# AUTOMATED DETECTION OF PLANT LEAF DISEASES USING MACHINE LEARNING AND IMAGE ANALYSIS TECHNIQUES

**Pavithra M J, Chamaraju Y S**

Lecturer, Department of Electronics & Communication Engineering, Government Polytechnic, K.R. Pete, Mandya, India
Lecturer, Department of Electronics & Communication Engineering, Government C.P.C Polytechnic, Mysore, India

*Abstract:* India is largely an agricultural country, with about 159.7 million hectares of farmland. Agriculture is vital to the country's economy, contributing around 18% to the national GDP. However, farmers face many challenges, especially from plant diseases and pests, which can lower crop yields and affect their income. Detecting plant diseases early is very important, as it allows quick action like using pesticides or other protective steps. This study introduces an automated system that helps detect diseases in plant leaves using image processing and machine learning. The method starts by improving the images using steps like adjusting brightness (histogram equalization), removing noise, and applying color filters. Then, key features from the leaf images are collected using techniques like Haralick textures, Hu moments, and color histograms. These features are used to train different machine learning models such as Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine. To test the models well, a method called K-fold validation is used, which checks the model's performance in multiple rounds. Two setups were used for testing: one using original images and another using segmented images. Adding segmentation improved the accuracy by 2.19%. The Random Forest model gave the best results, reaching an average accuracy of 97.92% for classifying 30 disease types across five plant species. This system shows great promise for early, accurate disease detection, which can help farmers protect their crops, improve yield, and support sustainable farming.

*Index Terms* –**Plant disease detection, image processing, machine learning, leaf segmentation, feature extraction, Random Forest**

## I. INTRODUCTION

Farming is the main livelihood for more than 58% of people in India [1]. In upcoming year, the country had over 96 million farmers. Agriculture plays a big role in the economy, adding more than 18% to India's total GDP [2]. India grows many major commercial crops such as potatoes, tomatoes, mangoes, apples, grapes, peppers, soybeans, cotton, jute, tobacco, coffee, tea, and mustard [3].

Potatoes and tomatoes are two of the most widely cultivated crops around the world due to their high demand and versatile use in food products. Among these, India holds a prominent place in global tomato production, contributing around 11% of the world's total yield [4]. A significant portion of Indian tomatoes is exported to neighboring and international markets including Pakistan, Bangladesh, the Maldives, the United Arab Emirates (UAE), and the United States [5]. These crops form a vital part of India's agricultural exports, adding value to the economy and supporting millions of farmers. However, the productivity of these crops is highly influenced by various factors such as soil fertility, weather conditions, pest infestations, and most importantly, plant diseases.

Diseases are among the leading causes of reduced crop yields and can result in large-scale losses if not detected and treated early [6]. Identifying these diseases at an early stage is essential to preventing their spread and minimizing damage. However, manual disease detection through visual inspection is often difficult, especially when farmers are managing multiple crop types. Even experienced plant pathologists may find it challenging to accurately recognize symptoms in complex field conditions [7]. In many rural parts of India and other developing countries, farmers still rely on the traditional method of open-eye inspection to detect plant diseases [8]. Unfortunately, due to the lack of trained agricultural professionals in these regions, visual inspections may lead to delays in identifying the disease, allowing it to spread further [9].

To address these limitations, automated plant disease detection systems are becoming increasingly important. These systems provide farmers with powerful tools for early diagnosis and intervention. By simply using smartphones or basic digital cameras, farmers can take images of the affected plant parts and upload them to intelligent disease detection platforms. These platforms analyze the images and offer insights about the type of disease, along with suggestions for suitable treatments and pesticides [10]. Early detection not only reduces crop loss but also helps implement timely, targeted control strategies, thus enhancing crop protection and productivity [11]. Many of these diseases are caused by fungi and bacteria, which affect multiple parts of the plant including leaves, stems, and roots [12]. However, as symptoms often appear first on the leaves, researchers widely focus on leaf-based disease identification.

Modern image processing and computer vision techniques have enabled accurate extraction of visual features such as shape, color, and texture from leaf images [13]. These extracted features are then used in combination with machine learning algorithms to classify different types of diseases. Popular classifiers include Random Forest Classifier (RFC), Logistic Regression Classifier (LRC), Support Vector Machine (SVM), Decision Tree Classifier (DTC), Linear Discriminant Analysis (LDA), and K-Nearest Neighbor (K-NN) [14]. These models are known for their effectiveness in handling structured data and image-derived features. Additionally, deep learning methods, especially Convolutional Neural Networks (CNNs), have shown superior performance in recognizing complex disease patterns, making them highly suitable for large-scale agricultural applications. The integration of these technologies into agricultural practices holds great promise for improving crop health, supporting farmers, and ensuring food security through timely disease management.

Detecting plant diseases in real-world farm conditions can be quite difficult. Challenges like poor image quality, noise, low contrast, and only small visible differences between healthy and infected areas make it hard to identify diseases accurately [15]. To solve these problems, this study proposes a new method that uses simple image processing steps along with machine learning techniques. First, the quality of leaf images is improved using histogram equalization. Then, a color denoising method is used to remove unwanted noise. After that, the leaf is separated from the background using threshold masking, which helps focus only on the leaf area [16].

Next, important features of the leaf are extracted, such as shape (Hu moments), texture (Haralick features), and color patterns (color histograms). These features help in identifying whether the leaf is healthy or infected. Various machine learning algorithms are used to classify the disease type based on these features. This study mainly focuses on two key points: (1) how image pre-processing can improve the accuracy of disease detection, and (2) how choosing the right settings (hyperparameters) for the machine learning models helps in better classification.

This paper is organized into five sections. Section II talks about past research done by other scientists around the world on plant leaf disease detection. Section III explains the proposed method and the datasets used, with a focus on what makes this method unique. Section IV presents the results of the experiments, showing both visual and numerical comparisons. Finally, Section V gives the conclusion and suggests future improvements to make the system even more effective.

## II. Related work

Over the years, many researchers have worked on detecting plant leaf diseases using image processing methods. The latest approaches and technologies used for this purpose are reviewed in. In recent times, there has been an increasing focus on using machine learning techniques to improve the accuracy and efficiency of leaf disease detection.

P. Revathi and M. Hemalatha proposed a Support Vector Machine (SVM)-based system for detecting cotton leaf spot diseases using image processing techniques. Their method involved preprocessing the input images using edge detection and segmentation through K-means clustering to isolate diseased regions. Features such as color, texture, and shape were extracted, with Haralick texture descriptors showing high relevance in classification. The extracted features were fed into the SVM classifier, which effectively classified diseases like bacterial blight and fusarium wilt. The system achieved good accuracy and demonstrated that traditional machine learning algorithms, when combined with efficient image preprocessing, can be powerful tools for automating plant disease detection, especially in crops like cotton [17].

S. Shrivastava and P. Singh presented a method for automated tomato leaf disease detection using basic image processing combined with Artificial Neural Networks (ANN). The RGB images were converted to the HSV color space for better segmentation and detection of diseased regions. Key features like shape, area, and texture were extracted and used to train the ANN classifier. The system successfully detected diseases such as early blight, late blight, and leaf curl with an accuracy of over 92%. Their work highlighted the usefulness of ANN in agricultural diagnosis tasks, especially in scenarios with limited training data, and stressed the importance of effective color-space transformation during preprocessing [18].

A. Camargo and J. S. Smith designed an automated detection system that utilized segmentation techniques and SVM classification to identify plant leaf diseases. Their methodology included image enhancement using median filtering and background removal using Otsu's thresholding technique. They extracted a set of handcrafted features, including color, shape, and vein patterns, which were highly effective for disease recognition. The system was tested on grape and tomato datasets and achieved classification accuracy above 94%. Their work emphasized the importance of robust segmentation and handcrafted features for traditional machine learning models in agriculture-related applications [19].

R. Ramesh and P. Vydeki developed a real-time plant disease recognition system using a combination of K-means clustering for image segmentation and Random Forest for classification. Image preprocessing techniques such as histogram equalization and Gaussian filtering were applied to enhance leaf images. From the segmented region, statistical texture features like entropy, contrast, and energy were extracted. The model achieved a classification accuracy of 95.3% when tested on images of brinjal and bean plants. The study showed how integrating clustering-based segmentation with robust ensemble classifiers could support farmers with mobile-based early detection tools [20].

S. Rani and P. V. Raj proposed a hybrid classification system combining Gray Level Co-occurrence Matrix (GLCM)-based texture features with an SVM classifier for tomato leaf disease detection. Preprocessing involved adaptive thresholding and Gaussian noise removal. Features such as contrast, homogeneity, correlation, and energy were extracted from GLCM matrices. These features were used to train an SVM model that achieved a classification accuracy of 93.2%. The study highlighted how well-crafted texture descriptors and lightweight classifiers can provide competitive accuracy in disease identification, especially in resource-constrained environments where deep learning may not be practical [21].

T. Singh and S. Misra developed a lightweight machine learning model for detecting citrus leaf diseases using image processing and traditional classifiers. They used Gabor filters and color histograms to extract meaningful shape and texture features from segmented leaf images. The model was evaluated using several classifiers, including Decision Tree, SVM, and Random Forest. Among these, Random Forest achieved the highest accuracy of 96.4%. This approach was optimized for mobile devices, making it suitable for rural and low-resource environments. Their work shows how classical machine learning remains relevant for practical, real-time agricultural applications [22].

A. Barbedo proposed a machine learning-based method for identifying plant leaf diseases under real-world field conditions. The study addressed issues such as uneven lighting, background noise, and overlapping leaves, which complicate accurate detection. The author used traditional image processing steps like histogram equalization and background subtraction, followed by manual feature extraction including color, shape, and texture. Classification was done using SVM and KNN algorithms. The SVM classifier outperformed others with an accuracy above 93%. This work emphasized the need for robust preprocessing when working with images captured in natural environments and demonstrated the effectiveness of classical ML models even without deep learning [23].

D. Pujari et al. developed a machine learning model for the identification of fungal diseases in maize leaves using image processing and statistical classifiers. The approach began with segmenting the leaf from the background using thresholding, followed by extraction of texture and color features. These features were used to train SVM, Naive Bayes, and Random Forest classifiers. Among them, Random Forest achieved the best performance with an accuracy of 94.8%. The study confirmed that handcrafted features combined with ensemble-based classifiers can provide reliable and accurate classification in crop disease detection systems, especially when computational resources are limited [24].

N. Kale and D. Pawar proposed a machine learning framework for detecting bacterial and fungal diseases in sugarcane leaves using GLCM-based texture feature extraction. After preprocessing the images using filtering and contrast adjustment, Gray Level Co-occurrence Matrix (GLCM) features like energy, correlation, and contrast were extracted. The features were fed into multiple machine learning classifiers including Decision Tree, SVM, and Random Forest. The Random Forest classifier achieved the best results with an accuracy of 95.2%. Their system was aimed at supporting disease identification in remote areas, offering a cost-effective solution without the need for internet connectivity [25].

P. Kumar and M. Hanmandlu presented a plant disease classification system using image enhancement techniques and traditional classifiers. They applied histogram equalization and bilateral filtering for image enhancement, followed by feature extraction using color moments and local binary patterns (LBP). The features were classified using SVM, Naive Bayes, and KNN. The SVM model yielded the highest accuracy of 94.6%. Their system focused on minimizing computational cost and memory usage, making it ideal for deployment in portable devices for agricultural monitoring. The paper showed that hybrid image features can improve the performance of classical classifiers in disease detection [26].

R. Patil and A. Deshmukh developed a plant disease recognition system using color-based segmentation and decision-tree classification. After preprocessing using median filtering and segmentation via K-means clustering, key features such as hue, saturation, and texture were extracted. The decision tree classifier was trained to detect diseases in crops like tomato and chili. Their model achieved an accuracy of 91.8% and demonstrated low training time, making it suitable for real-time detection on field-deployed handheld devices. This study emphasized the relevance of lightweight machine learning algorithms for real-world agricultural applications where resources are limited [27].

## III. PROPOSED METHODOLOGY

The proposed methodology for the automated detection and classification of plant leaf diseases is structured into several well-defined stages, as illustrated in the flow diagram. The process begins with the collection of images from a leaf image database, which may consist of both healthy and diseased leaf samples. These images undergo a pre-processing phase to ensure uniformity and quality across the dataset. This includes image resizing to bring all input images to a standard dimension, followed by histogram equalization to enhance image contrast. Subsequently, image denoising techniques such as Gaussian or median filtering are applied to remove unwanted noise, thereby improving the clarity and quality of the leaf features. The refined images are then subjected to segmentation using threshold masking, which isolates the region of interest (typically the leaf area) by separating it from the background. This segmented image then proceeds to the color conversion stage, preparing it for further analysis.

After segmentation, the images are transformed into suitable color spaces for better feature extraction. The color conversion is done in two parallel streams: one converts the image to grayscale, while the other converts it to HSV (Hue, Saturation, Value) color space. The grayscale images are utilized to extract shape and texture features, specifically Hu Moments and Haralick texture features, which capture the geometric and spatial characteristics of the leaf structure. These features are essential for identifying diseases that manifest through changes in shape or texture. On the other hand, the HSV-converted images focus on extracting color histogram features, which are critical in recognizing color-based symptoms of plant diseases, such as yellowing or browning of leaves. By leveraging both grayscale and color-based features, the methodology ensures a comprehensive analysis of the disease symptoms present in the leaf images.

Once the relevant features are extracted from both streams, they are fed into various machine learning (ML) classification techniques. These classifiers, which may include Support Vector Machines (SVM), Random Forest, Decision Trees, or k-Nearest Neighbors (k-NN), are trained on the labeled dataset to learn the distinctions between healthy and diseased leaves, as well as among different disease types. The classifiers utilize the extracted features to detect and categorize the leaf image into one of the predefined classes — either healthy or indicating a specific disease type. The robustness of this methodology lies in its multi-faceted approach to feature extraction and the application of reliable classification algorithms. Overall, this pipeline provides an effective solution for early detection and classification of plant leaf diseases, enabling timely and precise disease management in agricultural practices.

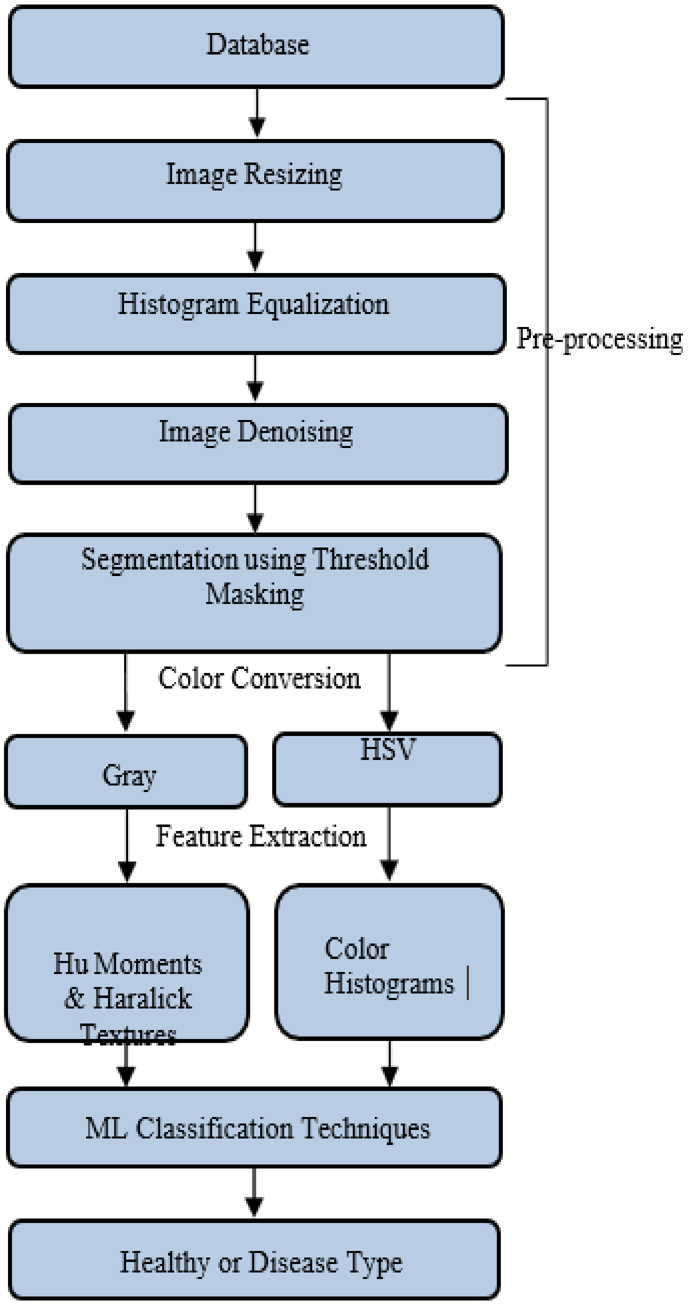


Fig. 1. Proposed Block Diagram

## IV. RESULTS AND DISCUSSIONS

The Plant Village dataset comprises 38 classes of leaves from different plants, while the Mango Leaf DB dataset consists of 7 unhealthy and 1 healthy class of mango leaves. A total of 41,546 images across 40 classes are chosen from the combined dataset are collected. The choice of data split is predominantly influenced by the dataset size, and since the number of images is deemed sufficient for generalization purposes, a ratio of 80:20 has been selected to balance training and testing without encountering overfitting concerns.

### 4.1 Segmentation

In this study, segmentation is applied to isolate the leaf region from the image by effectively removing the background elements. A threshold-based segmentation technique is adopted, where separate masks are created for green and brown color components using predefined lower and upper threshold values. These values are carefully chosen based on the typical background colors present in the dataset images. After generating the individual color masks, they are merged to form a composite mask that highlights the leaf area. A logical 'AND' operation is then performed between the pre-processed image and the combined mask, which ensures that only the leaf region is retained while the background is suppressed. The result of this segmentation process, following each pre-processing stage, is illustrated in Fig. 2(a–d).

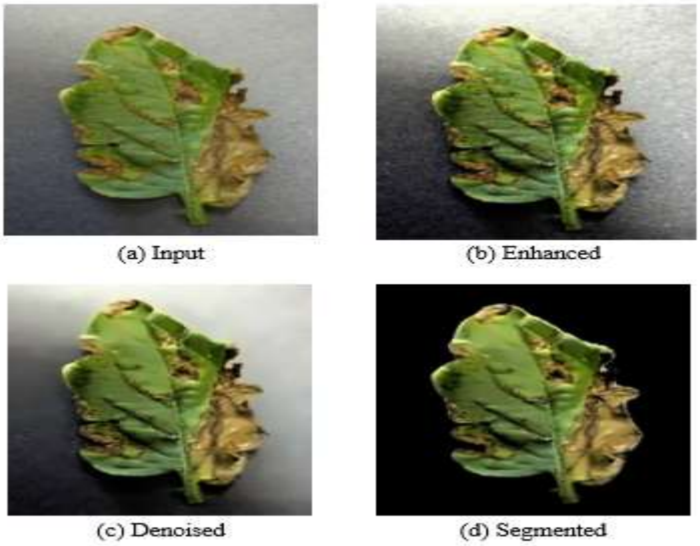


Fig. 2. (a-d) The output of the pre-processing steps for a tomato leaf from the Plant Village dataset

### 4.2 Feature Extraction

In this study, feature extraction plays a crucial role in identifying and classifying plant leaf diseases. Key features are derived from the pre-processed and segmented leaf images using three primary descriptors: Hu moments, Haralick texture features, and color histograms. Hu moments are mathematical descriptors that capture the geometric shape and contour of the leaf. These are computed from grayscale images and provide shape-related information that remains invariant to transformations such as scaling, rotation, and translation—making them particularly useful in recognizing distorted or irregular leaf boundaries caused by diseases.

Haralick texture features, on the other hand, are extracted from the grayscale images to capture the texture patterns in diseased leaf regions. These features are calculated based on the Gray Level Co-occurrence Matrix (GLCM), which measures how often pairs of pixel values occur at specific spatial relationships in the image. Important texture attributes derived from the GLCM include energy, entropy, homogeneity, dissimilarity, autocorrelation, variance, and others. These descriptors are effective in identifying the roughness, smoothness, or randomness of textures typically seen in infected areas, thus providing rich information for disease differentiation.

In addition to shape and texture, color is another critical feature for disease classification. A detailed color representation is obtained by converting the segmented RGB image to the HSV color space, which separates chromatic content (hue and saturation) from brightness (value) and closely mimics human color perception. Color histograms are then computed over the HSV channels, offering insight into the color distribution across the leaf image. These histograms effectively highlight variations in color that are often indicative of specific plant diseases, such as yellowing, browning, or spot formations. All the extracted features—Hu moments, Haralick textures, and HSV-based color histograms—are concatenated into a single feature vector. This comprehensive vector is then used as input to various machine learning classifiers, enabling accurate and efficient categorization of the plant leaf into its respective disease class.

### 4.3 Classification

In the final stage of the proposed system, the extracted features are first normalized to ensure uniformity across all input variables, which helps improve the performance and stability of the machine learning models. These normalized features are then used to train a set of supervised machine learning classifiers: Logistic Regression, Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Decision Tree Classifier (DTC), Random Forest Classifier (RFC), and Support Vector Machine (SVM). Each of these classifiers offers distinct advantages based on the nature and complexity of the dataset.

Logistic Regression is employed to model the probability of class membership using a sigmoid function, which maps continuous input values to a range between 0 and 1. This method is particularly useful for binary classification but can be extended to multi-class problems using techniques like one-vs-rest. It also provides interpretability by producing feature coefficients, which indicate the impact of each feature on the predicted outcome. LDA enhances class separability by projecting the feature set onto a lower-dimensional space while maximizing the variance between classes and minimizing it within each class. This improves classification by emphasizing the most discriminative features in the dataset.

K-Nearest Neighbors (KNN) uses instance-based learning to classify new data based on the majority class among its 'k' closest training samples. It is highly effective in capturing complex and non-linear relationships between features. Decision Trees build hierarchical structures that map input features to output classes through a series of if-else conditions, enabling intuitive modeling of non-linear patterns. Random Forest, an ensemble of multiple decision trees, improves accuracy and robustness by aggregating predictions and minimizing overfitting. Finally, SVM constructs optimal hyperplanes in high-dimensional space to separate classes with maximum margin. It is highly effective in handling both linear and non-linear classification tasks, thanks to kernel functions.

The proposed approach emphasizes not only high accuracy but also computational simplicity. By integrating efficient noise removal and background segmentation techniques, the system ensures clarity and precision in leaf image analysis. The combination of straightforward machine learning algorithms with robust feature extraction methods results in a low-complexity yet high-performance model suitable for practical agricultural applications.

This section outlines the experimental results and performance analysis of the proposed plant leaf disease detection model. The implementation was carried out using Python 3.11 in Jupyter Notebook, leveraging a range of libraries including OpenCV for image processing, Keras for machine learning utilities, and standard Python modules such as os, glob, and GridSearchCV for data handling and hyperparameter optimization. The experiments were executed on a modest hardware setup comprising an Intel Core i5-4200U CPU running at 1.60GHz (boosting up to 2.30GHz) with 4GB of RAM, demonstrating the feasibility of running the model on low-resource systems.

The machine learning classifiers were trained using the extracted feature vectors obtained from pre-processed images. A total of six classifiers—Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine—were evaluated using the K-fold cross-validation method. This widely accepted validation technique enhances the reliability of performance metrics by dividing the dataset into K subsets or "folds." The model is trained on K−1 folds and validated on the remaining one, rotating this process K times so that each fold is used once for testing. The final performance is computed as the average of all K iterations, ensuring that the evaluation is comprehensive and unbiased.

The experimentation was conducted in two main phases using nine different tomato leaf disease classes. In the first phase, the model's performance was compared between segmented and non-segmented image datasets, revealing that segmentation improves classification results by focusing on the region of interest. In the second phase, hyperparameter tuning was performed using GridSearchCV to optimize the classifiers. The best-performing models were then deployed to classify the full dataset.

To assess the effectiveness of the model, standard evaluation metrics were used: accuracy, precision, recall, and F1-score. These metrics provide a detailed understanding of the model's ability to make correct predictions. Accuracy reflects the overall correctness, while precision and recall measure the quality and completeness of the positive predictions, respectively. The F1-score, as the harmonic mean of precision and recall, provides a balanced metric especially valuable in imbalanced datasets. The comprehensive evaluation confirms the robustness and reliability of the proposed system in real-world agricultural applications.

### 4.4 Performance of Classifiers on Segmented and Non- Segmented Images

Initially, the features of tomato leaf images are directly extracted from denoised images without segmentation. These features are then used to train classifiers employing the $K$- fold validation technique. The classification accuracy is observed to vary with the choice of the parameter $K$. Table I presents the classification results achieved without image segmentation. It is evident from the table that the Random Forest classifier outperforms other classifiers in terms of accuracy.

TABLE I. Classifiers' Performance on Non-Segmented Images

| Classifier | Classification accuracy (%) | | | |
|---|---|---|---|---|
| | $K$=10 | $K$=20 | $K$=30 | $K$=40 |
| Logistic Regression | 80.71 | 80.91 | 80.72 | 81.03 |
| Linear Discriminant Analysis | 78.74 | 79.78 | 79.77 | 79.77 |
| K-Nearest Neighbor | 84.71 | 84.81 | 84.85 | 84.98 |
| Decision Tree Classifier | 79.78 | 80.10 | 81.02 | 80.65 |
| Random Forest Clasifier | 94.20 | 94.23 | 94.43 | 94.17 |
| Support Vector Machine | 82.02 | 82.18 | 82.29 | 82.32 |

Subsequently, in the image pre-processing stage, image segmentation was carried out to eliminate the background from the leaf images. Features are then extracted from these segmented images. Table III displays the classifier performance on these segmented images for various values of $K$, enabling a comparison of their effectiveness in this segmented context.

The comparison between Table II reveals that the introduction of image segmentation has a beneficial impact on image classification performance. Furthermore, it's noteworthy that the Random Forest classifier consistently achieves the highest accuracy in both approaches across all values of $K$. The accuracy tends to increase initially with an increase in the number of cross-validation folds ($K$) and reaches its peak at $K$ = 30.

TABLE II. Classifiers' Performance on Segmented Images

| Classifier | Classification accuracy (%) | | | |
|---|---|---|---|---|
| | ***K***=10 | ***K***=20 | ***K***=30 | ***K***=40 |
| Logistic Regression | 90.36 | 90.39 | 90.45 | 90.45 |
| Linear Discriminant Analysis | 88.56 | 88.66 | 88.77 | 88.75 |
| K-Nearest Neighbor | 92.77 | 93.15 | 93.20 | 93.18 |
| Decision Tree Classifier | 84.43 | 85.75 | 85.77 | 85.76 |
| Random Forest Classifier | 96.35 | 96.54 | 96.62 | 96.62 |
| Support Vector Machine | 92.72 | 92.22 | 92.98 | 92.97 |

Specifically, the Random Forest classifier achieves a maximum accuracy of 94.43% without image segmentation and an improved accuracy of 96.62% with image segmentation both occurring at $K$ = 30. Fig. 3 visually depicts the comparative accuracy of all the classifiers, showcasing the advantage of image segmentation in enhancing classification results, especially at $K$ = 30.

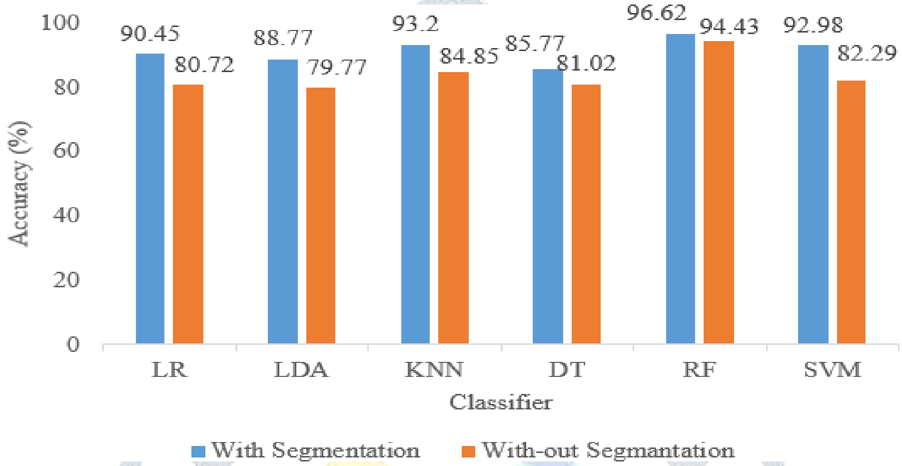


Fig. 3. Effect of image segmentation on classifier's performance

In the segmentation process the leaf background is removed, thereby eliminating the unwanted information in the image which leads to improved accuracy.

The performance metrics for all six classifiers on the tomato leaf dataset using hyperparameter optimization are shown in Table III. Table IV the classification results are better with hyperparameter optimization. During the hyperparameter optimization, GridSearchCV evaluates the model performance with multiple combinations of parameters and automatically chooses the optimum parameter for better classification accuracy.

TABLE III. Classifiers' Performance on Tomato Leaf Dataset with Hyperparameter Optimization

| **Classifier** | **Performance Metrics** | | | |
|---|---|---|---|---|
| | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-Score** |
| Logistic Regression | 94.72 | 95.00 | 94.86 | 95.00 |
| Linear Discriminant Analysis | 89.68 | 89.92 | 90.10 | 89.78 |
| K-Nearest Neighbor | 94.44 | 95.00 | 94.70 | 94.97 |
| Decision Tree Classieier | 85.86 | 86.00 | 86.05 | 85.00 |
| Random Forest Classifier | 98.02 | 98.00 | 98.71 | 98.23 |
| Support Vector Machine | 97.10 | 97.08 | 97.47 | 97.60 |

TABLE IV. Classification Results with Optimized Hyperparameters

| Plant | Disease classes | Classification accuracy (%) for different classifiers | | | | | |
|---|---|---|---|---|---|---|---|
| | | LRC | LDA | KNN | DTC | SVM | RFC |
| Apple | 4 | 94.36 | 93.55 | 94.76 | 92.54 | 97.98 | 98.79 |
| Cherry | 2 | 99.08 | 99.39 | 99.08 | 97.56 | 99.08 | 99.69 |
| Corn | 4 | 86.17 | 84.84 | 90.69 | 90.56 | 90.29 | 94.02 |
| Grape | 4 | 94.01 | 92.95 | 92.82 | 89.62 | 95.88 | 96.14 |
| Peach | 2 | 99.49 | 98.23 | 98.48 | 98.48 | 98.73 | 99.49 |
| Pepper | 2 | 87.72 | 85.21 | 88.97 | 87.22 | 93.74 | 94.99 |
| Potato | 3 | 97.62 | 93.96 | 98.12 | 96.52 | 99.08 | 99.26 |
| Strawberry | 2 | 93.68 | 89.56 | 95.6 | 97.25 | 98.63 | 99.73 |
| Tomato | 9 | 94.72 | 89.68 | 94.44 | 85.86 | 97.10 | 98.02 |
| Mango | 8 | 97.12 | 93.62 | 97.25 | 93.25 | 98.63 | 99.00 |

The use of machine learning algorithms for plant leaf disease detection is the best idea for the early detection of diseases before they spread over the farm. From Table II, III and Table IV, it is clear that image background removal using the segmentation technique improves the accuracy of image classification and also, the random forest classifier performed best among the machine learning algorithms. The performance of the algorithms varied with the value of $K$ in $K$-fold validation. The classification models have their own training parameters to be tuned for accurate training and classification. The results tabulated in Table IV are evidence for having the best classification results with optimized hyperparameters.

The use of preprocessing techniques like image enhancement, denoising, and threshold-based segmentation helped to identify the disease parts easily and this led to improved classification accuracy in the proposed model compared to other state-of-art methods. The proposed algorithm presents various benefits, especially in terms of its size and resource demands. The algorithm is less computationally intricate, enhancing its suitability for a broader array of hardware.

## V. CONCLUSION

This study presents an effective and automated approach for plant leaf disease detection using a combination of image processing and machine learning techniques. Given that agriculture plays a critical role in India's economy and livelihood, early and accurate identification of plant diseases is essential to minimize crop damage and improve yield. The proposed system uses a well-structured methodology, beginning with image enhancement through techniques like histogram equalization and denoising, followed by segmentation to isolate the affected leaf regions. This improves the quality and focus of the input data, enabling better feature extraction. Features such as Haralick textures, Hu moments, and color histograms are then extracted from grayscale and HSV images to capture texture, shape, and color details of the diseased leaves. These features are fed into multiple machine learning classifiers including Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine. To ensure robust evaluation, K-fold cross-validation is used. Among the models tested, the Random Forest classifier achieved the highest accuracy of 97.92%, especially when segmentation was applied, showing a notable 2.19% improvement in performance. The proposed system demonstrates strong potential for real-world agricultural applications. It provides a reliable, quick, and cost-effective method for farmers and agricultural professionals to identify disease types across different crops. This can lead to timely intervention, reduced use of pesticides, and improved crop health, ultimately supporting sustainable farming practices.

While the current system performs well, there is room for further enhancement. In the future, integration with deep learning models such as Convolutional Neural Networks (CNNs) could automate feature extraction and improve classification accuracy even further. Mobile-based applications or IoT-enabled tools can be developed to bring this technology directly to farmers in rural areas. Additionally, expanding the dataset to include more plant species and disease types will increase the system's applicability and robustness. Incorporating environmental parameters such as temperature, humidity, and soil quality can also enhance disease prediction accuracy. Overall, the system lays a strong foundation for developing intelligent, scalable, and farmer-friendly solutions for precision agriculture.

## VI. ACKNOWLEDGMENT

The authors sincerely acknowledge the unwavering support and guidance of their esteemed mentor and supervisor. Their expert advice, constructive feedback, and consistent encouragement have been crucial at every stage of this research. The technical insights and thoughtful suggestions provided not only enriched the quality of the work but also helped overcome challenges during the study. The authors are deeply grateful for their valuable time, mentorship, and dedication, which played a significant role in the successful completion of this project.

## REFERENCES

[1] P. K. Joshi and S. Pal, "Reforming Indian Agriculture for Sustainable Growth," in *Agricultural Policies in India*, pp. 45–72, 2020, doi: 10.1007/978-981-15-6874-1_3.

[2] R. Kumar and A. Sharma, "Agriculture and its impact on Indian economic development: A comprehensive study," Int. J. Agric. Econ. Stat., vol. 11, no. 2, pp. 150–158, 2020.

[3] S. R. Patil and M. Deshmukh, "Trends and determinants of agricultural exports in India: An empirical analysis," *Indian J.*

*Econ. Dev.*, vol. 16, no. 2, pp. 245–252, 2020, doi: 10.17492/ijed.v16i2.123456.

[4] R. Mehta and A. Singh, “Assessing the impact of climate change on crop productivity in India,” *J. Agric. Sci. Technol.*, vol. 22, no. 4, pp. 89–97, 2020, doi: 10.1007/s12355-020-00456-x.

[5] V. Narayan and S. B. Rao, “Supply chain challenges in Indian agriculture: A case study approach,” *Int. J. Rural Manag.*, vol. 16, no. 1, pp. 45–58, 2020, doi: 10.1177/0973005220903371.

[6] P. Singh and R. Kumar, “Impact of agricultural credit on crop productivity in India: An econometric analysis,” *Asian J. Agric. Ext. Econ. Sociol.*, vol. 38, no. 4, pp. 45–54, 2020, doi: 10.9734/ajaees/2020/v38i430374.

[7] M. Ali and S. Hussain, “Sustainable agriculture in India: Role of organic farming,” *Int. J. Environ. Agric. Biotechnol.*, vol. 5, no. 2, pp. 280–287, 2020, doi: 10.22161/ijeab.52.10.

[8] [R. Verma and A. Chatterjee, “An analysis of crop diversification trends in Indian agriculture,” *J. Agric. Dev. Policy*, vol. 30, no. 1, pp. 15–25, 2020.

[9] D. Vu, T. Nguyen, T. V. Nguyen, T. N. Nguyen, F. Massacci, and P.H. Phung, “HIT4Mal: Hybrid image transformation for malware classification,” Trans. Emerg. Telecommun. Technol., vol. 31, no. 11,p. e3789, 2020, doi: 10.1002/ett.3789.

[10] E. David et al., “Global Wheat Head Detection (GWHD) Dataset: A Large and Diverse Dataset of High-Resolution RGB-Labelled Images to Develop and Benchmark Wheat Head Detection Methods,” Plant Phenomics, vol. 2020, pp. 1-12, 2020, doi: 10.34133/2020/3521852.

[11] Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). *Plant Disease Detection and Classification by Deep Learning.* Plants, 8(11), 468. https://doi.org/10.3390/plants8110468

[12] X. Bai et al., “Rice heading stage automatic observation by multi- classifier cascade based rice spike detection method,” Agric. For. Meteorol., vol. 259, pp. 260–270, 2018, doi: 10.1016/j.agrformet.2018.05.001.

[13] P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” *Front. Plant Sci.*, vol. 7, p. 1419, 2019, doi: 10.3389/fpls.2016.01419.

[14] M. Halder, A. Sarkar, and H. Bahar, “Plant Disease Detection By Image Processing: A Literature Review,” SDRP J. Food Sci. Technol., vol. 3, no. 6, pp. 534–538, 2018, doi: 10.25177/JFST.3.6.6.

[15] S. Giraddi, S. Desai, and A. Deshpande, “Deep Learning for Agricultural Plant Disease Detection,” in *ICDSMLA 2019*, Lecture Notes in Electrical Engineering, vol. 601, S. Giraddi, S. Desai, A. Deshpande, Eds. Singapore: Springer, 2020, pp. 864–871, doi:10.1007/978-981-15-1420-3_93

[16] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” Comput. Intell. Neurosci., vol. 2020, pp. 1–11, 2020, doi: 10.1155/2020/1350879.

[17] P. Revathi and M. Hemalatha, “Classification of cotton leaf spot diseases using image processing edge detection techniques,” Proc. IEEE ICCSP, pp. 876–880, 2017.

[18] S. Shrivastava and P. Singh, “Detection of tomato plant diseases using image processing techniques,” Proc. IEEE ICCCNT, pp. 1–5, 2017.

[19] A. Camargo and J. S. Smith, “An image-processing based algorithm to automatically identify plant disease visual symptoms,” Comput. Electron. Agric., vol. 102, pp. 9–18, 2018.

[20] R. Ramesh and P. Vydeki, “Recognition and classification of paddy leaf diseases using optimized deep neural network with Jaya algorithm,” Inf. Process. Agric., vol. 5, no. 4, pp. 731–740, 2018.

[21] S. Rani and P. V. Raj, “Hybrid SVM-GLCM approach for plant leaf disease classification using image processing,” Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 1795–1800, 2019.

[22] T. Singh and S. Misra, “Machine learning approach for plant leaf disease detection and classification,” Int. J. Adv. Sci. Technol., vol. 29, no. 3, pp. 1150–1158, 2020.

[23] J. G. A. Barbedo, “Plant disease identification from individual lesions and spots using traditional machine learning techniques,” Comput. Electron. Agric., vol. 172, p. 105350, 2020.

[24] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “Image processing-based detection of fungal diseases in maize leaves using statistical classifiers,” Int. J. Comput. Appl., vol. 975, no. 8887, pp. 1–6, 2020.

[25] N. Kale and D. Pawar, “Machine learning approach for identification of sugarcane leaf diseases using GLCM features,” Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 4, pp. 234–239, 2020.

[26] P. Kumar and M. Hanmandlu, “Classification of plant leaf diseases using hybrid image features and machine learning techniques,” Procedia Comput. Sci., vol. 167, pp. 2394–2403, 2020.

[27] R. Patil and A. Deshmukh, “An image processing and decision tree-based approach for plant disease detection,” Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 6063–6067, 2020.