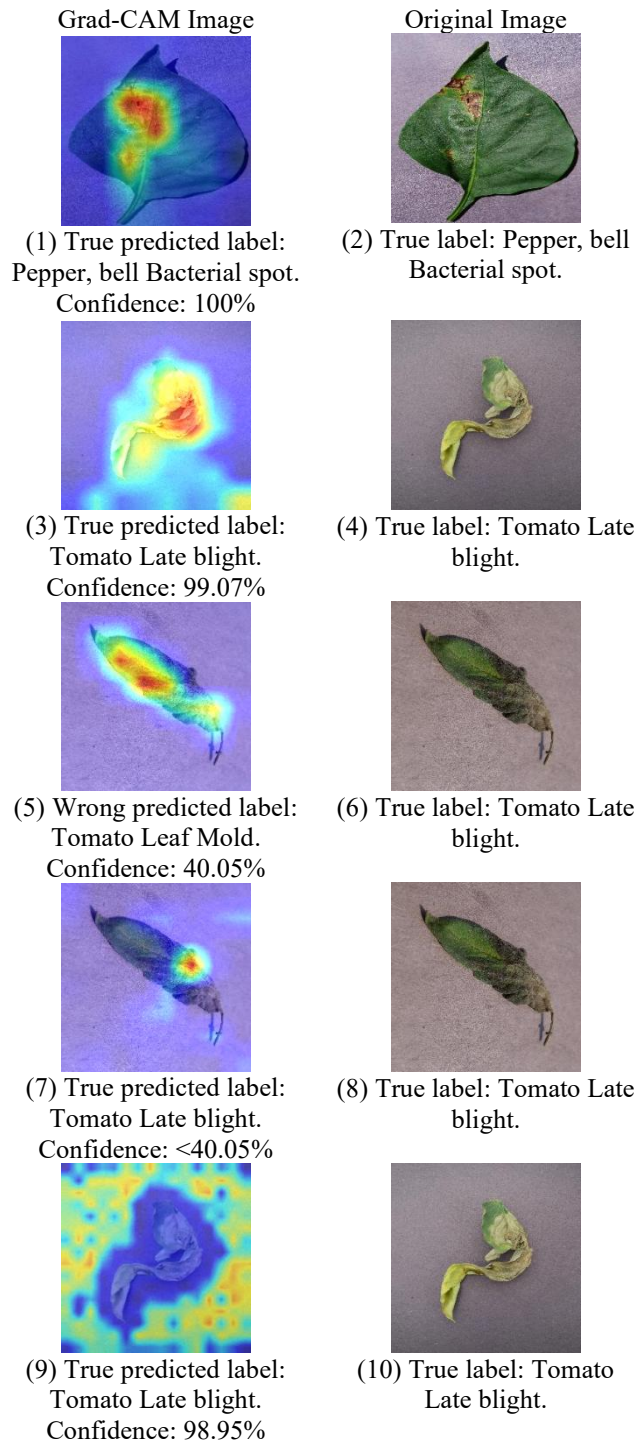
The model training parameters for the ResNet-152+CBAM, YOLO11, and DINOv2 DL models are summarized in Table 3. To ensure the interpretability of the DL predictions and address the "black-box" nature of these architectures, Gradient-weighted Class Activation Mapping (Grad-CAM) is applied. Unlike isolated heatmaps, Figure 3 presents a comparative composite visualization encompassing the original input image, the ground truth label, the predicted class, and the superimposed Grad-CAM heatmap. This structured visualization rigorously demonstrates whether the models' high confidence scores are correctly localizing the actual pathological lesions or spuriously focusing on background noise, offering a transparent view into both correct predictions and edge-case failures.

Table 2. Deep learning (DL) model's parameter counts, sizes

| Model           | Total Params (M) | Computation (GFLOPs) | Model Size (MB) |
|-----------------|------------------|----------------------|-----------------|
| ResNet-152+CBAM | 64.80            | 240.83               | 259.22          |
| YOLO11          | 28.40            | 111.0                | 108.36          |
| DINOv2          | 22.16            | 2.54                 | 86.55           |

Table 3. Model training parameters table for ResNet-152 + CBAM, YOLO11, and DINOv2 deep learning (DL) models

| Models/PARAMETERS     | ResNet-152 + CBAM | YOLO11                                     | DINOv2        |
|-----------------------|-------------------|--------------------------------------------|---------------|
| Pretrained Model      | none              | yolo11x-cls                                | dinov2_vits14 |
| Batch Size            | 64                | 128                                        | 64            |
| Initial Learning Rate | 1e-2              | 1e-2                                       | 1e-6          |
| MixUp                 | Used              | Used                                       | Not Used      |
| Optimizer             | AdamW             | SGD (momentum:0.937, weight decay: 0.0005) | Adam          |
| Dropout               | 0.0               | 0.1                                        | 0.0           |
| Scheduler (Gamma)     | 0.99              | linear (lrf = 0.01)                        | 0.98          |
| Epoch                 | 100               | 100                                        | 100           |
| Input Size            | 256 × 256         | 224 × 224                                  | 224 × 224     |



**Figure 3.** Grad-CAM Images: (1) Prediction with YOLO11. (3) Prediction with ResNet-152+CBAM. (5) Wrong prediction with ResNet-152+CBAM. (7) True prediction with ResNet-152+CBAM. (9) Prediction with DINOv2

## 5. RESULTS AND DISCUSSION

To provide a comprehensive evaluation across the 38-class problem and to strictly account for potential class imbalances, performance is reported utilizing macro-averaged metrics alongside weighted averages. Furthermore, to evaluate algorithmic stability, DL model training is repeated over 3 times, yielding mean  $\pm$  standard deviation scores across key metrics. A detailed per-class precision and recall analysis, supported by the Confusion Matrix (see Figure 4), reveals that

while overall accuracy is exceptionally high, the rare misclassifications primarily occur between visually homologous disease stages, such as the early necrotic lesions of 'Tomato Late Blight' and 'Tomato Leaf Mold'; for instance, only 1 of the 253 'Tomato Late Blight' samples was misclassified as 'Tomato Leaf Mold'.

While the XGBoost classifier achieved a respectable 91.59% accuracy using only structural and textural features, the integration of deep semantic representations (ResNet-50) paired with an SVM significantly elevated the diagnostic accuracy to 96.16%, highlighting the necessity of deep feature fusion for resolving visually ambiguous disease classes like Tomato Septoria Leaf Spot.

A detailed analysis conducted on a class-by-class basis reveals the limitations of a single-tree architecture in a high-dimensional feature space. Although the model demonstrated exceptional performance in visually distinct categories such as Corn Common Rust (F1 score: 94.30%) and Early Blight of Potato (F1 score: 87.61%), its predictive ability significantly decreased in classes with subtle and overlapping symptoms. In particular, the F1 scores for tomato Septoria leaf spot, pepper healthy, and apple scab dropped to 41.79%, 43.84%, and 45.07%, respectively. These key findings demonstrate that while a Decision Tree alone can capture the basic data distribution, it lacks the nonlinear classification capacity required for robust and large-scale plant pathology diagnosis. The optimized decision tree model is evaluated on an unseen test set (8,791 examples). As a weak learner on its own, the classifier achieved a moderate overall accuracy of 68.34%, while the macro-average and weighted-average F1 scores were 68.39% and 68.26%, respectively.

When evaluated on an unseen test set (8,791 examples), the optimized SVM model achieved an overall accuracy of 88.68%. The macro-average and weighted-average F1 scores were consistent at 88.62% and 88.63%, respectively. This represents a significant improvement over the baseline independent Decision Tree (68.34%) and demonstrates that morphological and textural features are better distinguished by hyperplanes that maximize the margin.

The Model, Accuracy, Weighted Average Precision, Weighted Average Recall, Weighted Average F1 Score, Macro Average Precision, Macro Average Recall and Macro Average F1 score values of ML and DL models in the classification of the New Plant Disease image dataset are provided in Table 4. Also, the best model's confusion matrix is given in Figure 4.

While state-of-the-art DL architectures such as YOLO11 and ResNet-152+CBAM achieve substantially higher accuracy than the proposed hybrid model, this study does not claim equivalent performance. On the 8,791-image test set, the deep learning models reached 99.89% (YOLO11), 99.68% (ResNet-152+CBAM), and 99.13% (Dinov2), exceeding the hybrid model (96.16%) by approximately 3.0 to 4.7 percentage points, and they therefore provide clearly superior classification accuracy. However, these models operate with tens of millions of parameters (e.g., YOLO11x with over 28 million parameters) and require substantial GPU memory and computational resources. By contrast, the proposed hybrid PCA-SVM architecture requires only 8,072,928 trainable parameters and occupies a mere 68.96 MB of physical storage. The advantage of the proposed framework therefore lies not in matching the accuracy of these deep networks, but in its compact model design, its feature-fusion strategy and its markedly lower resource requirements, which indicate



Although manual feature extraction initially entailed a high computational cost (approximately 1.8 hours), reducing the dimensionality to 256 using PCA made the model extremely fast during the testing/inference phase and suitable for real-time applications.

While this study establishes a robust comparative baseline regarding feature quality and theoretical computational cost, future research is fundamentally required to validate these findings on physical agricultural embedded systems. Subsequent studies will focus on deploying these optimized models—particularly the YOLO11 architecture—on hardware accelerators such as NVIDIA Jetson edge nodes to rigorously benchmark real-time inference latency Frame Per Second (FPS), thermal behavior, and power consumption in dynamic field environments. External field-condition validation is acknowledged as an important next step and is planned for future work, as no suitable in-domain field dataset was available within the current revision scope.

## 6. CONCLUSIONS

This study presented a comparative evaluation of traditional ML and modern DL architectures for plant leaf disease classification on the 38-class New Plant Disease dataset, with an emphasis on the trade-off between feature quality, classification accuracy, and model storage cost. To ensure a fair comparison, a hybrid pipeline was constructed in which deep semantic features extracted from a pre-trained ResNet-50 backbone were fused with handcrafted textural and color descriptors and reduced to 256 dimensions using PCA. The proposed hybrid model reached an accuracy of 96.16%, substantially improving upon the handcrafted-only baseline (91.59% with XGBoost) and approaching the performance of the DL models — YOLO11 (99.89%), ResNet-152+CBAM (99.68%), and DINOv2 (99.13%), averaged over three independent runs — while requiring only about 8.07 million parameters and 68.96 MB of storage.

These results indicate that fusing deep and handcrafted representations offers a competitive yet computationally efficient alternative to large end-to-end deep networks, which is particularly relevant for resource-constrained agricultural deployment. The GMM analysis, guided by the AIC and BIC criteria, further provided a probabilistic characterization of the dataset's structural distribution rather than a direct measure of class separability.

The present findings are, however, limited to a single, controlled, augmentation-based dataset and to the reported training configuration; the reported model sizes reflect FP32 parameter storage rather than on-device runtime cost. Accordingly, future work will focus on validating the models on external, field-acquired datasets, on benchmarking real-time inference latency, memory, and power consumption on embedded hardware such as NVIDIA Jetson edge nodes, and on extending the complexity analysis to the deep feature space.

## ACKNOWLEDGMENT

The authors thank the editor and two of the anonymous referees for carefully reading the manuscript and making some valuable comments which had greatly improved the earlier draft of the manuscript. This work did not receive any specific grant from funding agencies in the public, commercial or not-

for-profit sectors. During the preparation of this manuscript, the authors used a generative artificial intelligence (AI)-based language tool solely for language editing and readability improvement; the authors reviewed and edited the content and take full responsibility for it.

## CONFLICT OF INTEREST

The authors declare that there are no known conflicts of interest, financial or otherwise, regarding the publication of this article.

## REFERENCES

- [1] Martinelli, F., Scalenghe, R., Davino, S., et al. (2015). Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*, 35: 1-25. <https://doi.org/10.1007/s13593-014-0246-1>
- [2] Khirade, S.D., Patil, A.B. (2015). Plant disease detection using image processing. In 2015 International Conference on Computing Communication Control and Automation, Pune, India, pp. 768-771. <https://doi.org/10.1109/ICCUBEA.2015.153>
- [3] Beevi, L.S., Prathap, P.M.J., Reddy, K.V.V., Balaji, K., Nishith, M., Dani, W.V. (2024). Revolutionizing agriculture: Detection of plant disease using machine learning and deep learning algorithms. In 2024 4th International Conference on Sustainable Expert Systems (ICES), Kaski, Nepal, pp. 1446-1450. <https://doi.org/10.1109/ICES63445.2024.10763373>
- [4] Gupta, A., Chug, A., Singh, A.P. (2024). Leveraging feature extraction and machine learning for effective plant disease classification. In 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), New Delhi, India, pp. 1-5. <https://doi.org/10.1109/DELCON64804.2024.10866877>
- [5] Yang, C.Z. (2021). Plant leaf recognition by integrating shape and texture features. *Pattern Recognition*, 112: 107809. <https://doi.org/10.1016/j.patcog.2020.107809>
- [6] Gil, E.S., Dias, L.M.A., Borges, P.V.K., Drews, P.L.J., Brito, A.S., Oliveira, F.G. (2025). Vision-based plant disease identification for autonomous crop management systems. In 2025 Brazilian Conference on Robotics (CROS), Belo Horizonte, Brazil, pp. 1-6. <https://doi.org/10.1109/CROS66186.2025.11066155>
- [7] Galabuzi, C., Abdullah, H., Ahmad, N., Kaidi, H.M. (2024). EfficientNet-based deep learning neural network for accurate plant disease detection. In 2024 5th International Conference on Smart Sensors and Application (ICSSA), Penang, Malaysia, pp. 1-6. <https://doi.org/10.1109/ICSSA62312.2024.10788558>
- [8] Kotwal, J.G., Kashyap, R., Shafi, P.M. (2024). Artificial driving based EfficientNet for automatic plant leaf disease classification. *Multimedia Tools and Applications*, 83: 38209-38240. <https://doi.org/10.1007/s11042-023-16882-w>
- [9] Sun, X.W., Li, G.H., Qu, P.X., Xie, X.W., Pan, X.P., Zhang, W.D. (2022). Research on plant disease identification based on CNN. *Cognitive Robotics*, 2: 155-163. <https://doi.org/10.1016/j.cogr.2022.07.001>
- [10] Apleni, T., Isinkaye, F.O., Olusanya, M.O. (2025). Ensemble-based feature fusion for accurate plant disease

- classification using pre-trained models. *Scientific Reports*, 15: 41925. <https://doi.org/10.1038/s41598-025-25927-z>
- [11] Bunyang, S., Thedwichienchai, N., Pintong, K., et al. (2023). Self-supervised learning advanced plant disease image classification with SimCLR. *Advances in Computational Intelligence*, 3: 18. <https://doi.org/10.1007/s43674-023-00065-z>
- [12] Yadav, D., Akanksha, Yadav, A.K. (2020). A novel convolutional neural network based model for recognition and classification of apple leaf diseases. *Traitement du Signal*, 37(6): 1093-1101. <https://doi.org/10.18280/ts.370622>
- [13] Rangarajan, A.K., Purushothaman, R., Prabhakar, M., Szczepański, C. (2023). Crop identification and disease classification using traditional machine learning and deep learning approaches. *Journal of Engineering Research*, 11(1): 228-252. <https://doi.org/10.36909/JER.11941>
- [14] McQueen, R.J., Garner, S.R., Nevill-Manning, C.G., Witten, I.H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, 12(4): 275-293. [https://doi.org/10.1016/0168-1699\(95\)98601-9](https://doi.org/10.1016/0168-1699(95)98601-9)
- [15] Biswas, A., Banik, R. (2024). Machine learning integration in agriculture domain. In *Fog Computing for Intelligent Cloud IoT Systems*, pp. 71-97. <https://doi.org/10.1002/9781394175345.ch4>
- [16] Behmann, J., Mahlein, A.K., Rumpf, T., Römer, C., Plümer, L. (2015). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, 16: 239-260. <https://doi.org/10.1007/S11119-014-9372-7>
- [17] Acevedo-Sánchez, G., Alarcón-Paredes, A., Yáñez-Márquez, C. (2025). Effect of agriculture-related dataset complexity on classical machine learning and deep learning classifiers performance. *Computers and Electronics in Agriculture*, 239: 110941. <https://doi.org/10.1016/j.compag.2025.110941>
- [18] Orchi, H., Sadik, M., Khaldoun, M., Sabir, E. (2023). Automation of crop disease detection through conventional machine learning and deep transfer learning approaches. *Agriculture*, 13(2): 352. <https://doi.org/10.3390/agriculture13020352>
- [19] Mohanty, S.P., Hughes, D.P., Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7: 1419. <https://doi.org/10.3389/fpls.2016.01419>
- [20] Barbedo, J.G.A., Koenigkan, L.V., Halfeld-Vieira, B.A., Costa, R.V., Nechet, K.L., Godoy, C.V. (2018). Annotated plant pathology databases for image-based detection and recognition of diseases. *IEEE Latin America Transactions*, 16(6): 1749-1757. <https://doi.org/10.1109/TLA.2018.8444395>
- [21] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, Hyderabad, India, pp. 249-253. <https://doi.org/10.1145/3371158.3371196>
- [22] Prashant, Garg, T., Theodore, S.K.A., Theodore, M., S. (2024). Krishi-Konekt: Enhancing agriculture with crop care solutions. In *2024 2nd International Conference on Networking and Communications (ICNWC)*, Chennai, India, pp. 1-6. <https://doi.org/10.1109/icnwc60771.2024.10537505>
- [23] Nagar, A., Kumar, D., Kumar, B. (2025). Multiple leave disease detection using CNN in plants. In *Recent Trends in Intelligent Computing and Communication*, CRC Press, pp. 106-110. <https://doi.org/10.1201/9781003593034-17>
- [24] Hassan, S.M., Maji, A.K. (2022). Plant disease identification using a novel convolutional neural network. *IEEE Access*, 10: 5390-5401. <https://doi.org/10.1109/ACCESS.2022.3141371>
- [25] Geetharamani, G., Pandian, A.J. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering*, 76: 323-338. <https://doi.org/10.1016/j.compeleceng.2019.04.011>
- [26] Lin, F.D., Crawford, S., Guillot, K., et al. (2023). Mmst-vit: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5774-5784.
- [27] Li, H.Y., Bie, Y.H., Wang, Y.J. (2024). AgriST-Trans: A self-supervised Transformer pre-trained model for crop classification based on time-series remote sensing. In *2024 12th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Novi Sad, Serbia, pp. 1-6. <https://doi.org/10.1109/agro-geoinformatics262780.2024.10660832>
- [28] Hao, P.F., An, J.P., Cai, Q., et al. (2025). Data-efficient and accurate rapeseed leaf area estimation by self-supervised vision transformer for germplasms early evaluation. *Research Square*. <https://doi.org/10.1016/j.jia.2025.12.041>
- [29] Zhong, M., Wei, L., Mo, H. (2025). Research on lightweight detection of cotton leaf diseases based on self-supervised contrastive representation learning. *Journal of Integrative Agriculture*. <https://doi.org/10.1016/j.jia.2025.12.041>
- [30] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716-723. <https://doi.org/10.1109/TAC.1974.1100705>
- [31] Neath, A.A., Cavanaugh, J.E. (2011). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, 4(2): 199-203. <https://doi.org/10.1002/wics.199>
- [32] Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5): 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- [33] Quinlan, J.R. (1996). Learning decision tree classifiers. *ACM Computing Surveys*, 28(1): 71-72. <https://doi.org/10.1145/234313.234346>
- [34] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20: 273-297. <https://doi.org/10.1007/BF00994018>
- [35] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [36] Jocher, G., Qiu, J. (2024). Ultralytics YOLO11 (Version 11.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics>.
- [37] Oquab, M., Darcet, T., Moutakanni, T., et al. (2024). DINOv2: Learning robust visual features without

- supervision. arXiv preprint arXiv:2304.07193. <https://doi.org/10.48550/arXiv.2304.07193>
- [38] Haralick, R.M., Shanmugam, K., Dinstein, I.H. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6): 610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [39] Pietikäinen, M. (2010). Local binary patterns. Scholarpedia, 5(3): 9775. <http://doi.org/10.4249/scholarpedia.9775>