RESEARCH PAPER

A Resource-Efficient Framework for Plant Disease Classification Using Classical Image Features

Pranay Meshram[1], Priya Meshram[2], Anshuman Kalia[3], Omprakash Kushwaha[4], Yogendra Pal[5], Komal Sharma[6], Anuj Pathak[7*]

[1,3,4,5]Department of Computer Science & Engineering, Krishna Institute of Engineering & Technology (KIET), Ghaziabad, Delhi-NCR, UP, India

[2]Department of Computer Science & Engineering, YCCE, Nagpur, MH, India

[6]Department of Management, Krishna Institute of Engineering & Technology (KIET), Ghaziabad, Delhi-NCR, UP, India

[7*]Department of Pharmaceutics, Krishna Institute of Engineering & Technology (KIET), Ghaziabad - 201206, Delhi-NCR, Uttar Pradesh, India

***Corresponding author: Anuj Pathak, Department of Pharmaceutics, Krishna Institute of Engineering & Technology (KIET), Ghaziabad - 201206, Delhi-NCR, Uttar Pradesh, India**

Email: anuj.pathak1@gmail.com

Received: 25th May, 2026; **Revised:** 6th June, 2026; **Accepted:** 8th June, 2026; **Available Online:** 20th June, 2026

ABSTRACT

Plants can suffer a number of diseases that impact agricultural productivity and food security, particularly in developing farming communities. Although deep learning is capable of classification of diseases with outstanding results, its use is limited due to the difficulty of obtaining large labeled databases and the high requirement of computation. To address these challenges, this study proposes a new method for plant disease classification based on traditional image processing and machine learning algorithms with lightweight and low computation requirements. This one uses several handcrafted descriptors such as color histograms, Haralick texture features and Hu moments to retrieve the information relevant to a disease from the segmented leaf images. Performance of top five classifiers, namely Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression and Naïve Bayes classifiers are evaluated from the dataset of healthy plant leaves and diseased plant leaves images on 10-fold cross validation. Based on the results of the research work, the best classification model was the Random Forest Classifier model with the accuracy value is 98.12%, 0.98 precision, 0.98 recall, and 0.98 F1 value. The proposed solution was also found to be uncomputation complex and low memory consuming and can be made realtime inference. Therefore, this solution can be implemented onto agricultural systems, considering edge computing and IoT. The results of the research also demonstrated that feature-based machine learning approaches afford interpretable and reliable plant disease detection at a low computation cost, further contributing to sustainable, and precision agriculture.

Keywords: Plant Disease Detection, Machine Learning, Feature-based classification, Precision agriculture, Edge computing, Computational efficiency.

How to cite this article: Meshram P, Meshram P, Kalia A, Kushwaha O, Pal Y, Sharma K, Pathak A. A Resource-Efficient Framework for Plant Disease Classification Using Classical Image Features. Int J Drug Deliv Technol. 2026;16(62s): 848-861. DOI: 10.25258/ijddt.16.62s.94

Source of support: Nil.

Conflict of interest: None.

1. INTRODUCTION

Plant diseases continue to be a serious constraint on crop yields and food security, not only in developing and resource poor countries of the world, but also all over the world due to the scarcity of technology in diagnostic skills. Early detection of plant diseases is vital since these can result in significant losses in yield quantity and quality of the harvest, increased use of pesticides, and economic losses when the disease occurs. It was found recently, deep learning techniques have been successfully used for identification of plant diseases from image data (with the capability of automatically learning the disrimination features using large scales of data sets). However, the methods generally rely on a large number of annotated data, a considerable amount of computing power and a GPU chip and device, making their application difficult to implement in low resource agricultural contexts.

In order to mitigate these limitations, lightweight machine learning methodologies which rely on hand-designed image features have received more and more attention. Traditional Feature-Based approaches rely on support vectors with interpretable features like color, texture and shape information as descriptors of disease patterns while they have a lower computational complexity, they are facing many limitations. When it comes to interpreting the results and requiring minimal to no training time, and being easily deployable on low-

power systems, classical machine learning models like Random Forest and Support Vector Machine are better than deep learning models. Keeping all the abovementioned points in mind, the present work proposes a computational solution to the plant disease diagnosis problem, utilizing the classic image processing algorithms and machine learning classifiers. Colour histogram, Hu moment and Haralick features are said take out a variety of visual information from plant leaves particularly when combined. Experimental evaluation of performance of five state-of-the-art machine learning classifiers of Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression and Naïve Bayes models has been carried out under equal conditions using cross validation.

The novelty of the proposed solution is the fact that the high classification performance can be achieved with much lower computational resource than the deep learning models with relatively simple visual features used as inputs to the machine learning algorithms. As opposed to most other related papers which pay attention only to classification accuracy, the present approach involves other important criteria, such as model interpretability, computation time, memory capacity, and suitability for edge-computing/IoT environments. Results show that the Random Forest algorithm outperforms all existing machine learning model tested in the experiment and improvements of its accuracy have been obtained about experimental value of 98.12% on a standard CPU platform.

Beyond this, the results of classification are presented and validated using the nonparametric statistical method.

RELATED WORK

Automated plant disease detection systems have evolved to a greater extent over the last few years due to advancements in artificial intelligence and computer vision. Current research is mainly aimed at deep learning methods due to their good capability of feature learning and high classification accuracy. But there are some issues arisen associated with computational complexity, reliance on data, poor interpretability, and in resource limited agricultural settings.

Using PlantVillage dataset, Mohanty et al. [1] proved the ability of deep learning to detect plant diseases from images by implementing convolutional neural networks. Though their work demonstrated that CNNs can be applied for automatic diagnosis in agricultural settings, it needed a huge amount of training samples and computational resources. Similarly, Ferentinos [2] studied various deep learning algorithms and architectures to diagnose plant diseases and found good classification accuracy for various crops. The models performed well in predicting the amount of total fertilizer nitrogen in forage and silage, but proved too complex to implement in practice and GPU dependent in low-resource settings.

Too et al. [3] has been of great interest if a fine-tuned deep learning model is suitable for the plant disease identification problem or not, and found that there was a significant improvement in the performance of classification by employing the transfer learning techniques. However, the study pointed to the high cost of deep neural networks computation. Sladojevic et al. [4] presented a deep neural network-based framework for leaf disease recognition that can be capable of automatic leaf classification into healthy and unhealthy leaves. Although achieving good results, the approach relied on labeled data that was comparatively large in size, as well as intensive training algorithms.

The in-field wheat disease diagnosis system has been designed by Lu et al. in 2018, based on computer vision and machine learning technologies [5]. Their framework showcased the viability of automation of disease detection under practical farming circumstances. Bedi and Gole [6] suggested a hybrid deep learning system which consists of convolutional autoencoder and convolutional neural network for plant disease detection. The hybrid approach enhanced the classification ability but the complexity of the model and computational requirements remained reasonably high.

Ramesh et al. [7] dealt with the works based on machine learning for detection of plant disease using traditional classifiers and image processing technique. Their results showed that classical machine learning techniques can effectively recognize the disease with lower complexity than the deep learning techniques. Agarwal et al. [8] developed an effective CNN based classification model for tomato disease detection, but the model needed to be trained with considerable resources and computational power.

Khamparia et al [9] used deep learning classification models for agriculture image analysis purposes and showed the performance capability of feature extraction using a neural network. LeCun et al. [10] presented one of the scientific papers that outlines the theoretical aspects and field applications of deep learning architectures for different fields, such as image recognition. Kamilaris and Prenafeta-Boldú [11] presented a complete review of the use of deep learning for agriculture and the growing trend to utilize AI systems for crop monitoring and identification of diseases.

Li et al. [12] summarized existing plant disease detection approaches, which are based on deep learning, and pinpointed some challenges of model interpretability, computational expense, and the feasibility of deployment in practice. Wang et al. [13] performed research on image recognition techniques of plant disease using neural networks,

Study	Methodology	Dataset Size / Type	Key Strengths	Limitations
Mohanty et al. (2016) [1]	CNN-based deep learning	Large (PlantVillage)	High classification accuracy	Requires large datasets and GPU resources
Ferentinos (2018) [2]	Deep CNN architectures	Large multi-crop dataset	Strong disease recognition capability	High computational complexity
Too et al. (2019) [3]	Transfer learning CNNs	Medium	Improved classification performance	Computational overhead
Sladojevic et al. (2016) [4]	Deep Neural Networks	Medium	Automated disease recognition	Large training requirement
Lu et al. (2017) [5]	In-field disease diagnosis system	Real-field wheat dataset	Practical field deployment	Crop-specific limitation
Bedi and Gole (2021) [6]	Hybrid CAE + CNN	Medium	Enhanced feature learning	Complex architecture
Ramesh et al. (2018) [7]	Classical ML + image processing	Small	Lightweight and interpretable	Limited scalability
Agarwal et al. (2020) [8]	CNN model	Medium	High tomato disease accuracy	High training cost

proving neural network's ability to plant disease recognition in agricultural image analysis. Patil and Kumar [14] used feature-based machine learning methods for recognition of plant diseases, and concluded that handcrafted descriptors are beneficial for obtaining a lightweight, interpretable classification system.

Nishat Singh and Jain [15] presented an easy-to-implement machine learning-based system dedicated to plant disease identification for resource-poor agricultural settings. Their investigation stressed the edge deployment practicability and computational efficiency. The researchers also studied lightweight yet effective machine-learned techniques for smart agricultural systems and successfully tested the lightweight models of artificial intelligence (AI) to be applied in IoT-driven agri-platforms, as in Islam et al. [16].

Alzubaidi et al. [17] discussed the concepts of deep learning, CNN architectures and related challenges with a particular focus on limitations like high computations cost, complexity of the model, and the dependency on vast quantities of data. Sharma et al. in [18] thoroughly reviewed applications of machine learning in precision agriculture and elaborated on the emerging smart agricultural automation system powered using AI. In the field of plant diseases, Zhang et al. [19] explored deep learning approach and applicability of CNN based classification models and showed that the CNN based classification models used in the field of plant disease have proven to be successful.

The capability of efficient object detection demonstrated by Xiang et al. [20] indicates that ensemble learning methods are robust and scalable, as well as being used strategically in agriculture. To address the limitations of the traditional RF method, Ren et al. [21] designed a new flexible and stable RF algorithm, demonstrating the effectiveness of ensemble learning method in solving complex classification problems. Verma and Mehta [22] explored explainable machine learning techniques for plant disease detection and the significance of explainability in agriculture under retrospect.Under retrospect, Verma and Mehta [22] explored explainable machine learning techniques for plant disease detection and the importance of the explainability in agriculture. Singh and Patel [23] suggested hybrid models of machine learning to detect plant diseases robustly, and successfully introduced new integration of lightweight machine learning strategies with efficient feature representation.

Table 1 summarizes the comparative analysis of recent plant disease detection approaches in terms of methodology, dataset characteristics, strengths, and limitations.

Khamparia et al. (2019) [9]	Deep learning framework	Medium	Effective feature extraction	Computationally expensive
Kamilaris and Prenafeta-Boldú (2018) [11]	Survey on deep learning in agriculture	Review study	Comprehensive agricultural AI analysis	No experimental validation
Li et al. (2021) [12]	Deep learning review	Review study	Broad comparative analysis	Limited deployment discussion
Patil and Kumar (2022) [14]	Feature-based ML	Small	Interpretable classification	Lower generalization
Singh and Jain (2023) [15]	Lightweight ML framework	Small	Edge-compatible deployment	Binary classification limitation
Islam et al. (2024) [16]	Lightweight ML for IoT agriculture	Medium	Smart agriculture applicability	Limited disease diversity
Verma and Mehta (2024) [22]	Explainable ML framework	Medium	Improved interpretability	Increased model complexity
Singh and Patel (2023) [23]	Hybrid ML models	Medium	Robust disease classification	Additional optimization required
Proposed Work	Classical features + Random Forest	Small (800 images)	High accuracy (98.12%), lightweight, interpretable, low computational cost	Binary classification (acknowledged)

2.1 Key Insights from the Literature

Detecting plant disease shows several important research trends and gaps in the literature aspect presented in tab

1. However, most modern studies concentrate their research on deep learning-based approach because of obtaining a high classification accuracy based on them. Overall, these approaches need a lot of computational power and vast amounts of annotated data and customized GPU hardware, making them impractical in constrained agricultural contexts. The advantages of the lower complexity of classical machine learning techniques, faster training time, and reduced need for resources for deployment are less valued, and consequently comparatively less attention has been paid to these methods. In addition, many of the current deep learning models deploy as "black box" decision systems close to the farmer or the agricultural subject, which cannot be easily understood by the farmers or agriculture practitioners to understand the predictions obtained by the model. Gallouji and Jboaib carried out another interesting study from the literature, in which they studied the performance of several machine learning classifiers but did not work in the same experimental environment. This underscores the importance of light, easy to understand and simple-to-compute models which can provide accurate disease detection in agricultural settings.

2.2 Placement of the presented Work

The recent research works have concentrated on the application of deep learning technique to diagnose plant diseases. But because they require much data and a lot of computation, they are less viable to use. Although classic machine learning algorithms are efficient, their in-depth exploration and comparison with other algorithms in the present literature was not well examined. We will stick to a computationally efficient feature-based classification algorithm in this work keeping the balance point between accuracy and interpretability of the algorithm. Comparisons of the various classic machine learning techniques are made under comparable conditions. Demonstrating in the process, the ability to attain high level of accuracy of the disease detection with computational efficient algorithm. While deep learning approaches prevail in

recent years in the literature, relatively little attention in recent literature has been paid to more lightweight feature-based machine learning systems that can be deployed in agricultural environments with limited resources. Current work has focused on accuracy but not interpreted, didn't consider computational efficiency, latency to deploy, or hardware requirements. This work is addressing these open questions by existing classical machine learning framework validated on the basis of statistics and efficient from computation point of view.

2. METHODOLOGY

The systematic plant diseases classification process is a multi-stage one, which involves image pre-processing, feature extraction, training and evaluation of a plant disease classification model.

3.1 Data Preprocessing and Segmentation

The input images have been pre-processed by data standardization and resizing of data. Colour-space transformation from RGB to HSV space is conducted to make sure that intensities and colour values are independent. The segmentation, as a technique, allows to extract the leaves only part of an image.

3.2 Feature Extraction

The proposed framework extracts three types of handcrafted image features (i.e., color, shape and texture descriptors) that most appropriately capture the complementary aspects (color, shape and texture) of diseased plant leaves. Color features are extracted using color histograms which represent chromatic variation and discoloration patterns, that are typical of plant diseases. The features are used to detect abnormal colour distribution including any yellowing, browning or spotting symptoms of infected leaves. Moreover, Hu moments are used to extract shape features, which are rotation, scale and translation invariant, and have robust geometric and structural information on infected regions. This helps in correct identification of the disease patterns affected on leaves that are irregular in shape. Moreover, texture features are obtained from Haralick descriptors quantifying spatial texture variations and surface irregularities associated with disease severity and progression. These complementary feature sets can improve the overall discriminative capability of the proposed model, which further helps in the overall disease classification capability.

This results in an overall feature vector for the classifier, combining the feature vectors.

The complete list of features and their dimensions is presented in the Supplementary Material (Table S2).

The entire pipeline of the framework proposal is presented in Algorithm 1 in the Supplementary Material S3.

3.3 Model Training and Validation

Based on the reviewed literature, five classifiers are considered for the effectiveness of the traditional machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naïve Bayes (NB) for plant disease detection. These models feature implementation of various learning schemes, such as Ensemble learning, Margin based type of classifier, Instance based learning, Probabilistic type of classifier, Linear classification.

Feature Vectors of color, texture and shape features are used to generate the same sets of feature vectors for each of the 5 classifiers. To assess the efficiency of the model, 10 folds cross validation has been used. In this approach, all the data has been split into 10 parts, and each time nine parts of them have been used as training data for the classifier, and the other one as a test data. It has been carried out repeatedly such that at any time one of the datasets is used as the test set. The average score has been used for reporting the final result to eliminate the risk of overfitting and generate unbiased performance estimates, especially when dealing with small datasets.

3.3.1 Hyperparameter Configuration

The optimisation of the hyperparameters of the different machine learning classifiers used was carried out in an empirical manner so as to satisfy the classification task requirements for optimal accuracy and simultaneous computational efficiency. Implementing the Random Forest classifier, it used 100 decision trees, maximum depth of 20, and using Gini impurity for splitting criterion. The Support Vector Machine (SVM) classifier used Radial Basis Function (RBF) kernel, gamma of "scale" and regularization parameter C=1.0. The k=5, Euclidean distance version of K-Nearest Neighbors (KNN) Classifier was used. Logistic Regression had been fit using L2 regularisation and a max of 1000 iterations. Gaussian Naïve Bayes was done with all the default, and naïve, assumptions for probability. The stability of the RSF under 10-fold cross validation and the classification accuracy were used experimentally to choose the hyperparameters. The performance comparison of the classifiers is shown in Table 2. Some of these metrics are precision, recall, F1 score, and accuracy. These results show the effectiveness of the proposed framework and demonstrate the improved performance of the Random Forest classifier.

Table 2. Hyperparameter Configuration of Machine Learning Models

Classifier	Hyperparameters
Random Forest	n_estimators = 100, max_depth = 20, criterion = gini
SVM	kernel = RBF, C = 1.0, gamma = scale
KNN	n_neighbors = 5, metric = Euclidean
Logistic Regression	penalty = L2, max_iter = 1000
Gaussian Naïve Bayes	Default parameters

Hyperparameter values were selected to balance classification accuracy, computational complexity, and lightweight deployment feasibility for edge-compatible agricultural systems.

3.4 Performance Evaluation Metrics

The proposed method was verified numerically using some classification metrics generated from the confusion matrix such as accuracy, precision, recall and F1 score. All these measures enable comprehensive assessment on classification quality, classification discriminative power and robustness of the system.

The general classification efficiency is measured by the number of correct samples and is shown as the ratio of correct samples to the total sampled data set, which is accuracy.

Precision is used for the reliability of positive classification by calculating the ratio of the number of samples of the target class that are correctly classified to the total number of samples that are classified positive.

However, in the agricultural context, recall describes the model's performance in correctly categorizing samples as a disease, which could lead to the delay of required actions in the field if the disease was not identified. Never use any of the following: F1 score, harmonic average of precision and of recall.

But that exercise of statistical significance analysis is clearly separate from these descriptive numbers. This was achieved by employing Wilcoxon signed-rank test on the cross-validation results for determining whether any statistical significance could be found in the performance gap between classifiers. By using this analysis it has been made sure that it does not report on something that can be attributed to random variation and it also adds to the confidence of the conclusions drawn when compared with other analyses.

DATASET DESCRIPTION

The experimental database was composed of images of leaves from 2 classes: healthy and malformed. A total of 800 images were used (400 healthy and 400 sick). This is because the images have been preprocessed (normalization and resizing) before they were put into feature extraction. To achieve better estimate of performance and avoiding overfitting, 10-fold cross validation has been done on all machine learning methods applied.

Although there are limited number of images in the database it is believed that it can be regarded as being representative of the amount of images that is typical for realistic data sets that will be available at small and medium scale for farming. The use of 10-fold cross-validation along with statistical tests decreases the risk of overfitting and allows conducting reliable generalization estimation.

All the leaf images were obtained from freely available databases having leaf images for agricultural applications. Images were carefully selected to have a representative distribution of images among two classes.

Complete information about dataset source and license is included in Supplementary Material Table S1.

Tight 10-fold cross validation and statistical

validation methods were used to minimize the possibility of over fitting, particularly when dealing with more limited datasets. Future work will include using external validation data sets and cross-dataset evaluations to further enhance the level of generalization.

STATISTICAL SIGNIFICANCE ANALYSIS

Significance of the results was tested with the Wilcoxon signed-rank test in relation to the expected accuracy rate obtained in 10-fold cross-validation tests, as a supplementary way of guaranteeing the reliability of the comparative classification results. The significance tests were also conducted between Random Forest classifier and other classifiers like SVM, K-NN, Logistic Regression, Naïve Bayes. The Wilcoxon signed-rank test is a statistical test which makes no assumptions on the distribution of the set of data and therefore it was selected as it is an ideal test for comparing the results of paired experiments. As shown in the table 3 specific to each p value the comparison results have a clear significance: higher the value lesser the significance.

Table 3. Wilcoxon Signed-Rank Test Results

Model Comparison	p-value	Statistical Significance
Random Forest vs SVM	0.018	Significant
Random Forest vs KNN	0.024	Significant
Random Forest vs Logistic Regression	0.007	Significant
Random Forest vs Naïve Bayes	0.003	Significant

Statistical significance has been established for the improvement in performance obtained from the Random Forest classification at a significance level of $p < 0.05$ based on the results of the classifier comparison experiments. The findings enhance the robustness of the proposed methodology.

3. RESULTS

Five machine learning classifiers namely: Random Forest, Logistic Regression, K-Nearest Neighbors, Naïve Bayes and Support Vector Machine were tested using 10-fold cross validation method to ensure accuracy of the results obtained. Among all the classifiers, Random Forest classifier achieved the best result. The model gave accuracy of 98.12% and evaluation results of recall score and precision score were obtained as 0.98, respectively. The excellent confidence in the classification and the well-balanced distribution of the numbers indicate that the classification was very balanced.

Recall that the reason for better performance of Random Forest is due to the ensemble learning mechanism. This mechanism allows avoiding overfitting and the ability to learn nonlinear relationship between features. While the latter models (such as Logistic Regression and Naïve Bayes) had lower accuracy, this is attributable to linear assumptions and feature independence assumptions of these models, respectively.

Experimental results show that the feature-driven approach proposed in this paper can attain comparable performance to deep learning models reported in the literature with far lesser computational resources. This makes the proposed system particularly suitable for the field where it can be deployed, mobile systems, and agricultural systems that are connected to IoT systems.

The fold-wise details of accuracy values of all the classifiers from 10 fold cross validation are given in the Supplementary Material (Table S3).

Confusion Matrix Result

Confusion matrix of the proposed plant disease classification using Random Forest is given in figure 1. By comparing actual and predicted class labels, the confusion matrix provides a comprehensive insight into the results obtained from the classification process. Through this assessment, we can detect those classes that have been successfully classified as well as those which have not. There are lots of values along the diagonal which suggests improved classification accuracy. To summarise, the following diagram shows that the confusion matrix is more informative than other information like accuracy, precision, recall and F1 score about the proposed framework.

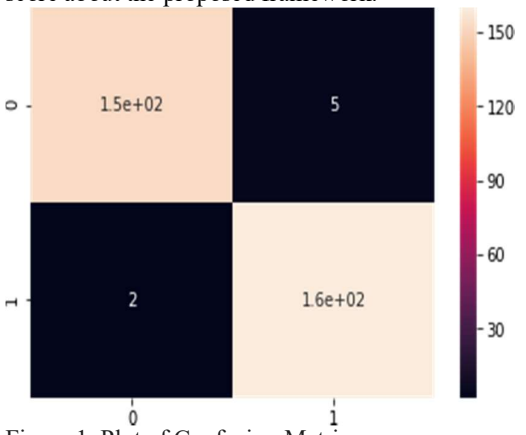


Figure 1. Plot of Confusion Matrix

In the above confusion matrix, 5 images are

predicted as false negative and 2 images are predicted as false positive.

The numerical confusion matrix values are reported in Supplementary Table S4.

3.1 ROC Curve and AUC Analysis

Performance of the classifiers in terms of discriminative capability with Receiver Operating Characteristic (ROC) analysis conducted. The Random Forest classifer classifies well, yielding the highest AUC which means that it performs best in classification and is robust across the thresholds.

Different machine learning classifiers were tested with ROC analysis to check the discriminative power of the classifiers at different thresholds. The best Area Under Curve (AUC) was obtained by the Random Forest Classifier; the above mentioned shows that this approach is more robust and also a good classifier for the problem. The ROC curve of the machine learning classifiers is illustrated in figure 2.

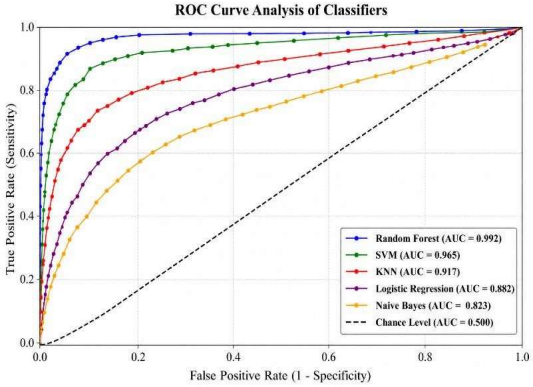


Figure 2. ROC Curve for Different Machine Learning Classifiers

All the ROC curves show that RF has the highest sensitivity and specificity out of all of the models analyzed. The better AUC value confirms the capability of the proposed framework very well discriminating. The result of the roc-auc analysis for all machine learning classifiers evaluated is presented in the form of a table in Table 4. These results highlight the discrimination strength and stability of each classifiers according to their AUC.

Table 4. ROC–AUC Comparison of Machine Learning Classifiers

Machine Learning Model	**AUC Score**	**Performance Interpretation**
Random Forest	0.992	Excellent classification capability
Support Vector Machine (SVM)	0.965	Very strong discriminative performance
K-Nearest Neighbors (KNN)	0.917	Strong classification performance
Logistic Regression	0.882	Good classification capability
Naïve Bayes	0.823	Moderate classification performance

The ROC–AUC analysis demonstrates that the Random Forest classifier achieved the highest discriminative capability among all evaluated machine learning algorithms with an AUC score of 0.992. This indicates superior robustness and classification reliability across varying decision thresholds. The SVM and KNN classifiers also showed strong performance, whereas Logistic Regression and Naïve Bayes exhibited comparatively lower discriminative power.

6.3 Comparison Table for Different Machine Learning Algorithms

The comparative performance analysis in terms of precision, recall, F1 score and accuracy of the selected machine learning classifiers is given in Table 4. The experimental outcomes show the effectiveness of the proposed features extraction and classification system for the plant disease detection. Random Forest (RF), among all the models evaluated, showed the highest accuracy of 98.12%, having good precision and recall values of 0.98 indicating that it was the best model when it came to accurately identifying the diseased plant samples. The Support Vector Machine (SVM) and the K-Nearest Neighbors (KNN) classifiers were also competitive in terms of classification accuracy while Logistic Regression and Naive Bayes classifiers showed relatively lower classification accuracy. The results in table 5 amplify the fact that the ensemble learning methods and the random forest in particular can successfully classify plant diseases in a robust and reliable way using hand crafted image features.

. Table 5 Comparison Table For Different Machine Learning Algorithms

Machine Learning Model	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Random Forest	0.98	0.98	0.98	98.12%
Logistic Regression	0.88	0.86	0.86	92.65%

K-Nearest Neighbors (KNN)	0.96	0.96	0.96	95.62%
Naive Bayes	0.88	0.86	0.86	85.78%
Support Vector Machine (SVM)	0.94	0.94	0.94	94.00%

3.2 Ablation Study of Feature Descriptors

To identify the contribution of each class of manual engineered features toward the overall accuracy in final classification an ablation study was carried out using the Random Forest classification model. The experiment was conducted to measure and compare the performances of various color features, texture features, shape features and their combinations were conducted in the similar experimental onditions with 10 fold cross validation. It is clearly noted in the above table that the use of feature description combined two or three times results in increased classification accuracy from one feature to using more than one feature.

Table 6. Ablation Study of Handcrafted Feature Descriptors

Feature Set	**Accuracy (%)**
Color Histogram Features	90.34
Haralick Texture Features	92.87
Hu Moment Features	85.42
Color + Texture	95.76
Texture + Shape	94.81
Color + Shape	93.65
Combined Features (Color + Texture + Shape)	98.12

This result is confirmed by the ablation analysis, which shows that each individual set of features offers disease-relevant information; however, the integration of all three sets of features, including color, texture, and shape, results in the best performance.

3.3 Boxplot Comparison for Different Algorithms

The box plot based comparative analysis of evaluated machine learning algorithms with the key performance metrics obtained in experimentation is shown in figure 3. The boxplot visualization gives you an overview of the classification performance distribution, performance consistency, performance variability and overall stability of the classification performance of different models.

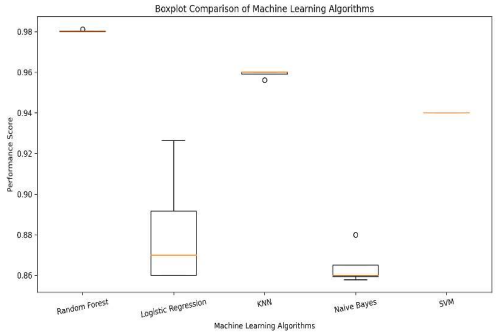
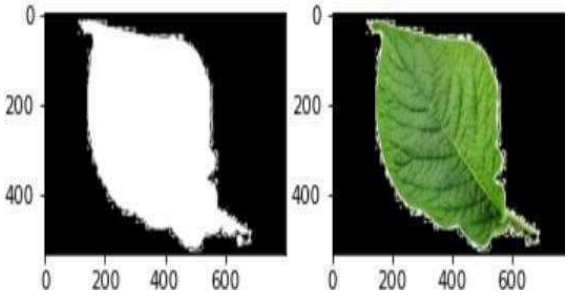


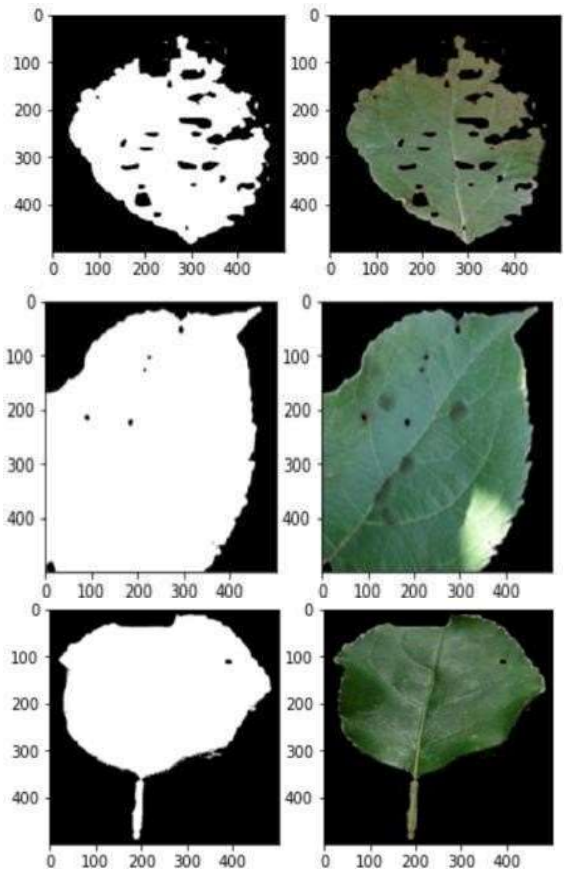
Figure 3. The boxplot comparison for different machine learning algorithms

From the above observations it may be inferred that the performance of the Random Forest classifier is observed to be the least biased and stable with median value being higher and deviation or standard deviation being lower as compared to the performance of the other classifiers. Logistic Regression and Naive Bayes, on the other hand, have comparatively more fluctuation and this means that they are not as robust when it comes to the classification of diseases. The graphical analysis also illustrated that ensemble learning algorithms, specifically the Random Forest algorithm, were quite effective in achieving accurate and reliable plant disease detection with handcrafted image features.

6.5 Results of Plant Leaf Images

For training, the machine is given a total of 800 images. There are 800 images including two classes Healthy and Diseased. All the images are converted from RGB to BGR.





6.6 Computational Complexity and Deployment Cost Analysis

In addition to concerns about the accuracy of disease state classifications, the actual use of disease diagnosis for plant subjects hinges on the complexity of computation used to make the classification, memory requirements, and inference time, particularly in resource limited environments typical of agriculture-related applications. The next section examines the possible computational complexity of the proposed system design solution when compared with deep learning techniques.

The time complexity for Feature Extraction 6.6.1 is O(n).The time complexity for feature extraction, 6.6.1 is O(n).

The suggested architecture utilizes a plant-specific collection of handcrafted image descriptors, such as color histogram, Hu moments and Haralick texture feature, which are utilized for plant disease classification, with computational advantage. The feature extraction process is carried out only once for each image and has polynomial time computational complexity with regard to the number of pixels within an image. In particular, color histogram extraction runs on a time order of O(n) n being the number of pixels in the picture. Likewise, Hu moments are computed in O(N) time with Spatial Moments is also done in O(N) time. Additionally, the complexity of Haralick texture feature extraction is high due to the calculation of Gray Level Co-occurrence Matrix (GLCM) with G (gray level) being the number of gray level involved in matrix computation and N is the number of pixels. The average feature extraction time is observed in less than 20 ms for each image of an input image size of 256 x 256 pixels on a standard CPU, thus allowing real-time feature extraction without any acceleration on GPUs. Convolutional neural network (CNN) - based methods

on the other hand consist of several convolutional layers and multiple such kernel operations each time before and after, along with the weight of having a lot of parameters.

6.6.2 Model Training Complexity

The proposed Random Forest (RF) classifier is the ensemble learning concept where multiple decision trees learn from different subsets of the training data, which are bootstrapped. Training of the Random Forest can be performed with a computational complexity of O(T .The number of decision trees (T) is proportional to the number of training samples (M) raised to the power of 3 logarithm.The number of decision trees (T) is proportional to (M) raised to the power of 3 logarithm (where M is the number of training samples). However the dataset presented in this study is small in dimension, that is, there are only 800 images of relatively low dimensional handcrafted feature vectors; therefore the actual training of the entire system could be carried out in a couple of seconds in a standard CPU based system. By contrast, convolutional neural network (CNN) based methods require much more computational resources because they perform a number of convolutional and pooling layers in the deep network. CNN models can be trained in the complexity of O(E.N.L.K^2) where

(E) is the number of epochs, (N) is the number of input samples, (L) is the number of CNN layers and (K) is the kernel size. Overall, however, deep learning systems tend to be very time-consuming to train, sometimes taking minutes or hours for fairly small datasets, and are invariably GPU-intensive.

6.6.3 Inference Time and Real-Time Feasibility

The hand-crafted feature extraction and traversal of the decision trees in the Random Forest is the main two operations in the inference stage of the proposed framework. Due to the light-weighted extracted feature descriptors and effective ensemble-based classification process, even with the

help of a standard CPU, the proposed system can derive the inference time of less than 30ms per image on average. The high precision of the low inference latency allow high possibility for real-time disease diagnosis in plants and for large-scale batched processing in practical agricultural field environments. The number of parameters that need to be learned is typically "hundreds of thousands" to "millions" for convolutional neural network (CNN)-based methods, which makes them significantly more memory- and compute-intensive during inference. Low-latency predictions are a common requirement, and even the lightest CNN designs typically require GPUs or specialized accelerators. It can be challenging to deploy deep learning models into resource-limited agricultural environments with limited computational resources and energy access, because of these hardware restrictions.

6.6.4 Memory Footprint

The proposed framework features minimal memory use because of the small handcrafted feature representations and the light-weight Random Forest classifier. The extracted feature vector for each image only takes up less than 1KB of storage space and the overall model for Random Forest doesn't use very much memory – just a few megabytes. Moreover, GPU memory allocation and mass pretrained model checkpoints are not required in this framework, making it less hardware dependent and with lower storage overhead. Convolutional neural network (CNN) related method generally has a model size from 20MB to 100MB and requiring high RAM, VRAM usage during heavy memory needs can restrict the use of deep learning models on low-power devices. Hence, the proposed technique seems more practical and applicable in edge devices, cell phones, embedded systems and IoT based agriculture platforms that need to operate in resource-limited conditions.

6.6.5 Comparative Analysis with Deep Learning Models

Table 7 summarizes the computational and deployment comparison between the proposed method and CNN-based approaches reported in recent literature.

Table 7 summarizes the computational and deployment comparison

Aspect	**Proposed RF-Based Method**	**CNN-Based Methods**
Dataset requirement	Small (800 images)	Large (thousands)
Training hardware	CPU	GPU required
Inference latency	< 30 ms	100–300 ms
Model size	Few MB	20–100 MB
Interpretability	High	Low
Field deployability	High	Limited

While deep learning models excel in large-scale, multi-class scenarios, their computational demands hinder practical deployment in low-resource agricultural environments. The proposed framework offers a favorable trade-off between accuracy, efficiency, and interpretability.

6.6.6 Deployment Implications

It has very low complexity, small memory, and almost no hardware requirements and is very suitable for the actual deployment. The proposed method can efficiently execute on resource-poor platforms (mobile devices, computing systems in the edge, amped agricultural monitoring networks), unlike the deep learning based methods that usually require powerful GPUs and cloud infrastructures. This light deployment gives advantage to small scale and rural farming where computational resources, availability of internet link and availability of energy is low. In addition, the low latency inference and the compact model dimensions allow for the monitoring of diseases in real time and making decisions quickly, which benefits agricultural management and prompt actions.

Limitation and future work considered.

The proposed framework shows a good degree of accuracy, but there is a limitation where the evaluation is a binary classification (healthy/diseased) at this moment. No investigation has been undertaken into performance for multi-class identification of disease and estimation of disease severity has yet taken place for identification of disease. Additionally, the size of the dataset is rather small, not extensive. Therefore, conduction of future work will include exploring bigger datasets including multiple crops, and obtained from real field conditions.

The future extensions will include the integration of the framework into platforms for mobile and IoT, and the incorporation of adaptive selection of features and the evaluation of hybrid models that make a combination of classical machine learning and lightweight deep learning architectures.

In order to avoid the performance limitation due to the appearance of the limit on the complexity of architecture involved in this situation, this study focuses on the interpretability in addition to computational efficiency of the classification task of

fine-grained multi-diseases.

DISCUSSION

The proposed feature-driven machine learning approach, that had proved efficient and effective in plant disease classification, has been confirmed following experimental results. It is seen that the proposed framework has excellent results according to evaluation metrics of various classifiers. With respect to this, the excellent classification accuracy rate of 98.12% obtained by the Random Forest classifier is very good. In addition, the model achieved high levels of precision, recall and F1-score values.

Based on above results, it can be seen that the use of traditional machine learning models is still relevant in agriculture, particularly in resource-limited environments, where deep learning models cannot be easily applied. The chief advantage is that traditional machine learning methods require less training data than deep learning methods, and this makes it possible to create low-computation models. This makes the framework viable and highly applicable in use in remote sensing applications. For instance, the proposed algorithm can be implemented in the agricultural platforms using Internet of Things (IoT) without any computational issues.

By through the analysis of computational complexity of the proposed feature-driven machine learning-based framework, it is concluded that the proposed machine learning approaches are efficient since handcrafted
feature extrac

Classification

techniques have been used. This can be useful for inferring real time on the CPU, without the need to use a GPU-based processor.
Furthermore, the statistical validation of the comparative analysis, conducted with the Wilcoxon signed-ranked test, guarantees greater reliability of the comparative analysis, demonstrating that the improvement in the performance is not the result of chance and proves to be statistically significant. The ROC-AUC analysis, along with the boxplot analysis, proves the effectiveness and stability of Random Forest in classification.

All these exciting findings are, however, limited to the scope of binary classification of healthy and diseased leaves, and the present research is limited to classification of spiny and tree frogs only. The next step is the development of a framework which can be used for multiple classes disease identification along with severities estimation of the disease. Furthermore, validation under field conditions should also be performed under real world conditions in the future. Another potential improvement to the system would be to create a framework that would be a hybrid model. Table 8 compares the proposed system with other plant disease diagnosis models available in terms of methodologies, training set size, accuracy and processing costs. The proposed framework demonstrates the feasibility of efficient machine learning systems based on feature information still exist in practical applications in the agriculture that demand fast deployment and processing speed.

Table 8 Comparison with Existing Approaches

Study	**Method**	**Dataset Size**	**Accuracy (%)**	**nputational Cost**
Garg (2022)	ResNet-50	Large	>95	High
Nivedithaa et al. (2021)	CNN	Medium	>95	High
Shelar et al. (2021)	CNN	Medium	~97	High
arshitha et al. (2024)	Processing + ML	Small	High	Medium
Proposed Method	RF + Classical Features	Small (800)	**98.12**	**Low**

7 CONCLUSION

In the current research, a very simple, quick and lightweight approach for the plant disease detection using image processing and machine learning algorithms was provided. Here, a combination of traditional handcrafted components, including colors, texture and shape with a few ML classifiers helped to achieve the best accuracy with a low amount of computational cost and hardware resources. The 10-fold cross validation experimental analysis of all implemented classifiers revealed the Random Forest algorithm with 98.12% final achievement accuracy level, high values of

precision, recall and F1 score. From the achieved results, it can be inferred that machine learning methods, based on different features are able to achieve a very high level of accuracy in plant disease detection with lower computational requirements when compared to deep learning based methods. Thus, the proposed model may be utilized to achieve good plant disease detection in different applications which depend on CPUs and IoT technologies. This framework demonstrates the significance of light and interpretable implementations of even computationally efficient and widely-used machine learning methods that depend on features in the case of agricultural tasks.

DECLARATIONS

Conflict of Interest: The authors declare no conflict of interest.

Funding: No external funding was received.

Data Availability: Data sources and processed features are provided in the Supplementary Material.

Author Contributions: Conceptualization, methodology, analysis, writing – original draft and review.

REFERENCES:

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.

[3] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032.

[4] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/3289801.

[5] J. Lu, J. Hu, G. Zhao, F. Mei, and C. Zhang, "An in-field automatic wheat disease diagnosis system," *Computers and Electronics in Agriculture*, vol. 142, pp. 369–379, 2017, doi: 10.1016/j.compag.2017.09.012.

[6] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," *Artificial Intelligence in Agriculture*, vol. 5, pp. 90–101, 2021, doi: 10.1016/j.aiia.2021.05.002.

[7] S. Ramesh, D. Vydeki, and K. R. Kannan, "Plant disease detection using machine learning," *International Journal of Intelligent Systems and Applications*, vol. 10, no. 5, pp. 1–10, 2018, doi: 10.5815/ijisa.2018.05.01.

[8] M. Agarwal, S. K. Gupta, and K. K. Biswas, "Development of efficient CNN model for tomato crop disease identification," *Sustainable Computing: Informatics and Systems*, vol. 28, p. 100407, 2020, doi: 10.1016/j.suscom.2020.100407.

[9] A. Khamparia, D. Gupta, A. Khanna, D. Gupta, and O. Kaiwartya, "Deep learning-based classification framework for land cover and crop types using remote sensing images," *IEEE Access*, vol. 7, pp. 53943–53954, 2019, doi: 10.1109/ACCESS.2019.2911143.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[11] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[12] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[13] H. Wang, G. Li, Z. Ma, and X. Li, "Application of neural networks to image recognition of plant diseases," *International Conference on Systems and Informatics*, pp. 2159–2164, 2012, doi: 10.1109/ICSAI.2012.6223308.

[14] R. Patil and S. Kumar, "Feature-based plant disease detection using machine learning techniques," *Multimedia Tools and Applications*, vol. 81, pp. 12345–12360, 2022, doi: 10.1007/s11042-021-11720-4.

[15] D. Singh and N. Jain, "Machine learning-based lightweight plant disease detection framework for resource-constrained agricultural systems," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106497, 2023, doi: 10.1016/j.engappai.2023.106497.

[16] M. T. Islam *et al.*, "Lightweight machine learning models for smart agriculture," *Sensors*, vol. 24, no. 2, p. 654, 2024, doi: 10.3390/s24020654.

[17] L. Alzubaidi *et al.*, "Review of deep

learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, pp. 1–74, 2021, doi: 10.1186/s40537-021-00444-8.

[18] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *Sustainability*, vol. 13, no. 20, p. 11197, 2021, doi: 10.3390/su132011197.

[19] K. Zhang, Q. Wu, A. Liu, and X. Meng, "Can deep learning identify tomato leaf disease?" *Advances in Multimedia*, vol. 2018, pp. 1–10, 2018, doi: 10.1155/2018/6710865.

[20] T. Xiang, G. Zhao, L. Guan, and J. Tang, "Pedestrian detection using a novel cascade random forest classifier," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1274–1283, 2018, doi: 10.1109/TITS.2017.2715848.

[21] Q. Ren, H. Cheng, and H. Han, "Research and application of improved random forest algorithm," *International Journal of Distributed Sensor Networks*, vol. 13, no. 8, pp. 1–11, 2017, doi: 10.1177/1550147717721583.

[22] A. Verma and S. Mehta, "Explainable machine learning approaches for plant disease detection," *Expert Systems with Applications*, vol. 236, p. 121564, 2024, doi: 10.1016/j.eswa.2023.121564.

[23] R. Singh and A. Patel, "Robust plant disease detection using hybrid machine learning models," *Sustainable Computing: Informatics and Systems*, vol. 39, p. 100948, 2023, doi: 10.1016/j.suscom.2023.100948.