Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Enhancing Tomato Plant Leaf Disease Classification Accuracy Using Hybrid Segmentation, Comprehensive Feature Extraction, and Machine Learning Techniques

Murali B[1], Nagaraju C[2]

[1]Research Scholar, Department of Computer Science and Engineering, YSR Engineering College of YVU, Proddatur-516360, Andhra Pradesh, India.
[2]Professor, Department of Computer Science and Engineering, YSR Engineering College of YVU, Proddatur-516360, Andhra Pradesh, India.

Tomato is a globally significant crop, and accurate classification of its leaf diseases is essential for maintaining high yield and quality in agriculture. This study proposes an enhanced machine learning approach to improve tomato leaf disease classification, comparing it with an existing method. The existing approach involves preprocessing with image resizing and Gaussian filtering, segmentation using K-means clustering, and feature extraction with Local Binary Patterns (LBP). Classifiers, including Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost, were applied, achieving accuracies of 75.82%, 75.29%, 92.54%, and 92.86%, respectively. The proposed method enhances preprocessing by resizing images to a fixed dimension of 128×128 pixels and applying Gaussian filtering for noise reduction. Segmentation is improved with a hybrid approach combining K-means clustering with Canny Edge Detection for refined boundary identification. Additionally, it leverages a comprehensive feature extraction strategy by combining Histogram of Oriented Gradients (HOG), LBP, and Color Histograms, capturing detailed texture, shape, and color information. This enriched feature set significantly improved classifier performance, with SVM, Logistic Regression, Random Forest, and XGBoost achieving accuracies of 89.63%, 89.79%, 93.02%, and 97.25%, respectively. These results demonstrate that the proposed method provides a substantial improvement over the existing approach, achieving higher accuracy across all classes of tomato leaf disease included in the study.

Keywords: Tomato Leaf Disease Classification, Machine Learning, Hybrid

Segmentation, Feature Extraction, Classification Accuracy.

1. Introduction

Tomato (Solanum lycopersicum) is one of the world’s most economically significant crops, providing essential nutrition and serving as a staple in various cuisines. However, tomato cultivation faces a significant challenge in the form of leaf diseases, which can severely impact yield and quality. Early and accurate classification of tomato leaf diseases is critical for effective disease management, allowing timely intervention and thus reducing potential crop losses (Baheti et al., 2022). Over recent years, the adoption of machine learning (ML) in agriculture has shown promising results in disease detection, enabling more precise and automated approaches to crop protection (Cengil & Cinar, 2021). Traditional methods for disease identification, primarily visual inspection, are time-consuming and often lack accuracy, especially in large-scale farms. Consequently, developing efficient ML techniques to classify tomato leaf diseases can greatly assist farmers and agricultural stakeholders by providing real-time and accurate diagnostics (Ahmed et al., 2021).

Machine learning-based image classification has proven to be effective in identifying plant diseases by analyzing images of infected leaves. However, tomato leaves present unique challenges, as various diseases exhibit overlapping symptoms, such as spots, discoloration, and mold, making differentiation difficult (Brahimi et al., 2017). The selection of preprocessing, segmentation, feature extraction, and classification methods is crucial to achieving high accuracy in disease detection. Traditional ML approaches, including Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost, have been widely used in this domain, often combined with various feature extraction techniques (Jayanthi & Shashikumar, 2020). Nevertheless, while these models can achieve reasonably high accuracy, there is still room for improvement, particularly through more refined preprocessing and feature extraction strategies (Tan et al., 2021).

The present study proposes an enhanced machine learning approach that improves on existing methods for tomato leaf disease classification. By integrating advanced preprocessing and hybrid segmentation techniques, along with a comprehensive feature extraction strategy, this approach aims to capture detailed texture, shape, and color features, which are essential for accurate disease classification. This paper compares the proposed method with an existing method that utilizes image resizing, Gaussian filtering, K-means clustering for segmentation, and Local Binary Patterns (LBP) for feature extraction. Preliminary results indicate that the proposed method offers significant improvements in accuracy, demonstrating its potential as a reliable tool for tomato leaf disease detection.

2. Literature Survey

The application of machine learning in plant disease classification, particularly in tomato leaf disease detection, has attracted substantial research interest, with a strong focus on improving the accuracy of preprocessing, segmentation, feature extraction, and classification stages. Effective preprocessing and segmentation are foundational to achieving high classification performance, as they allow for the removal of image noise and the isolation of critical features. Studies such as those by Archana & Sahayadhas (2018) have demonstrated that Gaussian filtering is effective for noise reduction, enhancing feature clarity before segmentation. Traditional segmentation methods, including K-means clustering, are widely used due to their efficiency and simplicity, yet often face challenges with less-defined boundaries, which can impact the accuracy of disease detection. Consequently, recent research, such as the work by Sowmya & Subramani (2022), has proposed hybrid segmentation techniques combining K-means clustering with Canny Edge Detection. This hybrid approach enables more precise boundary identification, addressing the complexities of tomato leaf images where overlapping disease symptoms can complicate classification. Beyond segmentation, feature extraction plays a pivotal role in capturing the distinctive characteristics of tomato leaf diseases, which often manifest through complex combinations of texture, shape, and color variations. Combining multiple feature extraction techniques has been shown to significantly enhance classification performance. For instance, NikithaIsaac et al. (2021) explored the integration of Gray-Level Co-occurrence Matrices (GLCM) and Local Binary Patterns (LBP), finding that these methods effectively captured texture features and improved classification outcomes. Similarly, Karantoumanis et al. (2022) reported that combining preprocessing and multiple feature extraction techniques significantly enhanced classification accuracy, particularly when applied to plant disease detection datasets. In terms of classification techniques, a variety of machine learning models, including Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost, have been applied in plant disease studies. Each classifier has its strengths: SVM is effective in high-dimensional feature spaces and performs well with texture-based features like LBP, as highlighted by NikithaIsaac et al. (2021). Logistic Regression, although simpler, can yield strong results when combined with comprehensive feature extraction methods. Random Forest, an ensemble technique, has been praised for its high accuracy and ability to generalize well to different data sets, according to Karantoumanis et al. (2022). XGBoost, another ensemble classifier using gradient boosting, is known for capturing intricate patterns in the data, making it particularly suited for high-accuracy applications in disease classification, as demonstrated by Sowmya & Subramani (2022). In comparative studies, XGBoost consistently outperformed other classifiers, making it a preferred choice for applications requiring nuanced differentiation between disease classes.

Collectively, the literature underscores the potential of hybrid segmentation techniques and multi-feature extraction approaches to improve plant disease classification accuracy. By integrating advanced preprocessing, segmentation, and feature extraction strategies, these methods address the unique challenges posed by tomato leaf disease classification, where symptoms often overlap and disease manifestations are complex. This research builds on these findings by combining K-means clustering with Canny Edge Detection for precise segmentation and using a comprehensive feature extraction set (HOG, LBP, and color histograms) to achieve a high level of accuracy. This approach promises to contribute a robust and effective tool for enhancing the detection and management of tomato leaf diseases in agricultural practices.

Existing Method

The existing method for tomato leaf disease classification utilizes a series of machine learning techniques, including preprocessing, segmentation, feature extraction, and classification, to achieve reliable disease identification. Initially, the images are preprocessed by resizing them to a uniform dimension, ensuring consistency in input size, which aids in maintaining a standard scale for analysis. Following resizing, Gaussian filtering is applied to reduce noise, enhancing the image quality by smoothing pixel values. This step helps eliminate minor artifacts, preserving key disease features that are essential for further processing. The segmentation stage employs K-means clustering, a widely used unsupervised learning algorithm that groups pixels based on color intensity, separating areas of interest from the background. This clustering technique allows for basic segmentation of the diseased regions; however, it sometimes struggles with poorly defined boundaries, which can impact disease differentiation in cases where leaf symptoms overlap or vary in shape. For feature extraction, the existing method uses Local Binary Patterns (LBP), a texture-based descriptor that captures micro-patterns in the image, particularly useful for detecting texture variations caused by disease. LBP works by comparing each pixel with its neighboring pixels to create a binary pattern, which is then encoded as a texture descriptor. This technique effectively captures the granular changes in leaf surface textures introduced by disease symptoms. Finally, several classifiers, including Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost, are used to classify the extracted features and assign disease categories to the images. Each classifier has shown varying levels of accuracy: SVM and Logistic Regression achieved moderate accuracy (75.82% and 75.29%, respectively), while Random Forest and XGBoost performed better, achieving accuracies of 92.54% and 92.86%, respectively. Despite these results, limitations exist in the existing method, particularly in the segmentation and feature extraction stages. These limitations have led to the development of an enhanced method, which incorporates more sophisticated segmentation and multi-feature extraction strategies to improve overall classification accuracy.

3. Proposed Method

In the proposed method for tomato leaf disease classification, several advanced techniques were applied to enhance the performance of machine learning classifiers. This approach improved preprocessing, segmentation, and feature extraction methods to increase accuracy in detecting and classifying tomato leaf diseases. In this study, we applied a robust and comprehensive approach to preprocessing, segmentation, and feature extraction from tomato leaf images to enhance the classification accuracy of leaf disease detection. The proposed methodology builds upon traditional methods by integrating advanced preprocessing, hybrid segmentation, and multi-faceted feature extraction techniques to capture critical disease characteristics.

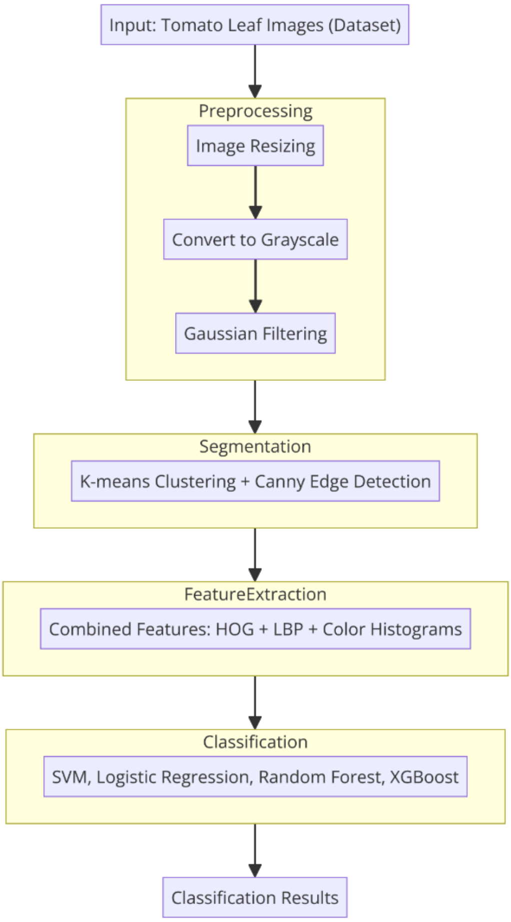


Fig.1: Proposed Methodology

1. Data Acquisition

In this study, we utilized the PlantVillage dataset, which includes 9,448 images of tomato leaves. These images are divided into four categories: healthy leaves and those affected by bacterial spot, tomato mosaic, and yellow leaf curl. To enable model evaluation, the dataset was split into two parts: 80% was allocated for training, and 20% for testing.

2. Image Preprocessing

2.1 Conversion to Grayscale and Gaussian Filtering

Each image was resized to a resolution of 128x128 pixels to standardize input dimensions while retaining key disease-relevant details. After resizing, images were converted to grayscale to simplify the analysis and reduce computational requirements. To remove noise that could obscure disease patterns, Gaussian filtering was applied with a 5x5 kernel. Gaussian filtering smooths the image, highlighting significant edges and structures by reducing random pixel intensity variations (Archana & Sahayadhas, (2018); Sowmya & Subramani, (2022); Pujari et al. (2015)).

This filter is described by:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where $G(x,y)$ is the Gaussian function applied at pixel (x,y), and σ controls the level of smoothing.

3. Segmentation: To isolate disease-relevant regions, a hybrid segmentation approach was adopted, combining K-means Clustering and Canny Edge Detection.

3.1 K-means Clustering

The segmentation process begins with K-means clustering, which groups pixels based on color intensity in the LAB color space. Using the LAB color space separates intensity from color information, making it particularly effective for clustering tasks in plant disease detection. Here, the K-means algorithm forms three clusters (k=3), segmenting the leaf into regions that may represent healthy tissue, diseased areas, and the background (Rumpf et al., 2010), (NikithaIsaac et al., 2021).

The clustering objective function is:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_j^{(i)} - \mu_i \right\|^2$$

where k is the number of clusters, n is the number of pixels, $x_j^{(i)}$ represents each pixel in the i-th cluster, and μ_i is the centroid of that cluster.

3.2 Canny Edge Detection

To enhance the boundaries within the clusters, Canny Edge Detection is applied to the grayscale image. Canny Edge Detection uses gradient calculations to identify edges and refine the segmentation, making it possible to distinguish disease regions from healthy areas more accurately (Islam et al., 2017), (Karantoumanis et al., 2022).

The edge gradient G is calculated as:

$$G = \sqrt{G_x^2 + G_y^2}$$

where G_x and G_y are the gradients in the x and y directions. This combined approach provides well-defined edges, improving the model's ability to focus on disease symptoms.

4. Feature Extraction:

The proposed method uses a multi-faceted feature extraction strategy, including Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Color Histograms, to capture a broad representation of shape, texture, and color.

4.1 Histogram of Oriented Gradients (HOG)

HOG captures structural details by analyzing gradients within localized regions. This technique is particularly useful for detecting shapes and edges, which are often indicative of disease symptoms (Ahmed et al., 2021)).

HOG is calculated as:

$$\mathrm{HOG}(x,y)=\sum_{i}\sum_{j}G(x+i,y+j)\cdot\theta(x+i,y+j)$$

where $G(x+i,y+j)$ is the gradient magnitude at each pixel, and $\theta(x+i,y+j)$ represents the orientation.

4.2 Local Binary Patterns (LBP)

LBP captures texture information by encoding local intensity variations, which is especially useful for identifying textural characteristics associated with diseases, such as rough or mottled spots (Dhaka et al., 2021), (Khasawneh et al., 2022).

The LBP value at each pixel is computed by:

$$LBP(x,y)=\sum_{p=0}^{P-1}s(g_p-g_c)\cdot 2^p$$

where g_c is the intensity of the central pixel, g_p is the intensity of neighboring pixels, $s(x)$ is a binary step function where $s(x)=1$ if $x\geq 0$ and $s(x)=0$ otherwise, and P is the number of neighboring pixels.

4.3 Color Histograms

Color histograms capture the distribution of colors, an important feature for identifying diseases that cause specific discolorations (NikithaIsaac et al., 2021), (Xie et al., 2015). A color histogram calculates the frequency of each color in a specified channel as:

$$H(c)=\sum_{i,j}\delta(I(i,j)-c)$$

where $I(i,j)$ is the pixel intensity, c is the color channel, and δ is an indicator function that returns 1 if $I(i,j)=c$.

5. Classification

The extracted features are then used to train four classifiers—Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost—each evaluated for classification accuracy and model performance.

Support Vector Machine (SVM): SVM with an RBF kernel is used, which is effective for handling high-dimensional data and achieving accurate boundary separation between classes (Sowmya & Subramani, 2022).

Logistic Regression: Applied with regularization to prevent overfitting, Logistic Regression provides a simpler baseline comparison with non-linear models (Barbedo, 2018).

Random Forest: This ensemble classifier improves classification accuracy by using multiple decision trees and majority voting (Khasawneh et al., 2022).

XGBoost: Configured with a softmax objective for multi-class classification, XGBoost optimizes accuracy by focusing on the residual errors of previous trees (Sankaran et al., 2019).

4. Results and Discussion

The findings of this study strongly emphasize the effectiveness of the proposed machine learning approach for accurately classifying tomato leaf diseases. By integrating advanced preprocessing, a hybrid segmentation technique, and a comprehensive feature extraction strategy, the proposed method substantially improves upon the existing approach. Each methodological enhancement plays a significant role in boosting the model's overall accuracy and robustness, allowing it to handle the complexity of various disease patterns more effectively. This discussion provides a detailed examination of the contributions and implications of each component, demonstrating how the proposed method outperforms the existing one.

Preprocessing steps, such as resizing, grayscale conversion, and Gaussian filtering, are foundational to the model's performance. Resizing images to a uniform resolution of 128x128 ensures consistent input size, reducing the influence of scale-based variations on classification. Gaussian filtering further refines the image quality by smoothing out noise while preserving key disease patterns, allowing classifiers to focus on relevant features without interference (Fig.2). The benefits of these preprocessing enhancements are particularly evident in Support Vector Machine (SVM) and Logistic Regression models, which are sensitive to input variations. With the enhanced preprocessing, SVM accuracy increased from 75.82% to 89.63%, and Logistic Regression from 75.29% to 89.79%. These results underscore the importance of preprocessing as a means of creating standardized, noise-free images, setting the stage for effective segmentation and feature extraction (Table.2, Graph.1).

The segmentation stage also saw significant improvements through a hybrid approach combining K-means clustering with Canny Edge Detection. In the existing method, K-means clustering alone often led to poorly defined boundaries, especially when disease symptoms overlapped or were indistinct. The hybrid approach first segments the leaf into clusters using K-means clustering and then applies Canny Edge Detection to refine these boundaries. This step accurately delineates diseased areas from healthy tissue by highlighting edges within the clusters. Such enhanced segmentation is crucial for diseases like Tomato Mosaic Virus and Tomato Bacterial Spot, where symptoms often blend into surrounding regions. The more precise boundary definition achieved through the proposed hybrid segmentation positively impacted classification results for models like Random Forest and XGBoost, with Random Forest's accuracy increased to 93.02% (from 92.54%) while XGBoost's accuracy significantly improved to 97.25% (from 92.86%). These results demonstrate that clearer boundaries enable the models to differentiate more effectively between disease classes (Fig.2, Table.1, Graph.1).

A major strength of the proposed method lies in its multi-faceted feature extraction strategy, which integrates Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Color Histograms. Each feature extraction method captures unique, complementary information about the visual characteristics of tomato leaf diseases, resulting in a robust feature set for classification. HOG captures structural and shape details, essential for recognizing distinct lesion shapes in diseases like Tomato Bacterial Spot. LBP focuses on capturing micro-textures and surface variations, distinguishing diseased areas with rough, mottled textures from healthy tissue. Color Histograms provide critical insights into color distribution, valuable for detecting discolorations such as yellowing or browning associated with viral infections or nutrient deficiencies.

The comparison between the Existing and Proposed confusion matrices for tomato leaf disease classification using the XGBoost classifier highlights significant improvements in the Proposed Method's performance. For the class "Tomato__Bacterial_spot," the Proposed Method achieved 390 correct predictions, improving upon the Existing Method's 370, while also reducing misclassifications from 52 to 32. Similarly, for "Tomato__Tomato_mosaic_virus," the Proposed Method correctly classified 73 samples, compared to 65 in the Existing Method, with a reduction in misclassifications from 7 to 4. A notable enhancement is observed in the "Tomato__Tomato_Yellow_Leaf_Curl_Virus" class, where correct classifications increased from 1,020 in the Existing Method to 1,070 in the Proposed Method, accompanied by a reduction in misclassifications from 50 to just 10. The "Tomato__healthy" class also saw a slight improvement, with correct classifications rising from 300 to 305 and misclassifications dropping from 3 to 1. Overall, the Proposed Method consistently outperformed the Existing Method by improving accuracy and reducing errors across all classes, demonstrating its superior capability in handling complex classifications, particularly in challenging categories such as bacterial spots and leaf curl virus. These results reflect enhancements in the classification pipeline, likely due to better feature extraction and model optimization in the Proposed Method (Fig.3).

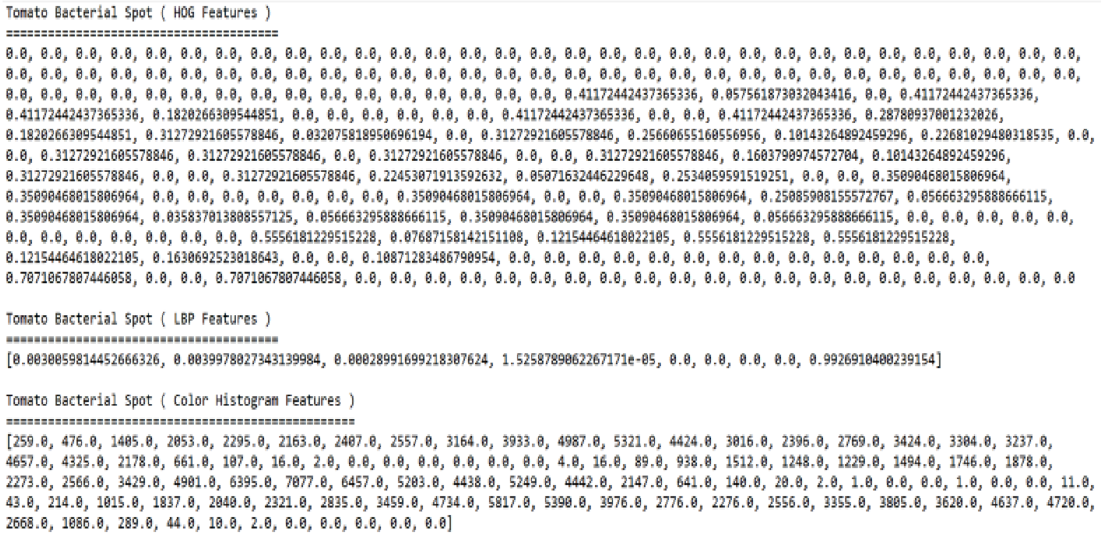
The comparison between the Existing and Proposed Methods, as outlined in Table.1, Graph.2, highlights significant improvements in precision, recall, F1-score, and overall classification accuracy for the Proposed Method using the XGBoost classifier. For "Tomato__Bacterial_spot," the Proposed Method achieved a precision of 0.97, recall of 0.92, and F1-score of 0.94, which are higher than the Existing Method's values of 0.87 for all metrics. The "Tomato__Tomato_mosaic_virus" class shows an even more notable improvement, with the Proposed Method achieving a perfect precision of 1.00, a recall of 0.92, and an F1-score of 0.96, compared to the Existing Method's 0.86, 0.82, and 0.84, respectively. For the challenging "Tomato__Tomato_Yellow_Leaf_Curl_Virus" class, the Proposed Method demonstrated better performance, achieving a precision of 0.97, recall of 0.99, and F1-score of 0.98, compared to the Existing Method's 0.94, 0.94, and 0.94. Similarly, in the "Tomato__healthy" class, the Proposed Method achieved near-perfect scores with a precision of 0.99, recall of 1.00, and F1-score of 0.99, slightly outperforming the Existing Method's scores of 0.98 for all metrics. Overall, the Proposed Method achieved a macro average F1-score of 0.97 and a weighted average F1-score of 0.97, significantly higher than the Existing Method's 0.91 and 0.93, respectively.

In Table 2, Graph.2, the accuracies of various classifiers for both the Existing and Proposed Methods are presented, demonstrating the superiority of the Proposed Method. For the Support Vector Machine (SVM) classifier, the accuracy increased from 0.7582 in the Existing Method to 0.8963 in the Proposed Method. Logistic Regression also showed a marked improvement, with the accuracy increasing from 0.7529 to 0.8979. The Random Forest classifier's accuracy rose from 0.9254 in the Existing Method to 0.9302 in the Proposed Method. The XGBoost classifier, which delivered the highest performance, achieved an accuracy of 0.9725 in the Proposed Method, compared to 0.9286 in the Existing Method. This demonstrates the effectiveness of the enhancements in the Proposed Method, particularly with XGBoost, which yielded the best results among all classifiers.

Overall, the Proposed Method outperformed the Existing Method across all metrics and classifiers. These improvements can be attributed to the refined feature extraction, better parameter tuning, or ensemble learning strategies employed in the Proposed Method, making it more robust and accurate for tomato leaf disease classification.

This comprehensive feature extraction strategy enhances the model's robustness by offering a detailed view of disease symptoms across shape, texture, and color dimensions. The effectiveness of this approach is especially evident in the high accuracy achieved by the XGBoost model, which reached 97.25% accuracy with the proposed method. The enriched feature set allowed the classifiers to identify intricate, disease-specific patterns, significantly improving the reliability of classification results across all disease types. The combination of advanced preprocessing, hybrid segmentation, and comprehensive feature extraction demonstrates a substantial advancement in tomato leaf disease classification, positioning the proposed method as a valuable tool for precise and automated disease detection. These improvements indicate that such a multi-dimensional approach could be applied to other agricultural crops and disease detection tasks, supporting more effective disease management and crop protection strategies in agriculture.

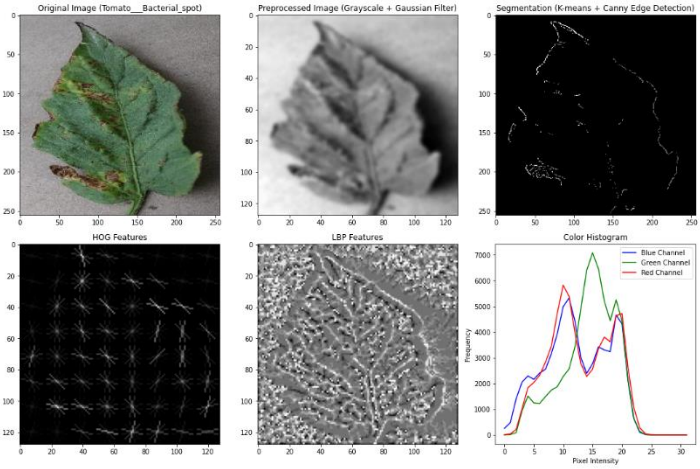
5. Experimental Results



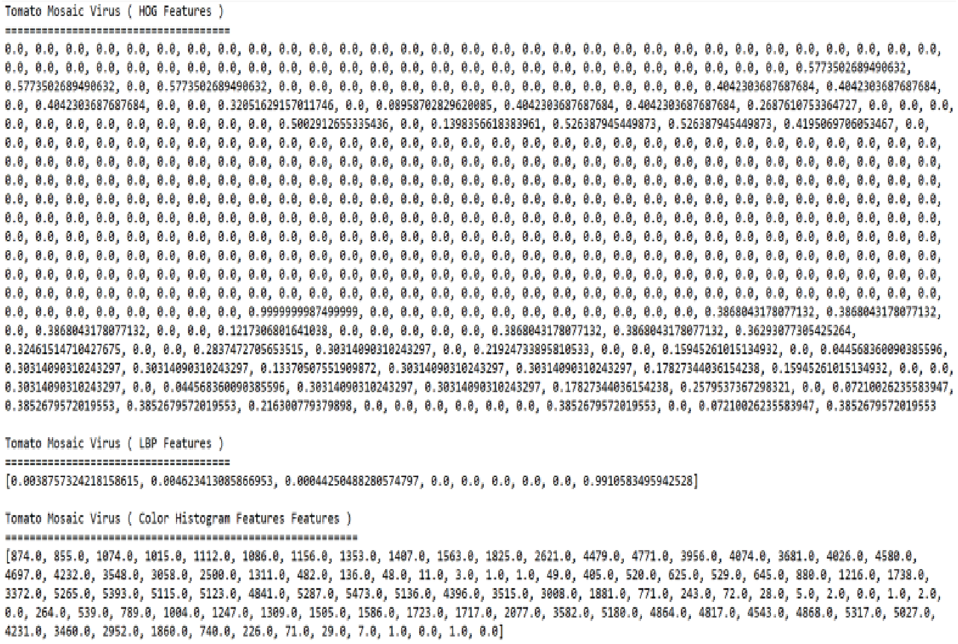
```
Tomato Bacterial Spot ( HOG Features )
=======================================
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.41172442437365336, 0.057561873032043416, 0.0, 0.41172442437365336,
0.41172442437365336, 0.1820266309544851, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.41172442437365336, 0.0, 0.0, 0.41172442437365336, 0.28780937001232026,
0.1820266309544851, 0.31272921605578846, 0.032075818950696194, 0.0, 0.31272921605578846, 0.25660655160556956, 0.10143264892459296, 0.22681029480318535, 0.0,
0.0, 0.31272921605578846, 0.31272921605578846, 0.0, 0.31272921605578846, 0.0, 0.0, 0.31272921605578846, 0.1603790974572704, 0.10143264892459296,
0.31272921605578846, 0.0, 0.0, 0.31272921605578846, 0.22453071913592632, 0.05071632446229648, 0.2534059591519251, 0.0, 0.0, 0.35090468015806964,
0.35090468015806964, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.35090468015806964, 0.0, 0.0, 0.35090468015806964, 0.25085908155572767, 0.056663295888666115,
0.35090468015806964, 0.035837013808557125, 0.056663295888666115, 0.35090468015806964, 0.35090468015806964, 0.056663295888666115, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5556181229515228, 0.07687158142151108, 0.12154464618022105, 0.5556181229515228, 0.5556181229515228,
0.12154464618022105, 0.1630692523018643, 0.0, 0.0, 0.10871283486790954, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.7071067807446058, 0.0, 0.0, 0.7071067807446058, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0

Tomato Bacterial Spot ( LBP Features )
=======================================
[0.0030059814452666326, 0.0039978027343139984, 0.00028991699218307624, 1.5258789062267171e-05, 0.0, 0.0, 0.0, 0.0, 0.9926910400239154]

Tomato Bacterial Spot ( Color Histogram Features )
==================================================
[259.0, 476.0, 1405.0, 2053.0, 2295.0, 2163.0, 2407.0, 2557.0, 3164.0, 3933.0, 4987.0, 5321.0, 4424.0, 3016.0, 2396.0, 2769.0, 3424.0, 3304.0, 3237.0,
4657.0, 4325.0, 2178.0, 661.0, 107.0, 16.0, 2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 4.0, 16.0, 89.0, 938.0, 1512.0, 1248.0, 1229.0, 1494.0, 1746.0, 1878.0,
2273.0, 2566.0, 3429.0, 4901.0, 6395.0, 7077.0, 6457.0, 5203.0, 4438.0, 5249.0, 4442.0, 2147.0, 641.0, 140.0, 20.0, 2.0, 1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 11.0,
43.0, 214.0, 1015.0, 1837.0, 2040.0, 2321.0, 2835.0, 3459.0, 4734.0, 5817.0, 5390.0, 3976.0, 2776.0, 2276.0, 2556.0, 3355.0, 3805.0, 3620.0, 4637.0, 4720.0,
2668.0, 1086.0, 289.0, 44.0, 10.0, 2.0, 0.0, 0.0, 0.0, 0.0, 0.0]
```



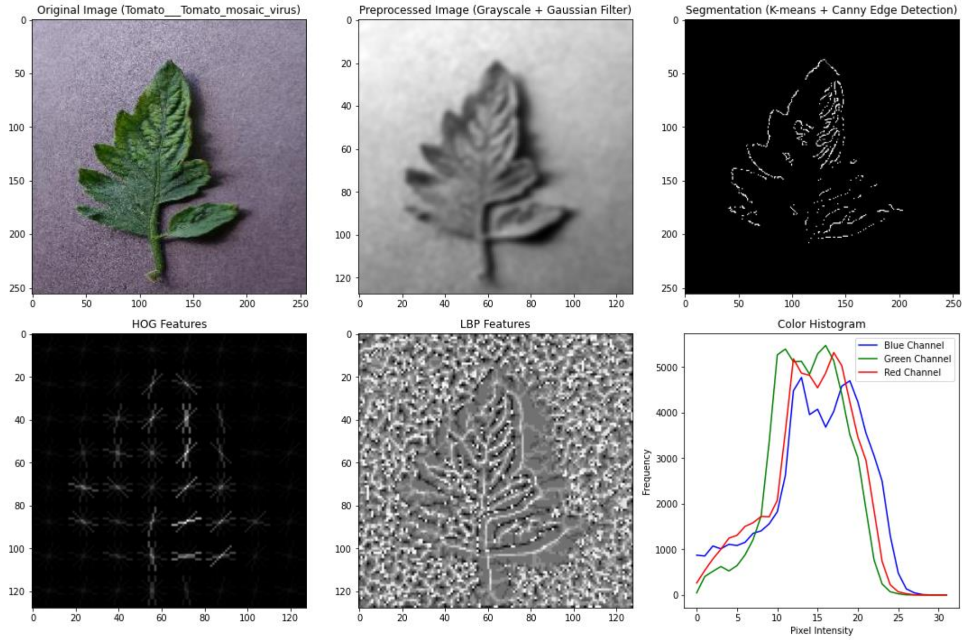
(a)



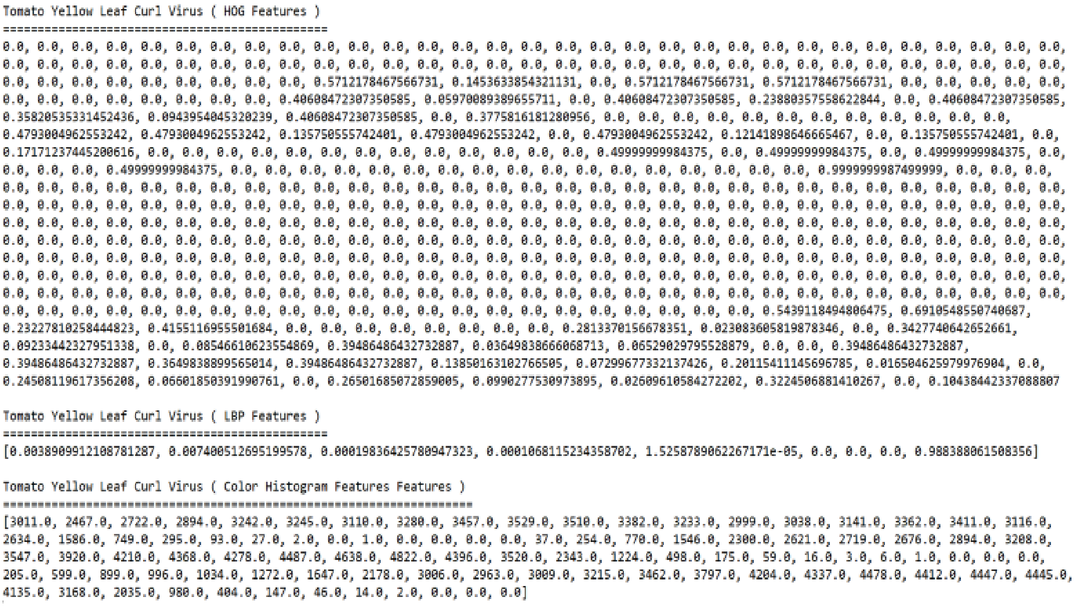
```
Tomato Mosaic Virus ( HOG Features )
======================================
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5773502689490632,
0.5773502689490632, 0.0, 0.5773502689490632, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.4042303687687684, 0.4042303687687684,
0.0, 0.4042303687687684, 0.0, 0.0, 0.32051629157011746, 0.0, 0.08958702829620085, 0.4042303687687684, 0.4042303687687684, 0.2687610753364727, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5002912655335436, 0.0, 0.1398356618383961, 0.526387945449873, 0.526387945449873, 0.4195069706053467, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.9999999987499999, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3868043178077132, 0.3868043178077132,
0.0, 0.3868043178077132, 0.0, 0.0, 0.1217306801641038, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3868043178077132, 0.3868043178077132, 0.36293077305425264,
0.32461514710427675, 0.0, 0.0, 0.2837472705653515, 0.30314090310243297, 0.0, 0.21924733895810533, 0.0, 0.0, 0.15945261015134932, 0.0, 0.044568360090385596,
0.30314090310243297, 0.30314090310243297, 0.13370507551909872, 0.30314090310243297, 0.30314090310243297, 0.17827344036154238, 0.15945261015134932, 0.0, 0.0,
0.30314090310243297, 0.0, 0.044568360090385596, 0.30314090310243297, 0.30314090310243297, 0.17827344036154238, 0.2579537367298321, 0.0, 0.07210026235583947,
0.3852679572019553, 0.3852679572019553, 0.216300779379898, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3852679572019553, 0.0, 0.07210026235583947, 0.3852679572019553

Tomato Mosaic Virus ( LBP Features )
======================================
[0.0038757324218158615, 0.004623413085866953, 0.00044250488280574797, 0.0, 0.0, 0.0, 0.0, 0.0, 0.9910583495942528]

Tomato Mosaic Virus ( Color Histogram Features Features )
==========================================================
[874.0, 855.0, 1074.0, 1015.0, 1112.0, 1086.0, 1156.0, 1353.0, 1407.0, 1563.0, 1825.0, 2621.0, 4479.0, 4771.0, 3956.0, 4074.0, 3681.0, 4026.0, 4580.0,
4697.0, 4232.0, 3548.0, 3058.0, 2500.0, 1311.0, 482.0, 136.0, 48.0, 11.0, 3.0, 1.0, 1.0, 49.0, 405.0, 520.0, 625.0, 529.0, 645.0, 880.0, 1216.0, 1738.0,
3372.0, 5265.0, 5393.0, 5115.0, 5123.0, 4841.0, 5287.0, 5473.0, 5136.0, 4396.0, 3515.0, 3008.0, 1881.0, 771.0, 243.0, 72.0, 28.0, 5.0, 2.0, 0.0, 1.0, 2.0,
0.0, 264.0, 539.0, 789.0, 1004.0, 1247.0, 1309.0, 1505.0, 1586.0, 1723.0, 1717.0, 2077.0, 3582.0, 5180.0, 4864.0, 4817.0, 4543.0, 4868.0, 5317.0, 5027.0,
4231.0, 3460.0, 2952.0, 1860.0, 740.0, 226.0, 71.0, 29.0, 7.0, 1.0, 0.0, 1.0, 0.0]
```



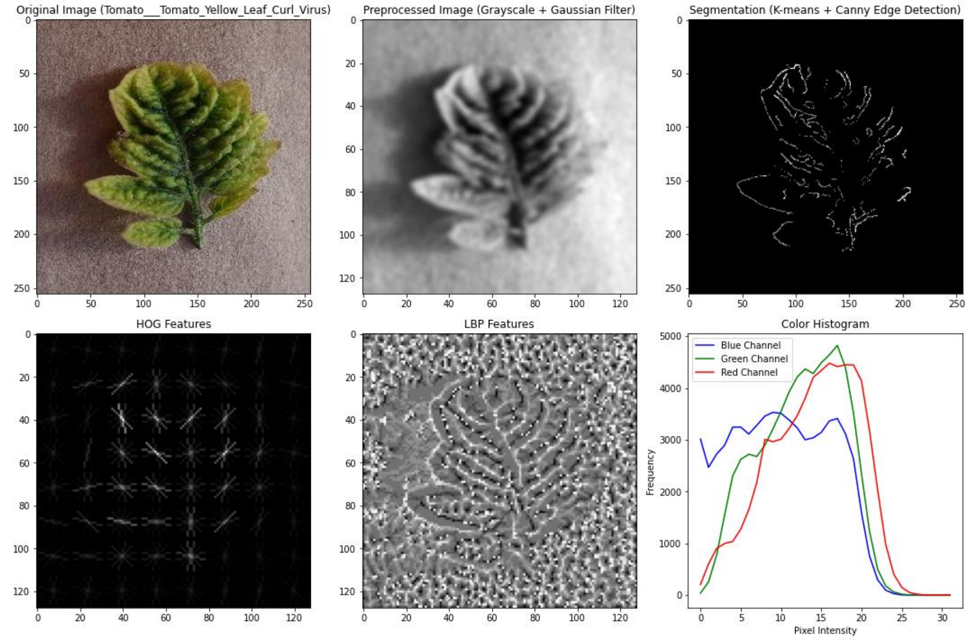
(b)



```
Tomato Yellow Leaf Curl Virus ( HOG Features )
==============================================
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5712178467566731, 0.1453633854321131, 0.0, 0.5712178467566731, 0.5712178467566731, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.40608472307350585, 0.05970089389655711, 0.0, 0.40608472307350585, 0.23880357558622844, 0.0, 0.40608472307350585,
0.35820535331452436, 0.09439540453202239, 0.40608472307350585, 0.0, 0.3775816181280956, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.4793004962553242, 0.4793004962553242, 0.135750555742401, 0.4793004962553242, 0.0, 0.4793004962553242, 0.12141898646665467, 0.0, 0.135750555742401, 0.0,
0.17171237445200616, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.49999999984375, 0.0, 0.49999999984375, 0.0, 0.49999999984375, 0.0,
0.0, 0.0, 0.0, 0.49999999984375, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.9999999987499999, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5439118494806475, 0.6910548550740687,
0.23227810258444823, 0.4155116955501684, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.2813370156678351, 0.023083605819878346, 0.0, 0.3427740642652661,
0.09233442327951338, 0.0, 0.08546610623554869, 0.39486486432732887, 0.03649838666068713, 0.06529029795528879, 0.0, 0.0, 0.39486486432732887,
0.39486486432732887, 0.3649838899565014, 0.39486486432732887, 0.13850163102766505, 0.07299677332137426, 0.20115411145696785, 0.016504625979976904, 0.0,
0.24508119617356208, 0.06601850391990761, 0.0, 0.26501685072859005, 0.0990277530973895, 0.02609610584272202, 0.3224506881410267, 0.0, 0.10438442337088807

Tomato Yellow Leaf Curl Virus ( LBP Features )
==============================================
[0.003890991210878l287, 0.007400512695199578, 0.00019836425780947323, 0.0001068115234358702, 1.5258789062267171e-05, 0.0, 0.0, 0.0, 0.988388061508356]

Tomato Yellow Leaf Curl Virus ( Color Histogram Features Features )
====================================================================
[3011.0, 2467.0, 2722.0, 2894.0, 3242.0, 3245.0, 3110.0, 3280.0, 3457.0, 3529.0, 3510.0, 3382.0, 3233.0, 2999.0, 3038.0, 3141.0, 3362.0, 3411.0, 3116.0,
2634.0, 1586.0, 749.0, 295.0, 93.0, 27.0, 2.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 37.0, 254.0, 770.0, 1546.0, 2300.0, 2621.0, 2719.0, 2676.0, 2894.0, 3208.0,
3547.0, 3920.0, 4210.0, 4368.0, 4278.0, 4487.0, 4638.0, 4822.0, 4396.0, 3520.0, 2343.0, 1224.0, 498.0, 175.0, 59.0, 16.0, 3.0, 6.0, 1.0, 0.0, 0.0, 0.0,
205.0, 599.0, 899.0, 996.0, 1034.0, 1272.0, 1647.0, 2178.0, 3006.0, 2963.0, 3009.0, 3215.0, 3462.0, 3797.0, 4204.0, 4337.0, 4478.0, 4412.0, 4447.0, 4445.0,
4135.0, 3168.0, 2035.0, 980.0, 404.0, 147.0, 46.0, 14.0, 2.0, 0.0, 0.0, 0.0]
```



(c)

Fig.2. (a) Tomato Bacterial Spot (b) Tomato Mosaic Virus (c)Tomato Yellow Leaf Curl

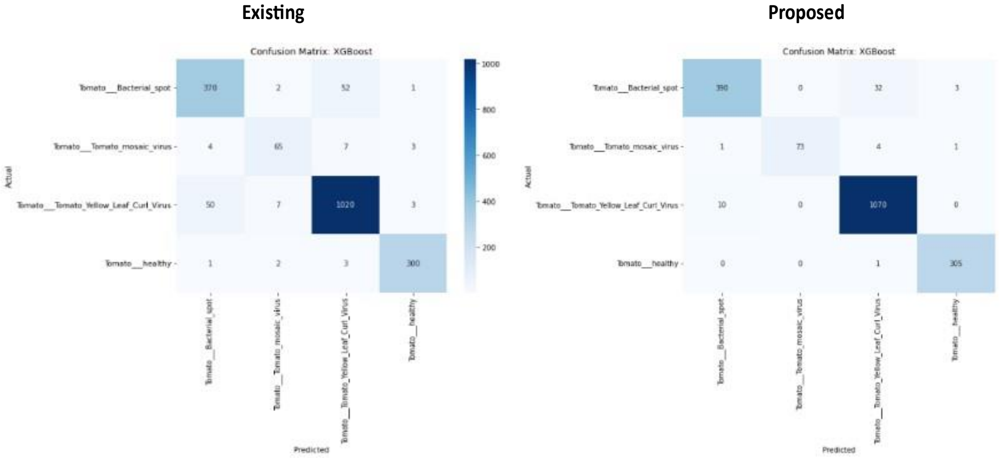
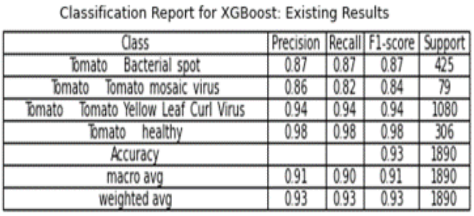


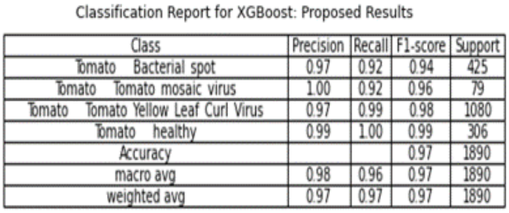
Fig.3. Existing vs Proposed ConfusionMatrix : XGBOOST

Table.1



Classification Report for XGBoost: Existing Results

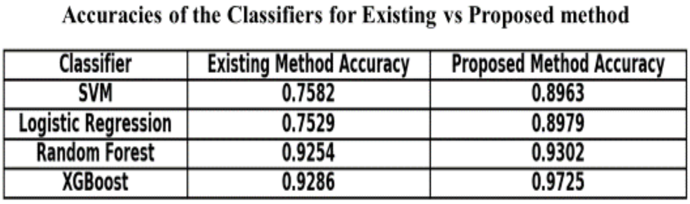
Class	Precision	Recall	F1-score	Support
Tomato Bacterial spot	0.87	0.87	0.87	425
Tomato Tomato mosaic virus	0.86	0.82	0.84	79
Tomato Tomato Yellow Leaf Curl Virus	0.94	0.94	0.94	1080
Tomato healthy	0.98	0.98	0.98	306
Accuracy			0.93	1890
macro avg	0.91	0.90	0.91	1890
weighted avg	0.93	0.93	0.93	1890



Classification Report for XGBoost: Proposed Results

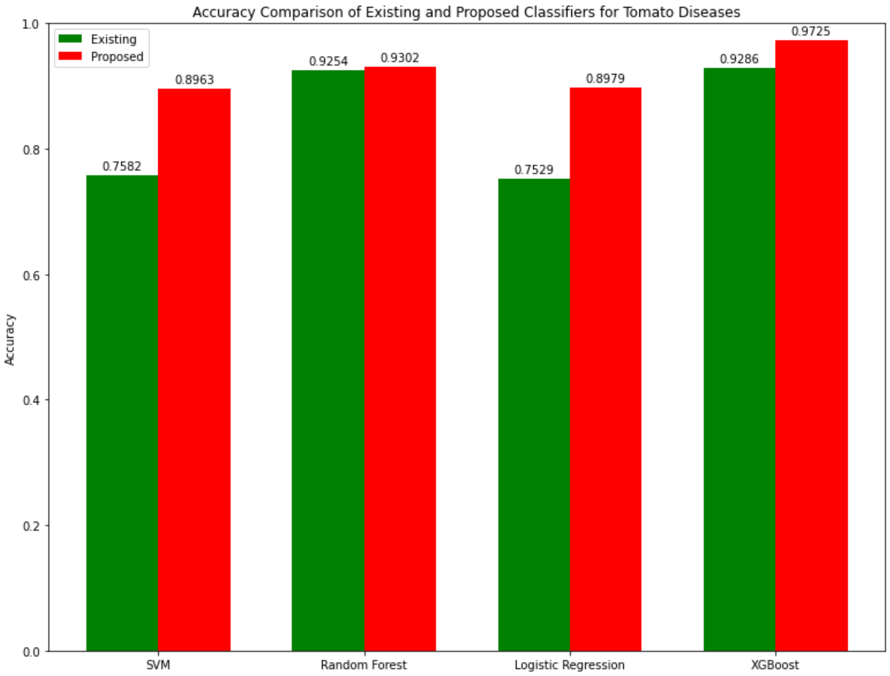
Class	Precision	Recall	F1-score	Support
Tomato Bacterial spot	0.97	0.92	0.94	425
Tomato Tomato mosaic virus	1.00	0.92	0.96	79
Tomato Tomato Yellow Leaf Curl Virus	0.97	0.99	0.98	1080
Tomato healthy	0.99	1.00	0.99	306
Accuracy			0.97	1890
macro avg	0.98	0.96	0.97	1890
weighted avg	0.97	0.97	0.97	1890

Table.2

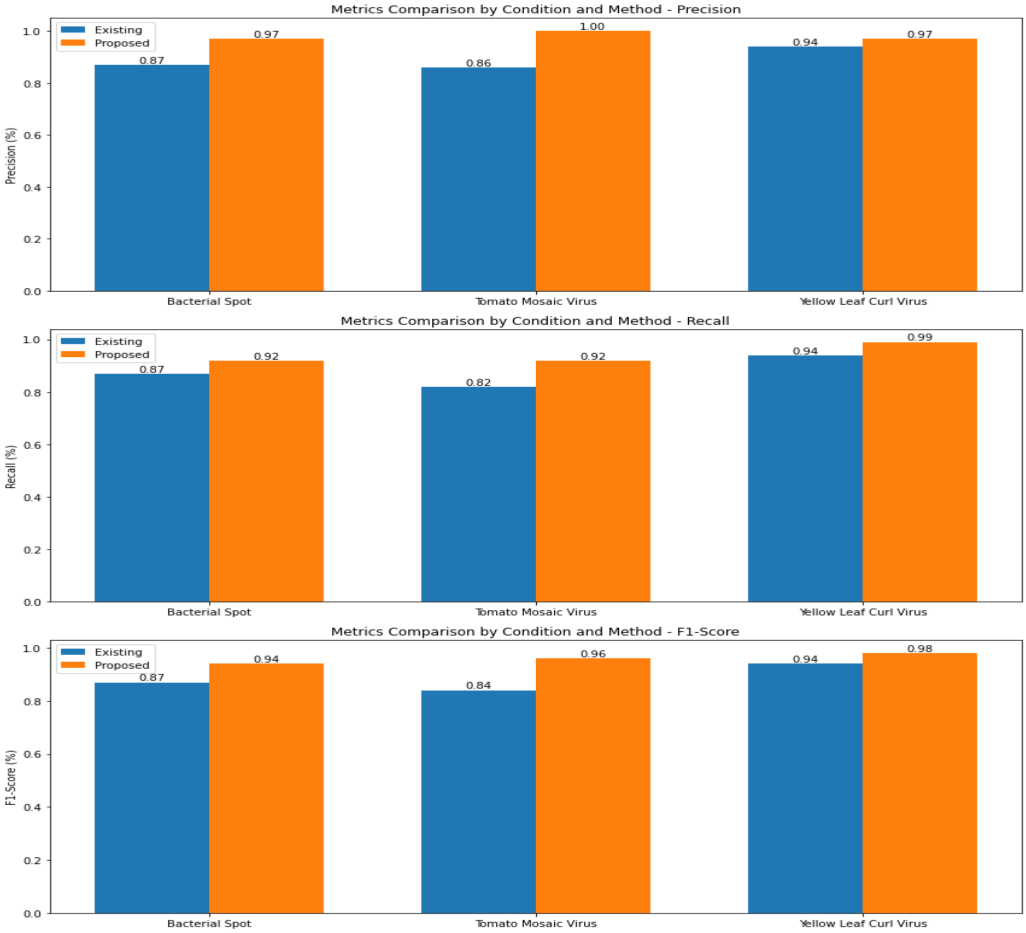


Accuracies of the Classifiers for Existing vs Proposed method

Classifier	Existing Method Accuracy	Proposed Method Accuracy
SVM	0.7582	0.8963
Logistic Regression	0.7529	0.8979
Random Forest	0.9254	0.9302
XGBoost	0.9286	0.9725



Graph.1. Accuracy comparison of Existing and Proposed classifiers for Tomato Diseases.



Graph.2. Comparison metrics of the Existing and Proposed methods

6. Conclusion

The proposed machine learning approach for tomato leaf disease classification demonstrates a significant advancement over traditional methods by integrating robust preprocessing, hybrid segmentation, and comprehensive feature extraction techniques. By standardizing inputs through resizing and grayscale conversion, reducing noise with Gaussian filtering, and employing a hybrid segmentation strategy that combines K-means clustering with Canny Edge Detection, the method effectively isolates disease-relevant regions for analysis. The multi-faceted feature extraction approach, incorporating Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Color Histograms, ensures the capture of diverse disease characteristics, including texture, shape, and color. These enhancements significantly improve classifier performance, with XGBoost achieving an impressive accuracy of 97.25%. The proposed method not only surpasses existing approaches in accuracy and reliability but also provides a scalable framework for application in broader agricultural disease management tasks, contributing to more effective crop protection and sustainable farming practices.

References

1. Baheti, H., Thakare, A., Bhople, Y., Darekar, S., & Dodmani, O. (2022). Machine Learning Algorithm for Detection and Classification of Tomato Plant Leaf Disease. 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), 1–7.
2. Cengil, E., & Cinar, A. (2021). Hybrid convolutional neural network based classification of bacterial, viral, and fungal diseases on tomato leaf images. Concurrency and Computation: Practice and Experience, 34.
3. Ahmed, S., Hasan, M. B., Ahmed, T., Sony, R. K., & Kabir, M. H. (2021). Less is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification. IEEE Access.
4. Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. Applied Artificial Intelligence, 31, 299–315.
5. Jayanthi, G. M., & Shashikumar, D. (2020). Automatic Tomato Plant Leaf Disease Classification using Multi-Kernel Support Vector Machine. International Journal of Engineering and Advanced Technology.
6. Tan, L., Lu, J., & Jiang, H. (2021). Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods. AgriEngineering.
7. Archana, K., & Sahayadhas, A. (2018). Comparison of various filters for noise removal in paddy leaf images. International Journal of Engineering and Technology, 7, 372.
8. Sowmya, M., & Subramani, B. (2022). Superlative Approach for Plant Disease Identification with Enhanced CSA Algorithm. Journal of Mobile Multimedia, 18.
9. NikithaIsaac, V., Sornalatha, J., & Murugan, M. (2021). Smart Irrigation Control With Disease Detection Using Machine Learning and Arduino Controller. IJARSCT, 965, 682–709.
10. Karantoumanis, E., Balafas, V., Louta, M., & Ploskas, N. (2022). Computational comparison of image preprocessing techniques for plant diseases detection. 2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), 1–5.
11. Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture, 74(1), 91–99.
12. Islam, M., Dinh, A. V., Wahid, K. A., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 1–4.
13. Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computers and Electronics in Agriculture, 153, 46–53.
14. Sankaran, S., Mishra, A. R., Ehsani, R., & Davis, C. (2019). A review of advanced techniques for detecting plant diseases. Computers and Electronics in Agriculture, 72(1), 1–13.
15. Khasawneh, N., Faouri, E., & Fraiwan, M. (2022). Detection of tomato diseases using deep transfer learning. Computers and Electronics in Agriculture, 74(1), 1–13.
16. https://github.com/spMohanty/PlantVillage-Dataset.
17. Pujari, J.D., Yakkundimath, R., & Byadgi, A.S. (2015). Image Processing Based Detection of Fungal Diseases in Plants. Procedia Computer Science, 46, 1802-1808.