Applied Fruit Science (2026) 68:133
https://doi.org/10.1007/s10341-026-01830-3

RESEARCH

Enhanced Lemon Leaf Disease Diagnosis via Gray Wolf Optimization and Machine Learning Techniques

Yavuz Unal[1] · **Yonis Gulzar**[2]

Received: 25 January 2026 / Accepted: 18 March 2026 / Published online: 20 April 2026

Abstract

This study proposes a practical and efficient framework for lemon leaf disease classification by combining deep feature extraction with optimization-based feature selection and machine learning classifiers. Instead of relying solely on end-to-end deep learning, the work focuses on improving feature quality and reducing redundancy. DenseNet121 is used to extract 1,024 deep features from a nine-class lemon leaf disease dataset, capturing relevant visual patterns from the images. These features are then refined using Gray Wolf Optimization (GWO), a wrapper-based feature selection method that reduces the feature space to 405 features, achieving a dimensionality reduction of 60.45%. To evaluate the effectiveness of the selected features, both Random Forest and Support Vector Machine (SVM) classifiers are applied to the original and optimized feature sets. The results show that Random Forest achieves 96.97% accuracy with original features and improves to 97.64% after feature selection. Similarly, SVM achieves the highest accuracy of 98.30% with original features and maintains competitive performance after optimization. The findings demonstrate that GWO effectively removes redundant information while preserving discriminative features, leading to improved efficiency without compromising performance. Compared to baseline approaches on the same dataset, the proposed hybrid framework achieves superior classification accuracy and reduced computational cost. This makes the method suitable for real-world agricultural applications where both accuracy and efficiency are important.

Keywords Precision agriculture · Plant disease detection · Deep learning in agriculture · Feature selection optimization · DenseNet121

Introduction

One of the most produced and traded citrus fruits in the world is the lemon (Niu et al. 2021). It is widely cultivated across temperate regions and holds significant economic value in food, cosmetics, and pharmaceutical industries (Iqbal et al. 2018). Citrus fruits are susceptible to many diseases caused by pathogens and fungi (Siam 2024) that spread rapidly between trees (Zhu et al. 2024), causing significant yield losses (J. Kotwal et al. 2023) and negatively affecting plant development and health (Iqbal et al. 2018). Leaf symptoms such as spots, yellowing, deformation, and curling indicate disease presence and reduce photosynthetic capacity, affecting plant growth (Shoaib et al. 2023). Early detection is essential to prevent significant agricultural losses (J. G. Kotwal et al. 2024; Gulzar 2025c). Traditional diagnostic methods are time-consuming, error-prone, and expensive for large areas, requiring expertise (Siam 2024). Machine learning and deep learning have been used successfully in plant disease detection (Kotwal et al. 2023), representing a paradigm shift for plant health inspection and early diagnosis (Yasin et al. 2025) (Gulzar 2025a). Deep learning models designed for image processing can accurately identify disease symptoms such as spots, color changes, or deformities, enabling farmers to reduce crop losses (Barman et al. 2020; Gulzar 2025b). Despite promising results, existing studies predominantly rely on end-to-end convolutional neural network (CNN) or transfer learning pipelines that feed all extracted features directly into

✉ Yavuz Unal
yunal@sinop.edu.tr

✉ Yonis Gulzar
ygulzar@kfu.edu.sa

[1] Faculty of Engineering and Architecture, Sinop University, Sinop, Turkey

[2] Department of Management Information Systems, College of Business Administration, King Faisal University, 31982 Al-Ahsa, Saudi Arabia

classifiers, overlooking feature redundancy and increasing computational cost. Furthermore, wrapper-based feature selection methods such as Gray Wolf Optimization (GWO) remain largely unexplored in lemon leaf disease classification, representing a clear research gap. This study aims to classify lemon leaf diseases using machine learning and image processing techniques. The study performs feature extraction using the DenseNet121 pre-trained network on lemon leaf images consisting of nine classes, followed by feature selection using the GWO algorithm. Random Forest (RF) and Support Vector Machine (SVM) algorithms are applied in the classification phase.

The main contributions of this study are as follows:

- DenseNet121 is employed as a deep feature extractor on a nine-class lemon leaf disease dataset.
- GWO is applied as a wrapper-based feature selector, reducing dimensionality by 60.45% (from 1024 to 405 features).
- SVM and RF classifiers are evaluated on both original and GWO-selected features.
- The proposed hybrid framework achieves up to 98.30% classification accuracy, outperforming the baseline study on the same dataset.

The rest of the study is organized as follows: the related studies on the classification of lemon and citrus leaf diseases are reviewed in the Literature Review section. The dataset and techniques used are explained in the Materials and Methods section. The findings are presented in the Results section, followed by a discussion of the results and suggestions for future research directions in the Discussion section.

Related Work

In one study, a dataset of lemon leaf images with nine classes was created (1354 original images), achieving 96.19% classification accuracy using pre-trained networks (Siam 2025). Another study classified lemon leaves in three classes using DenseNet-201, ResNet-50, ResNet-152V2, and Xception, with Xception achieving the highest accuracy of 94.43% (Pramanik et al. 2021). A citrus leaf dataset including *Citrus tristeza* virus (CTV)- and Huanglongbing (HLB)-infected leaves was developed, comparing MobileNet and SSCNN; MobileNet achieved 92% test accuracy while SSCNN reached 99% (Barman et al. 2020). Dang-Ngoc et al. (2021) preprocessed citrus leaf images from four classes and achieved 92.5% accuracy using an SVM-based classifier.

Using VGG16 and InceptionV3 on a four-class citrus dataset, Islam et al. (2024) achieved 98% and 99% accuracy, respectively. Luaibi et al. (2021) used a four-class dataset with data augmentation, achieving 95.83% and 97.92% accuracy with ResNet and AlexNet. A five-class lemon leaf dataset achieved 98.63% accuracy using MobileNetV2 (Reddy and Thirunavukkarasu, 2024).

Saygili (2023) investigated transfer learning and data augmentation effects using VGG-16, ResNet50, and DenseNet201, achieving the highest accuracy (98.29%) with DenseNet201. Sharif et al. (2018) performed feature selection using principal component analysis (PCA), skewness, and entropy on a seven-class citrus dataset with multi-class SVMs, reporting 97%, 89%, and 90.4% accuracy on different datasets.

Çetiner (2022) developed a CNN-based architecture for citrus disease detection, including blackspot, canker, and HLB, achieving an average accuracy of 96%. Duth and Bhat (2022) classified a four-class citrus dataset using ResNet50, ResNet101, and VGG-19, with VGG-19 achieving the highest success (97.95%). Suryavanshi et al. (2023) combined CNN and RF to detect Alternaria Brown Spot disease, achieving 97% accuracy. Yuan (2023) classified healthy and aphid-damaged lemon leaves using DenseNet and MobileNet, achieving 80.47% and 79.81% accuracy. Zhu et al. (2024) classified a six-class citrus dataset using a Multi-Models Fusion Network (MMFN) after data augmentation, achieving 98.68% accuracy. Butt et al. (2025) achieved 99.6% accuracy using DenseNet-201 and AlexNet with transfer learning on citrus leaves.

Beyond citrus, several studies applied deep learning to plant disease classification (Gulzar and Ünal 2025). Mia et al. (2021) proposed a You Only Look Once (YOLO)-based model for real-time recognition of medicinal plant leaves, achieving approximately 95% accuracy. Mahmud et al. (2022) employed CNN-based transfer learning for early detection of lychee leaf diseases, with InceptionV3 achieving 92.67% accuracy. Al Fahim et al. (2023) reviewed automated seed classification approaches using deep neural networks. Bitto et al. (2023) compared CNN architectures for potato leaf disease recognition, with ResNet-50 achieving the highest accuracy, followed by Inception-v3 (94.25%). Islam et al. (2025) proposed a model for dragon fruit leaf disease classification, achieving 92.83% accuracy with CNN + transfer learning. Sarker et al. (2025) created a flower dataset for species recognition, achieving 99.81% accuracy with MobileNetV2. Akash et al. (2024) developed a decision support system for crop recommendation, achieving 98.11% accuracy with RF.

Most existing studies rely on end-to-end CNN or transfer learning pipelines, where all extracted features are fed directly into classifiers without selection. This approach ignores feature redundancy and increases computational cost. Moreover, few studies have examined wrapper-based optimization methods for feature selection in citrus disease classification, leaving a gap that this study addresses. This study

Table 1 Performance comparison of existing studies on lemon and citrus leaf disease classification using deep learning approaches

Reference	Methods	Fruit/crop name	Class	Classification accuracy (%)
Siam et al. (2024)	DenseNet-121	Lemon	9	96.19
Pramanik et al. (2021)	Xception, DenseNet-201, ResNet-50, ResNet-152V2	Lemon	3	94.43, 92.45, 88.68, 79.25
Barman et al. (2020)	MobileNet, SSCNN	Citrus	3	92, 99
Çetiner (2022)	MobileNet, ResNet50, Custom CNN	Citrus	4	96
Dang-Ngoc (2021)	SVM	Citrus	4	92.5
Islam et al. (2024)	InceptionV3, VGG-16	Citrus	4	99, 98
Luaibi et al. (2021)	ResNet, AlexNet	Citrus	4	95.83, 97.92
Reddy et al. (2024)	VGG-16, MobileNetV2	Lemon	5	93.84, 98.63
Saygili (2023)	VGG-16, ResNet-50, DenseNet201	Lemon	2	81.70, 96.59, 98.29
Sharif et al. (2018)	M-SVM, W-KNN, Ensemble Boosted Tree, Decision Tree, Linear Discriminant Analysis	Citrus	7	97, 89, 90.4
Duth and Bhat (2022)	ResNet50, ResNet101, VGG19	Citrus	4	94.61, 95.21, 97.95
Suryavanshi et al. (2023)	CNN + Random Forest	Lemon	4	97
Yuan (2023)	DenseNet, MobileNet	Lemon	2	80.47, 79.81
Zhu et al. (2024)	Multi-Models Fusion Network (MMFN)	Citrus	6	98.68
Butt et al. (2025)	DenseNet-201, AlexNet + Hybrid Meta-Heuristic	Citrus	4	99.6

adopts a hybrid approach: DenseNet121 extracts discriminative deep representations, producing a 1024-dimensional feature vector refined using GWO. The selected features are evaluated with SVM and RF classifiers. Unlike existing studies that use DenseNet solely as an end-to-end classifier, this study decouples feature extraction from classification and introduces GWO as a wrapper-based selector to identify the most discriminative subset, resulting in a more efficient and interpretable pipeline.

It is worth noting that wrapper-based feature selection methods, despite their proven effectiveness in other domains, remain rarely applied in citrus and lemon leaf disease classification, making this study one of the few to systematically evaluate GWO as a wrapper-based selector in this context.

Additional related studies in plant disease classification are summarized in Table 1.

Materials and Methods

This section details the dataset, proposed model, technical methodologies, evaluation metrics, and algorithms used in the study.

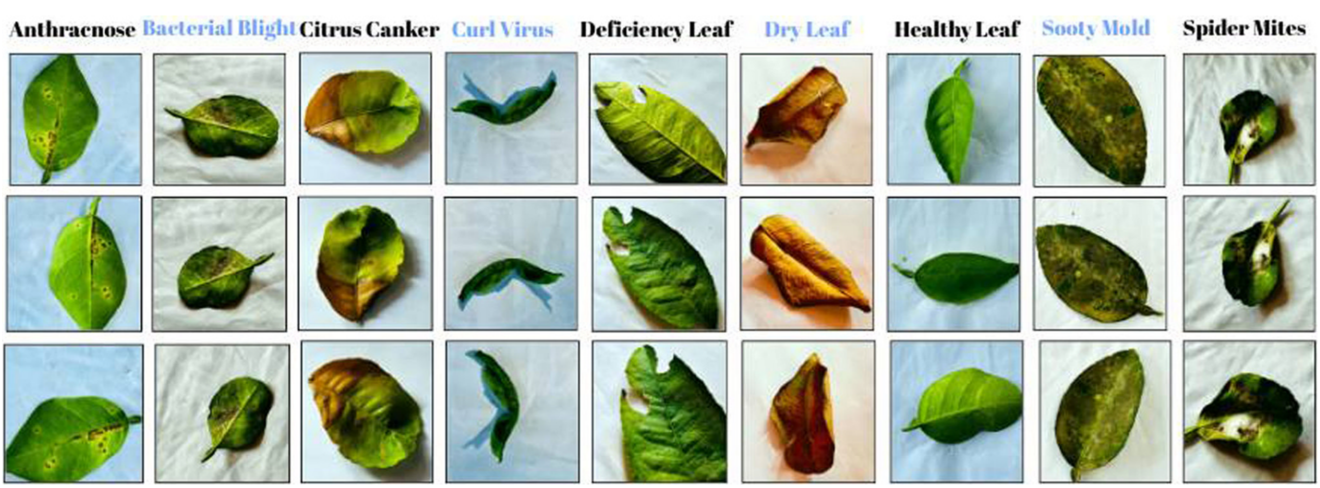


Fig. 1 Sample images of each class in the Lemon Leaf Disease dataset

Table 2 Class distribution of the original Lemon Leaf dataset

Class	Original count
Anthracnose	100
Bacterial Blight	105
Citrus Canker	178
Curl Virus	115
Deficiency Leaf	193
Dry Leaf	186
Healthy Leaf	210
Sooty Mold	153
Spider Mites	114
Total	1354

Dataset

In this study, the Lemon Leaf Diseases dataset (Siam 2024) was used. The dataset is publicly available in two versions: the original dataset comprising 1354 raw images and an augmented version comprising 9000 images generated via data augmentation (translation, rotation, zoom, cropping, and brightness adjustment). For this study, we used only the original dataset (1354 images) to ensure fair comparison with the baseline study (Siam 2025) and to evaluate the effectiveness of our feature selection approach without relying on artificially augmented samples. The images are in red–green–blue (RGB) format, with a resolution of 800×800 pixels, and were collected in 2024 from Jamalpur District, Bangladesh. The dataset contains nine classes of lemon leaf conditions.

Some of the sample citrus leaf diseased images used in this study are presented in Fig. 1.

The classes of lemon leaf images in the original dataset, along with their numerical labels, are presented in Table 2. There are nine classes in the database.

The classes defined in the dataset include Anthracnose, Bacterial Blight, Citrus Canker, Curl Virus, Deficiency Leaf, Dry Leaf, Healthy Leaf, Sooty Mold, and Spider Mites, respectively. The number of images belonging to each class is as follows: Anthracnose (100), Bacterial Blight (105), Citrus Canker (178), Curl Virus (115), Deficiency Leaf (193), Dry Leaf (186), Healthy Leaf (210), Sooty Mold (153), and Spider Mites (114). The original dataset comprises 1354 images of lemon leaves. The augmented version of this dataset includes 9000 images. The dataset comprising 1354 images was used in the study. Figure 2 presents a histogram of the number of samples per class in the dataset.

Proposed Method

In this study, a model integrating image processing, deep learning, and machine learning techniques is proposed for the classification of diseases observed in lemon leaves of the citrus family. The proposed model comprises feature extraction from the pre-trained DenseNet121 network, fea-

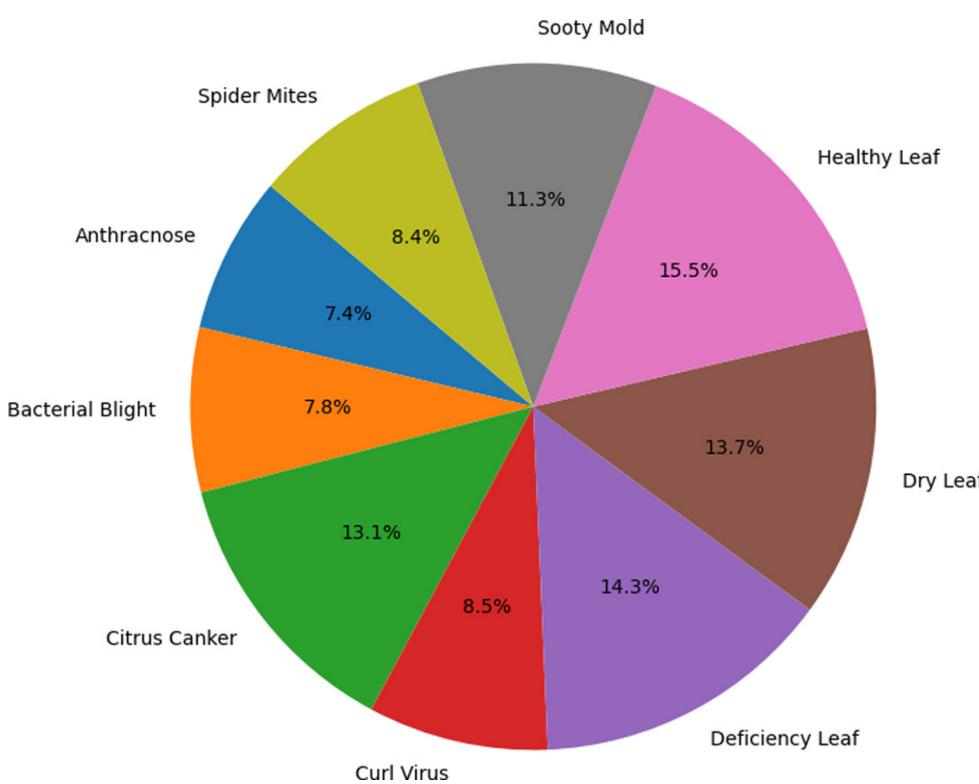


Fig. 2 A pie chart illustrates the distribution of leaf conditions and diseases in the dataset

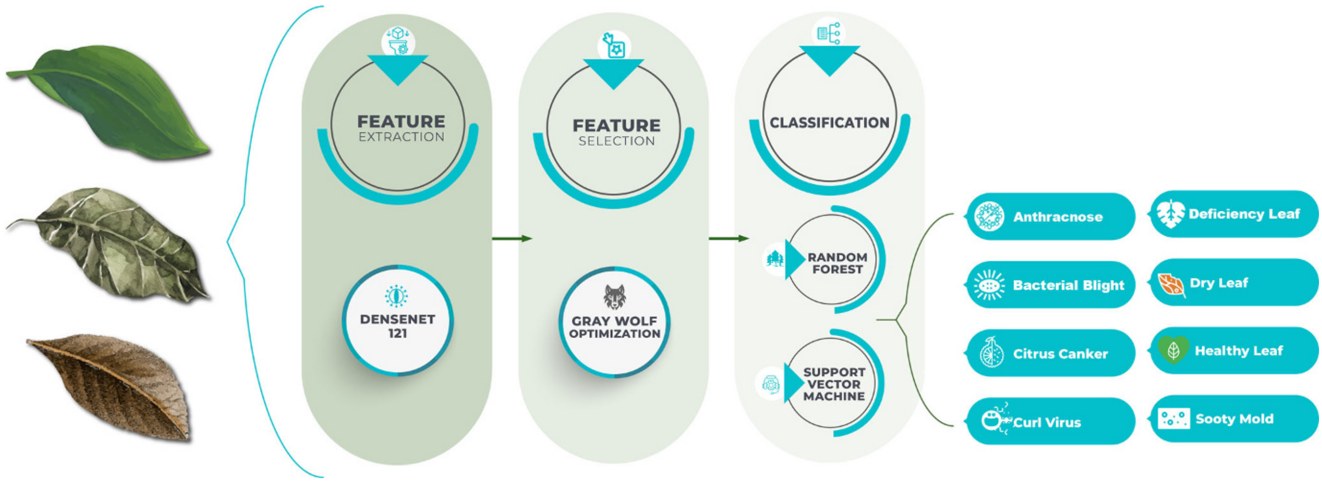


Fig. 3 Flowchart of the proposed methodology

ture selection using the GWO algorithm, classification with SVM and RF classifiers, model training, and performance evaluation. The model's general architecture is presented in Fig. 3.

In the feature extraction step, DenseNet121, a pre-trained network, was used as a feature extractor to capture relevant image features while reducing data size and increasing model efficiency. The model's final classification layer (the 1000-unit softmax layer) was removed, and features were extracted from the last global average pooling layer (pool5), which produces a 1024-dimensional feature vector. This layer captures high-level semantic representations while retaining spatial context from the preceding dense block. Global average pooling converted variable-sized feature maps into fixed-length vectors. Images were resized to 224×224 pixels and normalized according to the model's preprocessing standards, resulting in 1024-dimensional feature vectors.

For feature selection, GWO, a meta-heuristic algorithm inspired by wolf hunting behavior, was used to identify informative features and reduce dimensionality. GWO simulates wolf social hierarchy and prey searches, where wolf locations represent feature subsets.

Each location is evaluated using a fitness function that balances classification accuracy and the number of selected features. The fitness function is formally defined as:

$$F = \alpha \cdot E(K) + \beta \cdot (|S|/|F|)$$

where E(K) is the classification error rate of the selected feature subset K, |S| is the number of selected features, |F| is the total number of features, and α and β are weighting parameters controlling the trade-off between accuracy and dimensionality reduction ($\alpha = 0.99$, $\beta = 0.01$). Feature subsets are evaluated by training an RF classifier on an internal validation subset, defined as 20% of the training data held out via stratified sampling. This validation subset was used solely during GWO optimization and was not part of the final test set.

The algorithm iteratively improves fitness scores by adjusting feature subsets. GWO was initialized with 10 wolves and run for 50 iterations using a binary search space (1 = present, 0 = absent). These parameters were selected based on preliminary experiments: population sizes below 10 showed premature convergence, while sizes above 10 provided no significant improvement. Similarly, the fitness value stabilized before the 50th iteration in all trials, as confirmed by the convergence curve (Fig. 4), indicating that 50 iterations are sufficient for this problem size. It reduced the feature set from 1024 to 405 features (60.45% reduction) with a best fitness value of 0.0177.

In the classification phase, SVM with Radial Basis Function (RBF) kernel and RF with 100 trees were used. The RBF kernel was selected for SVM because it effectively handles non-linear decision boundaries in high-dimensional

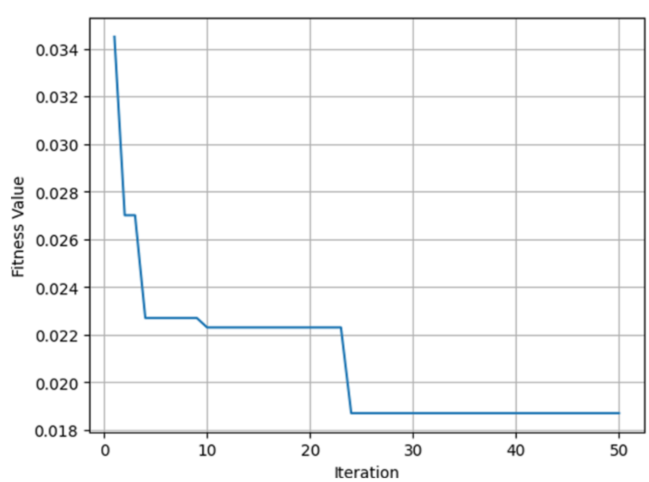


Fig. 4 Convergence curve of the Gray Wolf Optimization algorithm

feature spaces. The regularization parameter C = 1.0 and kernel coefficient gamma = 'scale' were set based on cross-validation. For RF, 100 trees were chosen as a balance between computational efficiency and ensemble stability; preliminary experiments showed no significant accuracy gain beyond this number. The maximum depth was left unrestricted to allow full tree growth, with bootstrap sampling enabled. The dataset was split 70% training and 30% test sets with stratified sampling to balance class distribution. This ratio was selected to provide sufficient training samples for GWO-based feature selection while retaining a representative test set for evaluation. Given the relatively small dataset size (1354 images), an 80/20 split would reduce the test set to approximately 270 images, potentially increasing evaluation variance, while a 60/40 split would limit training data for reliable feature selection. Additionally, five-fold stratified cross-validation was applied to further assess model stability. All preprocessing steps including resizing to 224 × 224 pixels and normalization using ImageNet statistics were applied after data splitting, independently on the training and test sets, to prevent data leakage. Similarly, GWO-based feature selection was performed exclusively on the training set. The test set was held out entirely during feature selection and model training, ensuring unbiased evaluation.

Performance was evaluated using five-fold stratified cross-validation. RF with original features achieved 96.97% ± 0.82%, RF with GWO-selected features achieved 97.64% ± 0.72%, SVM with original features achieved 98.30% ± 0.44%, and SVM with GWO-selected features achieved 97.34% ± 0.79%.

The low standard deviation values across all configurations (0.44%–0.82%) indicate stable and consistent model performance, suggesting that the observed accuracy differences are reliable. Formal significance testing was not performed due to the limited number of folds (k = 5); however, the consistency of results across folds supports the validity of the reported comparisons.

DenseNet121 was chosen because DenseNet architectures effectively capture fine-grained color and texture variations on limited agricultural datasets, achieving high accuracy in plant disease recognition (Haque et al. 2025). Its dense connectivity enables feature reuse and improves gradient flow, reducing overfitting when training images are limited. GWO was selected as a wrapper-based method to optimize both dimensionality reduction and classification performance. It balances maximizing accuracy while minimizing selected features, performing well on agricultural classification problems. GWO reduced dimensionality by 60.45% while maintaining or improving RF accuracy, confirming the selected subset preserved the most informative features for lemon leaf disease recognition.

Technical Methodologies

In this section, the theoretical foundations of the machine learning models used for classification and the optimization techniques employed in the study, which together constitute the research's conceptual framework, are discussed.

Support Vector Machine

This is a supervised learning method that performs well on regression and classification tasks. The main purpose of this method is to find the optimal hyperplane that separates the data into distinct classes. Support vectors are the data points that lie closest to the hyperplane's classification boundary. The space between the support vectors and the hyperplane is called the margin. The SVM method maximizes this margin (Yu et al. 2025). If the data can be linearly separated, SVM finds the optimal hyperplane that separates the data. If the data cannot be linearly separated, it is mapped to a higher-dimensional space via kernel functions and then made linearly separable in that space. SVM is effective in high-dimensional spaces. It is resistant to overfitting. It can be used on both linear and nonlinear data. In addition to these advantages, it has disadvantages, such as slow processing of large datasets (Gambella et al. 2021).

Random Forest

This ensemble learning method works by combining several decision trees. In short, it aims to build many decision trees and then integrate their predictions to produce a model that is more accurate and stable. One of the best features of the RF is that it can address both classification and regression problems, which are the building blocks of other machine learning algorithms (Breiman 2001). Each decision tree in the RF is constructed in parallel, and these trees may be either classification or regression trees. The nodes in each decision tree are partitioned using the best features that yield the most appropriate solution among all features (Mohan et al. 2019; Speiser et al. 2019).

Gray Wolf Optimization Algorithm

Metaheuristic optimization techniques have become increasingly popular in recent years (Mirjalili et al. 2014). These techniques are increasingly used not only by computer scientists but also by researchers in many other fields. The GWO algorithm is a metaheuristic optimization algorithm developed by Seyedali Mirjalili and his team in 2014. It is a nature-inspired algorithm that mimics the hunting and social hierarchy behaviors of gray wolves. GWO is a swarm intelligence algorithm for solving global optimization problems. The algorithm is based on the social structures and hunting strategies of gray wolves (Lu et al. 2025). Gray wolf packs exhibit a hierarchical structure. The leader wolf is called the alpha wolf, the assistant wolf is called the beta wolf, and there are also wolves called delta and omega in the pack. Each of these wolves is used to model different solutions and their effects on one another (Mirjalili et al. 2025).

Evaluation Metrics

We employed several different evaluation indicators to assess the effectiveness of our model in this investigation. The main criteria were accuracy, precision, recall, and the F1-score. Accuracy indicates the proportion of samples that were correctly classified, whereas precision and recall assess how well the model identifies positive cases. The F1-score is the harmonic mean of precision and recall. It provides a balanced assessment of the model's performance, which is helpful when the data is not evenly distributed.

Accuracy is a widely used metric for evaluating a model's overall performance in classification problems. This metric is defined as the proportion of correctly classified examples out of the total number of examples. It is represented by the following formula:

$$\text{Accuracy} = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\right) \times 100$$

Another important metric used in classification is precision. Precision is defined as the ratio of true positive examples to the total number of positive examples classified by the model. The formula is given below.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Another widely used performance metric in classification problems is sensitivity (recall). This metric indicates the proportion of truly positive examples correctly classified as positive by the model. In other words, sensitivity measures how well the model identifies examples from the positive class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

The F1-score allows the evaluation of precision and sensitivity together. It is calculated as the harmonic mean of these two values. The F1-score yields more meaningful results, particularly when the data are imbalanced, i.e., when there are very few positive instances (Haridasan et al. 2023).

$$\text{F1} - \text{Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}.$$

Results

The proposed method was implemented in Python 3.10 using TensorFlow 2.12 and Scikit-learn 1.2. Experiments were performed on an NVIDIA Tesla T4 16 GB GPU provided by Google Colab, with 12.7 GB RAM. DenseNet121 weights were loaded from the Keras applications module pre-trained on ImageNet. The random state was fixed at 42 for all stochastic operations to ensure reproducibility.

Model performance was evaluated using two approaches: (1) a single 70/30 stratified train-test split for generating confusion matrices and computing final accuracy metrics, and (2) five-fold stratified cross-validation to assess model stability and generalization capability.

The convergence curve of the GWO algorithm is presented in Fig. 4. The sharp decrease in early iterations shows rapid improvement of initial solutions. The fitness value decreases through the 30th iteration and stabilizes thereafter, indicating convergence to the optimal solution.

To evaluate GWO-based feature selection benefits, RF and SVM classifiers were trained on two feature sets: the full 1024-dimensional vectors (original features) and GWO-selected features. The dataset was split into 70% training and 30% test sets using stratified sampling (random state = 42). The accuracy values reported in the confusion matrices and classification tables (Tables 4–7) refer to this single holdout test set evaluation. In addition, five-fold stratified cross-validation was performed independently to assess model stability, with results reported as mean ± standard deviation. GWO reduced features from 1024 to 405, achieving 60.45% reduction (Table 3).

Classification performance was evaluated using confusion matrices across all four configurations. Figure 5 presents the confusion matrix for RF with original DenseNet121 features.

RF achieved 96.97% accuracy; Bacterial Blight, Citrus Canker, and Dry Leaf were perfectly classified, while Anthracnose and Curl Virus showed minor misclassifications. Table 4 presents detailed precision, recall, and F1-score metrics.

Table 3 Summary of selected feature count and selection rate with Gray Wolf Optimization

Original features Sizes	Selected features size	Feature reduction rate
1024	405	60.45%

Table 4 Classification performance: Random Forest with complete DenseNet121 features

Class	Precision	Recall	F1-score	Support
Anthracnose	0.90	0.90	0.90	30
Bacterial Blight	1.00	1.00	1.00	32
Citrus Canker	0.96	1.00	0.98	53
Curl Virus	0.97	0.89	0.93	35
Deficiency Leaf	1.00	0.97	0.98	58
Dry Leaf	1.00	1.00	1.00	56
Healthy Leaf	0.94	0.98	0.96	63
Sooty Mold	0.94	0.98	0.96	46
Spider Mites	1.00	0.94	0.97	34

Table 4 shows weighted average precision, recall, and F1-score all achieved 0.97, demonstrating strong overall performance. Three classes (Bacterial Blight, Dry Leaf, and Deficiency Leaf) achieved perfect or near-perfect precision (1.00). However, Anthracnose achieved 0.90 across all metrics and Curl Virus obtained 0.89 recall, suggesting that diseases with distinctive visual features are easier to classify, while those with subtle or overlapping symptoms require more discriminative features.

The confusion matrix for SVM with original features (Fig. 6) shows the SVM classifier achieved the highest overall accuracy of 98.30% among all methods. Bacterial Blight, Citrus Canker, Dry Leaf, and Healthy Leaf achieved perfect classification. Minor misclassifications remained in Anthracnose, Curl Virus, and Sooty Mold, likely due to overlapping visual symptoms. Table 5 demonstrates SVM achieved exceptional performance with the complete feature set. Weighted average precision, recall, and F1-score all reached 0.98, surpassing RF performance. Four classes achieved perfect recall (1.00). Anthracnose improved to 0.93 recall compared with RF's 0.90, and Curl Virus increased from 0.89 to 0.94 recall. SVM's strong performance on the full feature set suggests it handles high-dimensional spaces more effectively than RF without requiring dimensionality reduction.

This behavior is consistent with SVM's theoretical properties: the RBF kernel implicitly maps features to a higher-dimensional space, making it naturally suited for dense, high-dimensional inputs such as DenseNet121 features. In contrast, RF relies on random feature subsampling at each split, which may reduce its ability to exploit all 1024 features simultaneously. GWO-based selection mitigates this by retaining only the most discriminative 405 features, allowing RF to achieve more focused and accurate splits.

Figure 7 presents the confusion matrix for RF with GWO-selected features, achieving 97.64% accuracy. The matrix demonstrates strong performance with most classes achieving 93–100% recall. Three classes (Citrus Canker, Dry Leaf, and Healthy Leaf) achieved perfect recall (1.00). Minor misclassifications were observed in Anthracnose (93% recall) and Sooty Mold (93% recall). Compared with the original features, GWO-based selection improved accuracy from 96.97% to 97.64% while reducing the feature set by 60.45%.

Table 6 demonstrates GWO's impact on RF performance. With only 405 selected features, the classifier achieved 97.64% accuracy, a 0.67 percentage point improvement over the complete feature set. Weighted average metrics all reached 0.97. Dry Leaf maintained perfect classification (1.00), while Healthy Leaf achieved perfect recall despite a slight precision decrease. Anthracnose showed improved precision (0.97 vs 0.90) but slightly lower recall (0.93 vs 0.90), suggesting GWO selected more specific features for this disease class. Curl Virus demonstrated enhanced precision (1.00 vs 0.97) while maintaining comparable recall. GWO thus achieved simultaneous dimensionality reduction and accuracy improvement for RF. Figure 8 presents the confusion matrix for SVM with GWO-selected features, achieving 97.34% accuracy. Three classes (Citrus Canker, Dry Leaf, and Healthy Leaf) achieved perfect classification. Compared with SVM using all features (98.30%), accuracy decreased by only 0.96 percentage points despite 60.45% feature reduction, demonstrating the effectiveness of GWO.

Table 7 shows that SVM's performance with GWO-selected features achieved 97.34% accuracy with only 405 features compared to 98.30% with 1024 features. This represents a minimal accuracy reduction (0.96 percentage points) despite a 60.45% decrease in dimensionality. Weighted average metrics all reached 0.97. Four classes maintained near-perfect or perfect recall. Anthracnose showed enhanced recall (0.97 vs 0.93), indicating GWO selected more sensitive features. These results demonstrate substantial dimensionality reduction with minimal performance compromise, favorable for applications requiring reduced computational complexity. Figure 9 compares SVM and RF performance across classes. Both approaches achieve high accuracy, with both models achieving 100% accuracy for Dry Leaf and Healthy Leaf. Other classes show accuracy ranging from 93% to 99%, demonstrating that both models successfully classify plant diseases.

Figure 10 compares training times for classifiers trained on original and GWO-selected feature sets. RF training time was reduced from 13.33 to 7.76 s (41.8% reduction) and SVM from 2.32 to 0.70 s (69.8% reduction) after GWO feature selection. All experiments were conducted on an NVIDIA Tesla T4 16 GB GPU via Google Colab (Python 3.10, TensorFlow 2.12, Scikit-learn 1.2, random state = 42) to ensure reproducibility.

Table 8 presents results compared with studies on the same dataset. Siam et al. achieved 96.19% using DenseNet121. Among proposed models, DenseNet121 + SVM achieved the highest accuracy (98.30%). GWO-optimized models achieved 97.64% (RF) and 97.34% (SVM), demonstrating that the proposed approaches achieve competitive or improved accuracy compared to the baseline study on the same dataset, though direct comparison with other studies is limited due to differences in dataset, class count, and experimental setup.

Gradient-weighted Class Activation Mapping (Grad-CAM) visualization (Fig. 11) identifies regions the model focused on during classification. For the Deficiency Leaf class, the heatmap highlights yellowish discoloration and wilting areas in red and orange, indicating where DenseNet121 focused on detecting nutrient deficiency.

Multi-class receiver operating characteristic (ROC) analysis (Fig. 12) for SVM with GWO-selected features shows most classes achieve high area under the curve (AUC) values (above 0.94), indicating the feature-selection step preserves disease category separability even after dimensionality reduction.

Discussion

To contextualize our approach, we compared results with recent studies on citrus and lemon leaf disease detection (Table 1). Our method achieved 98.30% accuracy using SVM with original features and 97.64% using RF with GWO-selected features, demonstrating competitive performance on this dataset. It should be noted that the dataset used in this study consists of 1354 images across nine classes, which, while sufficient for demonstrating the proposed framework's effectiveness, represents a relatively limited sample size. The high accuracy values may partly reflect the controlled image acquisition conditions of the dataset (Ayoub et al. 2025). Generalization to larger, more diverse, or field-collected datasets remains to be validated in future work. Furthermore, the absence of external validation on an independent dataset is acknowledged as a limitation of this study. The five-fold cross-validation results (standard deviation: 0.44%–0.82%) suggest that overfitting is unlikely within the current dataset; however, external validation would be necessary to confirm the model's robustness and generalizability across different imaging conditions, geographic regions, and lemon cultivars. Among studies using the same dataset, Siam et al. (2024) achieved 96.19% accuracy using DenseNet-121 with end-to-end training. Our approach improves this by 2.11 percentage points while offering three advantages: (1) computational efficiency—traditional machine learning classifier training requires seconds versus hours for deep network fine-tuning, (2) feature interpretability—GWO identifies discriminative feature subsets enabling analysis of critical representations, and (3) deployment feasibility—dimensionality reduction facilitates implementation on resource-constrained edge devices.

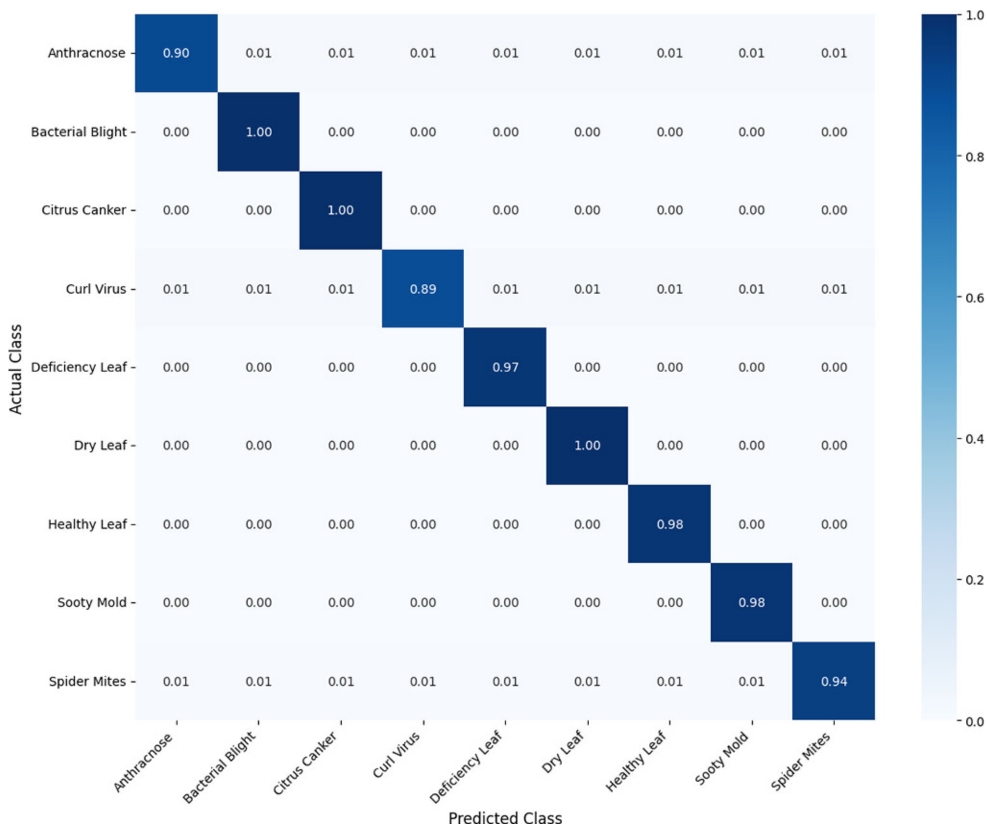


Fig. 5 Normalized confusion matrix for Random Forest with original DenseNet121 features

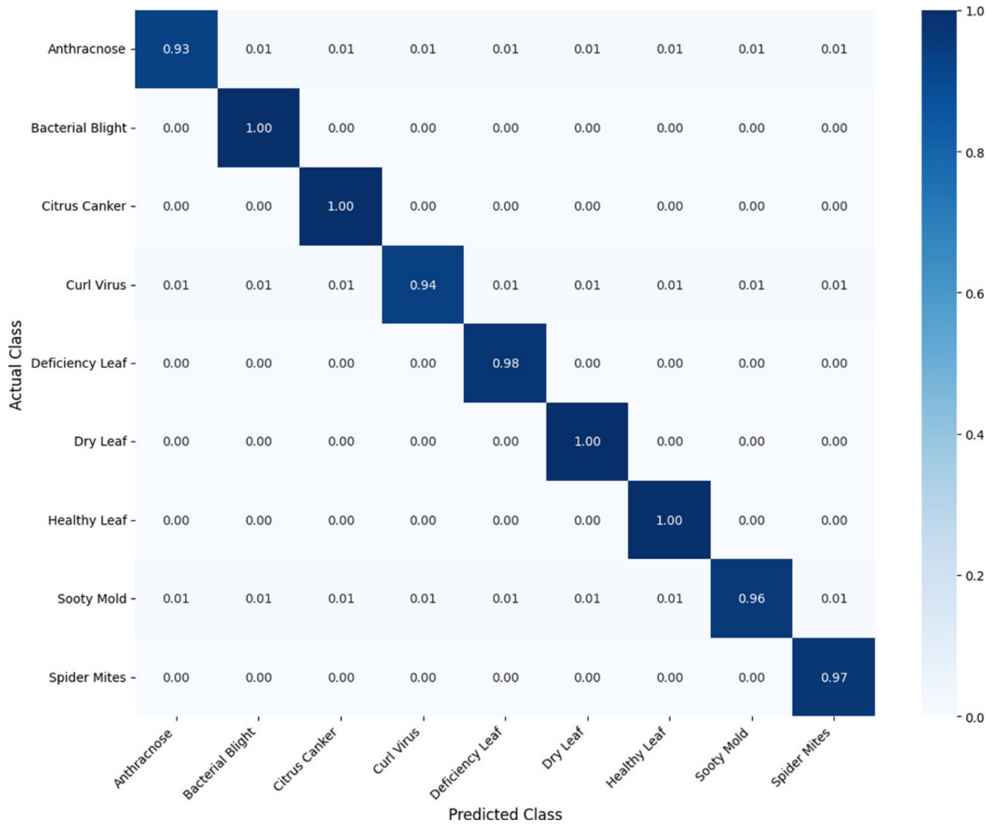


Fig. 6 Normalized confusion matrix for Support Vector Machine using original DenseNet121 features (1024 features)

The effect of GWO-based feature selection on performance and complexity was analyzed through confusion matrices (Figs. 5, 6, 7 and 8). Removing redundant features improved RF performance from 96.97% to 97.64%. DenseNet121's dense connectivity encourages feature reuse across layers, which, while beneficial for gradient flow, tends to produce correlated and partially redundant features in the final 1024-dimensional vector. GWO effectively identifies and discards these redundant dimensions, retaining only the most discriminative subset. This explains why a 60.45% reduction in feature count did not degrade—and in fact slightly improved—RF classification accuracy. For SVM, a slight decrease occurred after feature selection, which can happen when the selected subset excludes features that are individually weak but jointly useful for margin-based decision boundaries. Despite this, reducing the 1024-dimensional feature vector lowers computational cost for training and inference. Multi-class ROC curves (Fig. 12) confirm that class separability is largely preserved after dimensionality reduction.

While some studies report higher accuracies (Islam et al. 2024: 99% with InceptionV3, Zhu et al. 2024: 98.68% with Multi-Models Fusion Network), they address simpler four to six class problems. Our 98.30% accuracy on a nine-class problem represents substantially greater classification difficulty. Unlike studies applying complete deep features or

Table 5 Classification performance metrics for Support Vector Machine using DenseNet121-extracted features without feature selection

Class	Precision	Recall	F1-score	Support
Anthracnose	0.93	0.93	0.93	30
Bacterial Blight	1.00	1.00	1.00	32
Citrus Canker	1.00	1.00	1.00	53
Curl Virus	0.94	0.94	0.94	35
Deficiency Leaf	0.98	0.98	0.98	58
Dry Leaf	1.00	1.00	1.00	56
Healthy Leaf	1.00	1.00	1.00	63
Sooty Mold	0.96	0.96	0.96	46
Spider Mites	0.97	0.97	0.97	34

Table 6 Classification performance metrics for Random Forest with Gray Wolf Optimization-selected features

Class	Precision	Recall	F1-score	Support
Anthracnose	0.97	0.93	0.95	30
Bacterial Blight	0.97	0.94	0.95	32
Citrus Canker	0.96	1.00	0.98	53
Curl Virus	1.00	0.94	0.97	35
Deficiency Leaf	1.00	0.97	0.98	58
Dry Leaf	1.00	1.00	1.00	56
Healthy Leaf	0.93	1.00	0.96	63
Sooty Mold	0.96	0.93	0.95	46
Spider Mites	0.97	0.97	0.97	34

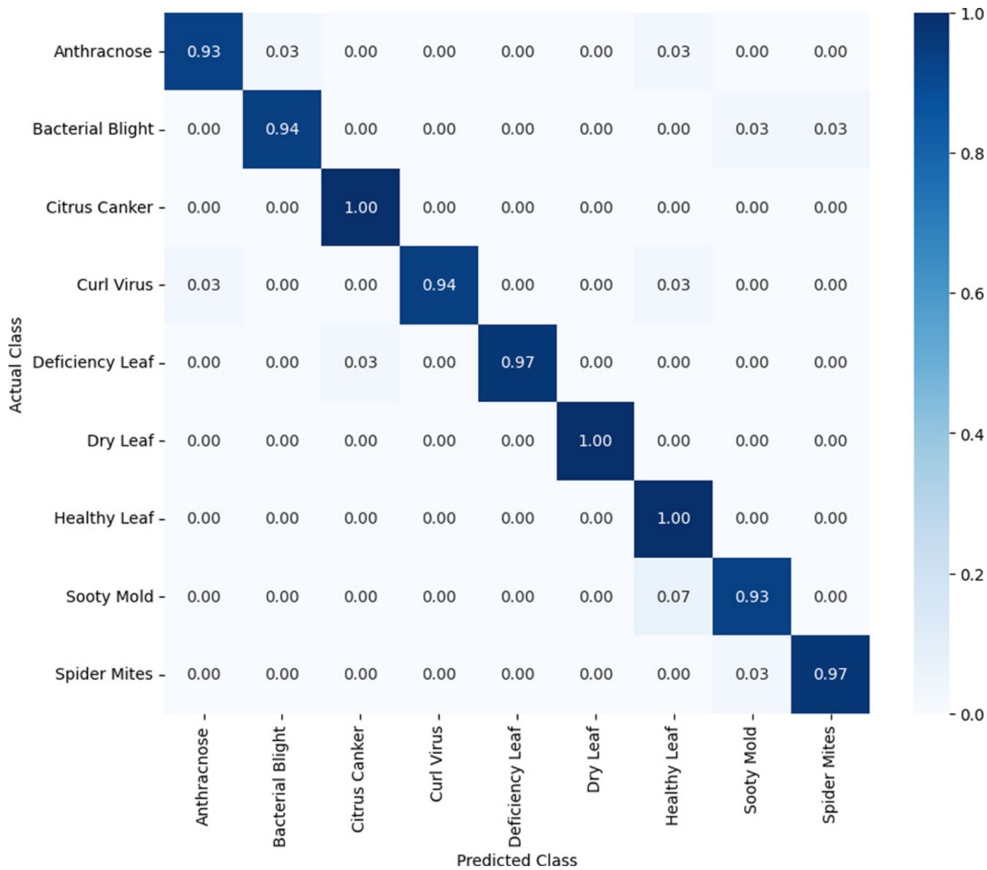


Fig. 7 Confusion matrix obtained from classification using Random Forest after feature selection

end-to-end learning, we demonstrate that bio-inspired optimization can improve performance while reducing dimensionality. This counterintuitive finding suggests that careful feature selection from pre-trained models can match or exceed exhaustive feature usage.

Our contribution demonstrates that systematic optimization of established techniques achieves competitive accuracy while addressing real agricultural constraints: limited training data (1354 images), computational resources (mobile/Internet of Things [IoT] devices), and real-time inference requirements, while recent advances in agricultural imaging continue to open new possibilities for more sophisticated learning frameworks (Haque et al. 2025; Gulzar et al. 2024). This represents a shift from accuracy-maximization to a balance between accuracy and deployability, critical for practical agricultural artificial intelligence (AI) systems.

The integration of DenseNet121 with GWO-based feature selection offers several advantages. First, GWO reduced dimensionality by 60.45% while maintaining high

Table 7 Classification performance metrics for Support Vector Machine with Gray Wolf Optimization-selected features

Class	Precision	Recall	F1-score	Support
Anthracnose	0.94	0.97	0.95	30
Bacterial Blight	0.97	0.97	0.97	32
Citrus Canker	0.96	1.00	0.98	53
Curl Virus	1.00	0.94	0.97	35
Deficiency Leaf	1.00	0.97	0.98	58
Dry Leaf	1.00	1.00	1.00	56
Healthy Leaf	0.98	1.00	0.99	63
Sooty Mold	0.98	0.96	0.97	46
Spider Mites	0.94	0.97	0.96	34

Table 8 Classification accuracy of different methods

Literature	Methods	Classification accuracy (%)
Siam et al.	DenseNet121	96.19
Proposed model	Feature Extraction (DenseNet121) + Random Forest	96.97
Proposed model	Feature Extraction (DenseNet121) + GWO + Random Forest	97.64
Proposed model	Feature Extraction (DenseNet121) + SVM	98.30
Proposed model	Feature Extraction (DenseNet121) + GWO + SVM	97.34

GWO Gray Wolf Optimization, *SVM* Support Vector Machine

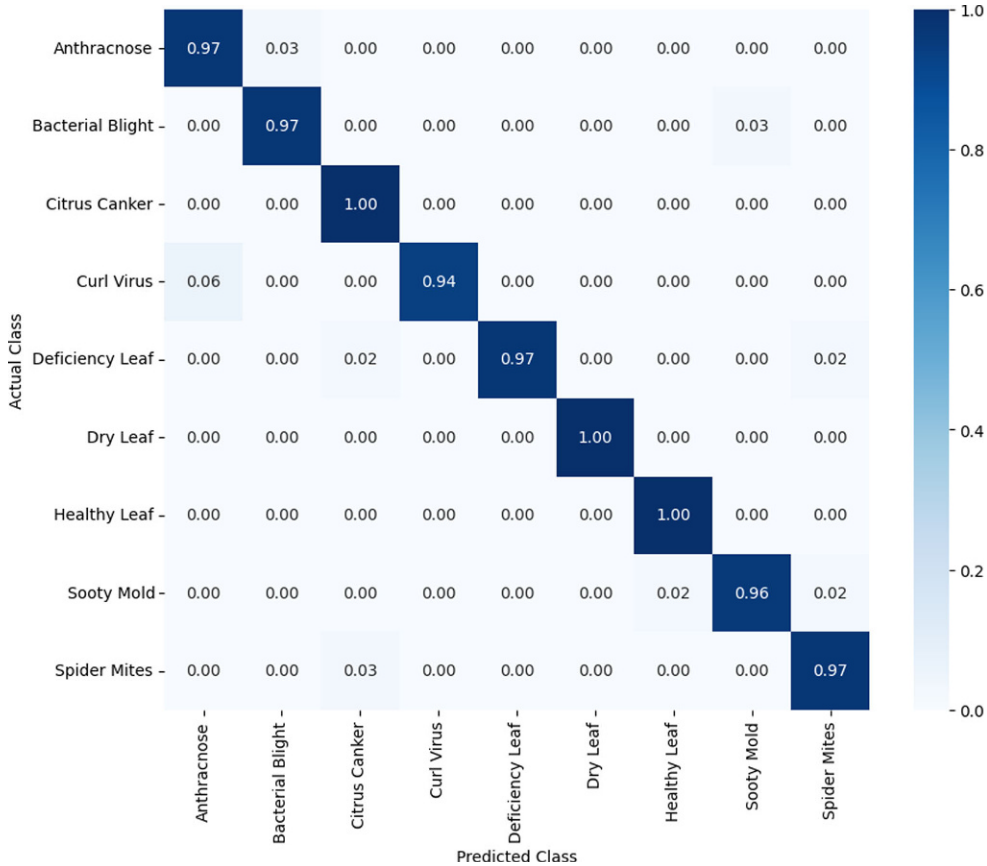


Fig. 8 Confusion matrix of the Support Vector Machine classifier after Gray Wolf Optimization feature selection

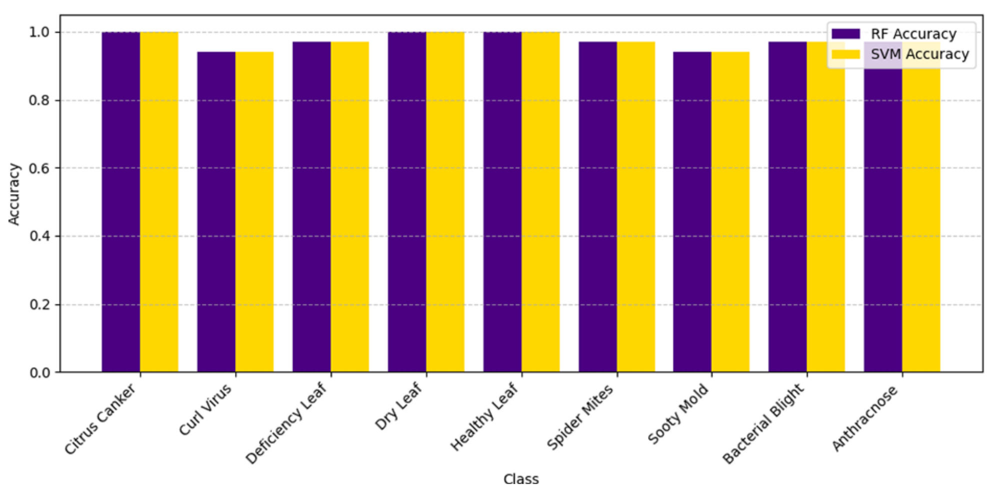


Fig. 9 Random Forest (*RF*) and Support Vector Machine (*SVM*) model classification accuracies for each class

accuracy, improving computational efficiency. Second, our method leverages pre-trained features with traditional classifiers, reducing computational requirements compared to end-to-end deep learning. Third, the approach achieves consistently high performance across multiple classifiers (RF: 97.64%, SVM: 98.30%), demonstrating robustness and generalizability.

GWO-based feature selection provides substantial computational advantages for practical deployment. Training time decreased by 27% for RF (1.92 s → 1.41 s) and 70% for SVM (0.26 s → 0.08 s). Cross-validation time was reduced by 42% for RF and 70% for SVM. These improvements, combined with 60.45% memory reduction, enable real-time on-device processing (~150–200 ms per image) on smartphones and edge devices. For RF, feature selection improved accuracy (+0.67 pp), revealing that deep representations contain redundant features. While SVM experienced a modest reduction in accuracy (–0.96 pp), the accuracy–efficiency trade-off is favorable for deployment scenarios that prioritize computational constraints.

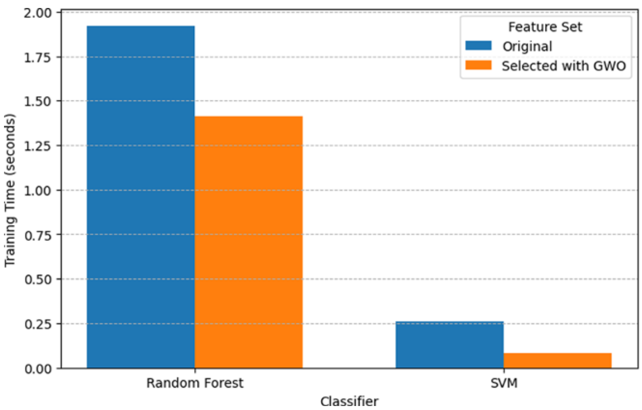


Fig. 10 Training times of the Random Forest and Support Vector Machine (*SVM*) models for both the original and Gray Wolf Optimization (*GWO*)-selected feature sets

Limitations

Despite promising results, this study has several limitations. The publicly available dataset was collected under controlled conditions in a single geographic region (Jamalpur district, Bangladesh) in 2024 and contains 1354 images with 100–210 samples per class. Cross-dataset validation with independent datasets from different geographic locations, climate zones, or lemon varieties was not performed, which limits the assessment of real-world generalization. While this sample size is adequate for traditional machine learning classifiers using pre-extracted features, it remains relatively small for training modern deep learning architectures from scratch.

The dataset also lacks variability typically encountered under field conditions, such as diverse lighting environments, complex backgrounds, partial leaf occlusions, and physical leaf damage. Furthermore, the dataset assumes a single disease label per image. In real agricultural environments, multiple diseases may appear simultaneously on the same leaf, which would require multi-label classification approaches.

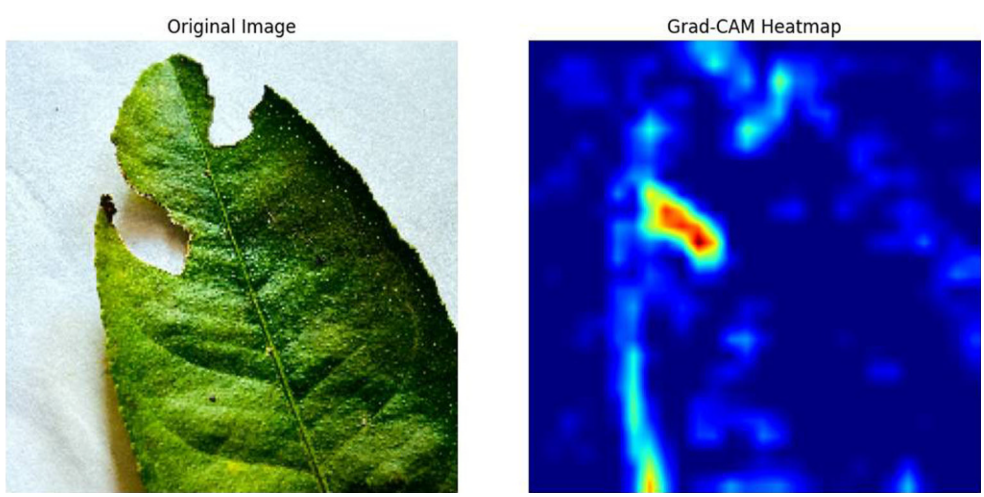
Several methodological limitations should also be acknowledged. Statistical significance tests such as t-tests or Wilcoxon signed-rank tests were not conducted to determine whether performance differences between models were statistically meaningful. In addition, feature selection was limited to the GWO algorithm, without comparison to alternative optimization approaches such as Genetic Algorithms, Particle Swarm Optimization, or traditional dimensionality-reduction techniques including PCA and mutual information. Benchmarking against recent deep learning architectures such as Vision Transformers, EfficientNet, or ConvNeXt was also not performed. Although GWO reduced the feature dimensionality by 60.45%, deployment of the full DenseNet121 feature extraction pipeline on resource-constrained IoT devices would require additional optimization through model compression techniques such as quantization, pruning, or knowledge distillation. Finally, this study focuses on algorithm development and evaluation rather than the implementation of a complete farmer-oriented application with real field validation.

Future Direction

Several research directions could extend this work toward practical agricultural deployment. Future studies should evaluate the proposed framework on larger and more diverse datasets collected under different environmental conditions to better assess generalization capability. Development of multi-label classification frameworks would allow simultaneous detection of multiple diseases occurring on the same leaf.

Further research may also explore systematic a comparison of alternative feature selection and optimization methods to determine the most effective dimensionality-reduc-

Fig. 11 Gradient-weighted Class Activation Mapping (*Grad-CAM*) visualizations for the Deficiency Leaf class. Left: original image, right: Grad-CAM heatmap



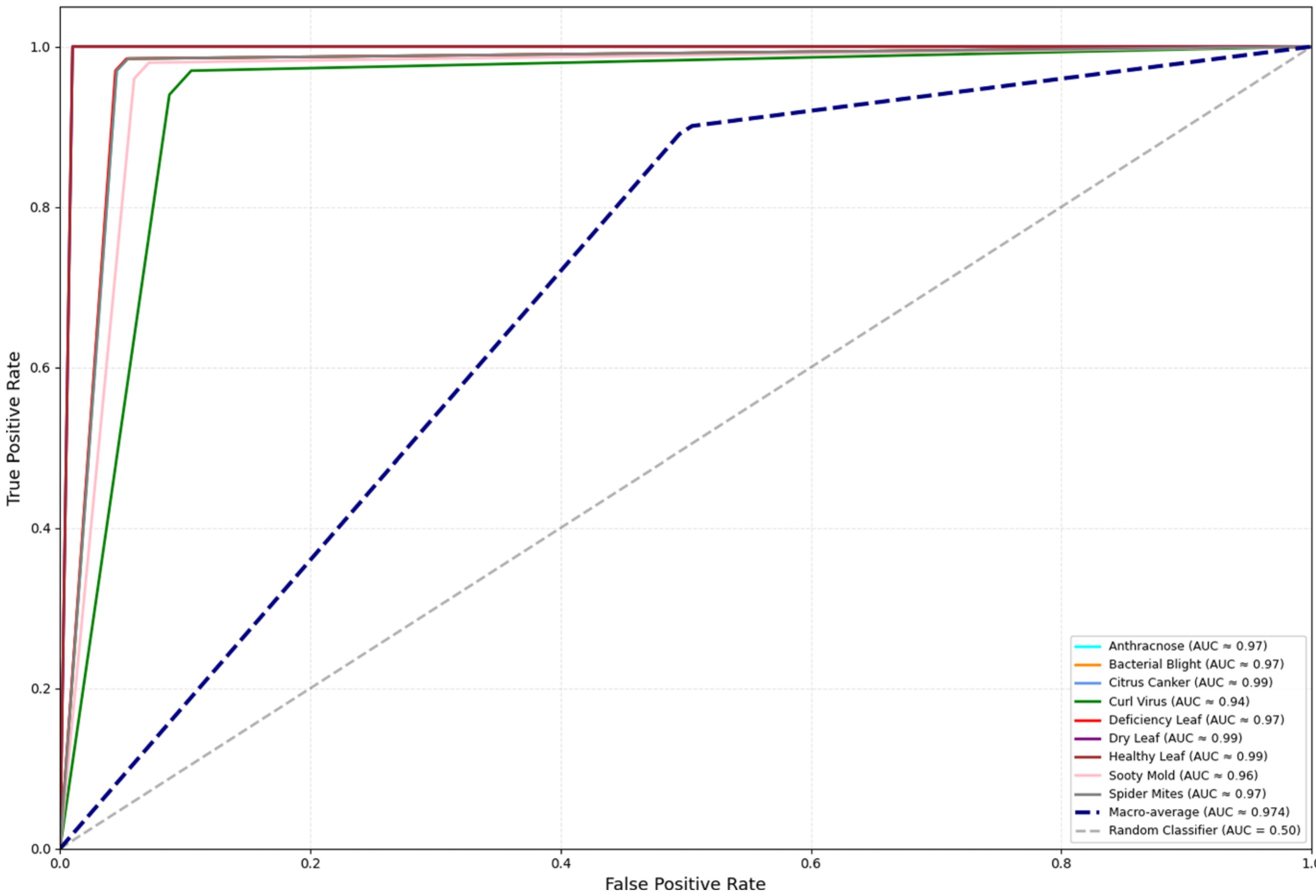


Fig. 12 Multi-class receiver operating characteristic curves for the Support Vector Machine classifier using Gray Wolf Optimization-selected DenseNet121 features. (*AUC* Area under the curve)

tion strategies for plant disease classification. Benchmarking the proposed framework against modern deep learning architectures such as Vision Transformers and EfficientNet would provide additional insight into performance differences between hybrid and end-to-end approaches (Zangana et al. 2025).

In addition, lightweight model optimization techniques should be investigated to support deployment on edge devices and mobile platforms used in agricultural monitoring systems. Integration of the proposed model into mobile applications, combined with field validation studies and collaboration with agricultural extension services, would facilitate translation of the proposed framework into practical tools for farmers and crop management professionals.

Conclusions

This study presented a hybrid framework for classifying lemon leaf diseases using a combination of deep learning feature extraction and machine learning classification. Deep features were extracted from lemon leaf images using a pre-trained DenseNet121 network and subsequently refined through Gray Wolf Optimization (GWO) for feature selection. The selected features were evaluated using Random Forest (RF) and Support Vector Machine (SVM) classifiers. Experimental results demonstrate strong classification performance across both models. RF achieved an accuracy of 96.97% using the original feature set and improved to 97.64% after GWO-based feature selection. SVM achieved the highest performance with 98.30% accuracy using the full feature set and 97.34% using the reduced feature subset. These results indicate that GWO can significantly reduce feature dimensionality (60.45%) while maintaining high classification accuracy.

Gradient-weighted Class Activation Mapping (Grad-CAM) visualization further confirmed that DenseNet121 effectively focuses on disease-relevant regions of the leaf images, improving interpretability of the model predictions. In addition, GWO-based feature selection reduced training time, storage requirements, and computational complexity, demonstrating the potential of the proposed framework for efficient agricultural disease detection. Overall, the results show that combining deep feature extraction with wrapper-based feature selection provides a practical and efficient approach for plant disease classification. The proposed method offers a balance between classification accuracy and computational efficiency, which is important for real-world agricultural monitoring systems.

Funding This work was supported by the Deanship of Scientific Research, the Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under Grant No. KFU261318.

Author contributions Yavuz UNAL and Yonis Gulzar conceived and designed the study. Yavuz UNAL conducted the experiments, analyzed the data, and drafted the manuscript. Yonis Gulzar provided scientific guidance, critically reviewed the methodology and results, and contributed to manuscript revision. Both authors read and approved the final manuscript.

Data Availability The data that support the findings of this study are openly available in Mendeley Data athttps://doi.org/10.17632/44nrn4593f.1

Competing interests The authors declare no competing interests.

Ethics, Consent to Participate, and Consent to Publish Declarations Not applicable.

References

Akash K, Atik A, Khan N, Rahim MA, Alam J, Bijoy HI, Bhuiyan T (2024) Machine learning-driven approch of crop selection for sustainable agriculture in Bangladesh. https://doi.org/10.1109/ICCIT64611.2024.11021993

Al Fahim H, Hasan MA, Bijoy MHI, Reza AW, Arefin MS (2023) Seeds classification using deep neural network: a review. In: Vasant P et al (ed) Lecture notes in networks and systems, vol 852. Springer, Cham https://doi.org/10.1007/978-3-031-50330-6_17

Ayoub S, Baig I, Ashraf M, Okasha M (2025) Impact of dataset quality on deep learning models for dragon fruit and leaf health classification. Impact Agric 1:1. https://doi.org/10.65500/agriculture-2025-001

Barman U, Choudhury RD, Sahu D, Barman GG (2020) Comparison of convolution neural networks for Smartphone image based real time classification of citrus leaf disease. Comput Electron Agric 177:105661. https://doi.org/10.1016/j.compag.2020.105661

Bitto AK, Bijoy MI, Das A, Rahman M, Rabbani M (2023) Potato-net: classifying potato leaf diseases using transfer learning approach. In: Satu MS, Moni MA, Kaiser MS, Arefin MS (eds) Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 490. Springer, Cham https://doi.org/10.1007/978-3-031-34619-4_1

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Butt N, Iqbal MM, Ramzan S et al (2025) Citrus diseases detection using innovative deep learning approach and hybrid meta-heuristic. PLoS ONE 20(1):e316081. https://doi.org/10.1371/journal.pone.0316081

Çetiner H (2022) Citrus disease detection and classification using based on convolution deep neural network. Microprocess Microsyst 95:104687. https://doi.org/10.1016/j.micpro.2022.104687

Dang-Ngoc Hanh, Cao Trang N.M, Dang-Nguyen Chau (2021) Citrus leaf disease detection and classification using hierarchical support vector machine https://doi.org/10.1109/ISEE51682.2021.9418680

Duth PS, Bhat SG (2022) Disease classification in citrus leaf using deep learning. IEEE https://doi.org/10.1109/ICDSIS55133.2022.9915847

Gambella C, Ghaddar B, Naoum-Sawaya J (2021) Optimization problems for machine learning: a survey. Eur J Oper Res 290(3):807–828. https://doi.org/10.1016/j.ejor.2020.08.045

Gulzar Y (2025a) Applications of transfer learning in sunflower disease detection: advances, challenges, and future directions. Turk J Biol 49(5):534–549. https://doi.org/10.55730/1300-0152.2763

Gulzar Y (2025b) Papaya leaf disease classification using pre-trained deep learning models: a comparative study. Appl Fruit Sci 67(4):1–10. https://doi.org/10.1007/S10341-025-01533-1

Gulzar Y (2025c) Papnet: an AI-driven approach for early detection and classification of papaya leaf diseases. Appl Fruit Sci 67(4):1–11. https://doi.org/10.1007/S10341-025-01466-9

Gulzar Y, Ünal Z (2025) Optimizing pear leaf disease detection through PL-densenet. Appl Fruit Sci 67(1):1–13. https://doi.org/10.1007/s10341-025-01265-2

Gulzar Y, Ünal Z, Kızıldeniz T, Umar UM (2024) Deep learning-based classification of alfalfa varieties: a comparative study using a custom leaf image dataset. MethodsX 13:103051. https://doi.org/10.1016/J.MEX.2024.103051

Haque Z, Koklu M, Mirza MA, Omar M, Mukhayyokhon S (2025) Mitigating spatial scale loss in CNN-based fine-grained image classification: application to date fruit grading. Impact Agric 1:5. https://doi.org/10.65500/agriculture-2025-005

Haridasan A, Thomas J, Raj ED (2023) Deep learning system for paddy plant disease detection and classification. Environ Monit Assess 195(1):120. https://doi.org/10.1007/s10661-022-10656-x

Iqbal Z, Khan MA, Sharif M, Shah JH, Ur Rehman MH, Javed K (2018) An automated detection and classification of citrus plant diseases using image processing techniques: a review. Comput Electron Agric 153:12–32. https://doi.org/10.1016/j.compag.2018.07.032

Islam KMT, Symoom F, Ahmed R, Islam S, Islam DA, Imam Bijoy H (2025) Dragon Fruit Leaf Disease Detection With Transfer Learning. In: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), Rangpur, Bangladesh, pp 1–6. https://doi.org/10.1109/QPAIN66474.2025.11172103

Islam S, Chakrabarty N, Uddin MS (2024) Citrus leaf disease detection through deep learning approach. Bull Electr Eng Informatics 13(4):2643–2649. https://doi.org/10.11591/eei.v13i4.4521

Kotwal J, Kashyap R, Pathan S (2023) Agricultural plant diseases identification: from traditional approach to deep learning. Mater Today: Proc 80(Part 1):344–356. https://doi.org/10.1016/j.matpr.2023.02.370

Kotwal J, Kashyap R, Pathan MS, Kimbahune V (2024) Enhanced leaf disease detection: unet for segmentation and optimized efficientnet for disease classification. Softw Impacts 22:100701. https://doi.org/10.1016/j.simpa.2024.100701

Kotwal JG, Kashyap R, Shafi PM (2024) Artificial driving based efficientnet for automatic plant leaf disease classification. Multimed Tools Appl 83:38209–38240. https://doi.org/10.1007/s11042-023-16882-w

Lu Y, Yu X, Hu Z, Wang X (2025) Convolutional neural network combined with reinforcement learning-based dual-mode Grey Wolf optimizer to identify crop diseases and pests. Swarm Evol Comput 94:101874. https://doi.org/10.1016/j.swevo.2025.101874

Luaibi AR, Salman TM, Miry AH (2021) Detection of citrus leaf diseases using a deep learning technique. Int J Electr Comput Eng (IJECE) 11(2):1719–1727. https://doi.org/10.11591/ijece.v11i2.pp1719-1727

Mahmud P, Ali A, Akter S, Bijoy HI (2022) Lychee tree disease classification and prediction using transfer learning. In: Proc. ICCCNT 2022 https://doi.org/10.1109/ICCCNT54827.2022.9984286

Mia J, Bijoy HI, Uddin S, Raza DM (2021) Real-time herb leaves localization and classification using YOLO https://doi.org/10.1109/ICCCNT51525.2021.9579718
Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007
Mirjalili S et al (2025) Optimized feature selection using Gray Wolf and particle swarm algorithms for corn seed image classification. J Food Compost Anal. https://doi.org/10.1016/j.jfca.2025.107738
Mohan S, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554. https://doi.org/10.1109/ACCESS.2019.2926593
Niu YH, Wang L, Wan XG, Peng QZ, Huang Q, Shi ZH (2021) A systematic review of soil erosion in citrus orchards worldwide. CATENA 206:105558. https://doi.org/10.1016/j.catena.2021.105558
Pramanik A, Khan AZ, Biswas AA, Rahman M (2021) Lemon leaf disease classification using Cnn-based architectures with transfer learning. IEEE
Reddy KRG, Thirunavukkarasu KS (2024) Deep learning approach for disease detection in sweet lemon leaves to enhance citrus fruit yield. IEEE https://doi.org/10.1109/ICECA63461.2024.10800763
Sarker T, Sikder M, Datta S, Noyan SI, Biswas N, Bijoy HI (2025) Automated flower species detection and classification using image preprocessing and deep learning https://doi.org/10.1109/QPAIN66474.2025.11171999
Saygili A (2023) The efficiency of transfer learning and data augmentation in lemon leaf image classification. Eur J Eng Appl Sci 6(1):32–40. https://doi.org/10.55581/ejeas.1321042
Sharif M, Khan MA, Iqbal Z, Azam MF, Lali MIU, Javed MY (2018) Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection. Comput Electron Agric 150:220–234. https://doi.org/10.1016/j.compag.2018.04.023
Shoaib M et al (2023) An advanced deep learning models-based plant disease detection: a review of recent research. Front Plant Sci 14:1158933. https://doi.org/10.3389/fpls.2023.1158933
Siam et al., (2024) Comprehensive lemon leaf disease dataset for advanced detection and sustainable agriculture. Mendeley Data. https://doi.org/10.17632/44NRN4593F.1
Siam et al., (2025) A comprehensive image dataset for the identification of lemon leaf diseases and computer vision applications. Data Brief 58:111244. https://doi.org/10.1016/j.dib.2024.111244
Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 134:93–101. https://doi.org/10.1016/j.eswa.2019.04.042
Suryavanshi A, Kukreja V, Dogra A, Joshi J (2023) Advanced ABS disease recognition in lemon—a multi-level approach using CNN and random forest ensemble. IEEE https://doi.org/10.1109/ICTACS59847.2023.10389937
Tanzil Islam K, Symoom F, Ahmed R, Islam S, Islam DA, Bijoy HI (2025) Dragon fruit leaf disease detection with transfer learning https://doi.org/10.1109/QPAIN66474.2025.11172103
Yasin ET, Ropelewska E, Yasin R (2025) Optimized feature selection using Gray Wolf and particle swarm optimization for classification of corn varieties. J Food Compost Anal 145:107738. https://doi.org/10.1016/j.jfca.2025.107738
Yu J-R, Lin C-Y, Lien D (2025) Cost-effective and accuracy-oriented -norm support vector machine for enhanced feature selection. Measurement 253:117506. https://doi.org/10.1016/j.measurement.2025.117506
Yuan Y (2023) Computer vision and deep learning for precise agriculture: a case study of lemon leaf image classification. J Phys Conf Ser. https://doi.org/10.1088/1742-6596/2547/1/012024
Zangana HM, Li S, Wani S (2025) Diffusion models for agricultural imaging: a systematic review of methods, applications and future prospects. Impact Agric 1:3. https://doi.org/10.65500/agriculture-2025-003
Zhu H, Wang D, Wei Y, Zhang X, Li L (2024) Combining Transfer Learning and Ensemble Algorithms for Improved Citrus Leaf Disease Classification. Agriculture 14(9):1549. https://doi.org/10.3390/agriculture14091549

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.