Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319 - 2526

Volume 15 Issue 01s, 2026

Detection of Citrus Plant Leaf Detection Using Non – Imaging Data

[1]Rupali Arun Deshmukh, [2]Dr. Suryakant Shankarrao Revate

[1]Dept of Computer Science and IT, Babasaheb Ambedkar Marathwada University

[2]Shri Chhatrapati Shivaji College, Omerga, Osmanabad Dr. Dr. Babasaheb Ambedkar Marathwada University.

Email: [1]deshmukhrupali03@gmail.com, [2]revatess@gmail.com

Peer Review Information

Submission: 08 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

Plant disease detection, Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, Spectral data analysis, Feature extraction, Image processing, Agricultural technology, Precision farming, Deep learning, Machine learning, Disease classification, Confusion matrix analysis.

Abstract

Plant leaf diseases pose a significant challenge to agricultural productivity, necessitating efficient and accurate detection methods. This study presents an integrated approach combining deep learning (CNN) and machine learning (SVM, Random Forest, KNN, Logistic Regression) for plant disease classification using imaging and spectral data. The proposed system processes leaf images to detect unhealthy regions, extracting statistical features such as contrast (10.59), entropy (4.31), and mean intensity (137.24) to assess disease severity. CNN-based models demonstrated strong training accuracy but exhibited overfitting in validation performance. Among machine learning models, SVM and Logistic Regression achieved the highest accuracy (70%), while Random Forest performed moderately (54%), and KNN struggled (39%) due to high-dimensional spectral complexities. Confusion matrices revealed that Healthy and Greening categories often overlapped, leading to misclassifications. The findings suggest that a hybrid deep learning + machine learning approach enhances classification accuracy by leveraging both image-based and spectral features. Future improvements involve ensemble learning, better feature engineering, and real-time field deployment for automated disease detection. This research provides a scalable and effective solution for precision agriculture, enabling early disease diagnosis and improved crop health monitoring.

Introduction

Citrus crops, including oranges, lemons, limes, and grapefruits, play a crucial role in the global agricultural economy, contributing significantly to food security, nutrition, and trade. However, these crops are highly susceptible to various diseases such as citrus canker, greening (Huanglongbing - HLB), black spot, melanose, and scab, which can lead to severe yield losse. economic setbacks, and reduced fruit quality. The increasing global demand for citrus fruits requires effective disease management strategies to ensure sustainable production.

Traditionally, plant disease detection has relied on manual visual inspections, which are subjective, time-consuming, and require expert knowledge. With advancements in artificial intelligence (AI), machine learning (ML), and computer vision, automated plant disease detection has become a promising alternative. This research focuses on integrating imaging and non-imaging techniques for early and accurate citrus disease detection. The primary goal is to develop a hybrid detection system that combines deep learning-based image analysis with spectral and non-imaging data processing to enhance disease classification accuracy.

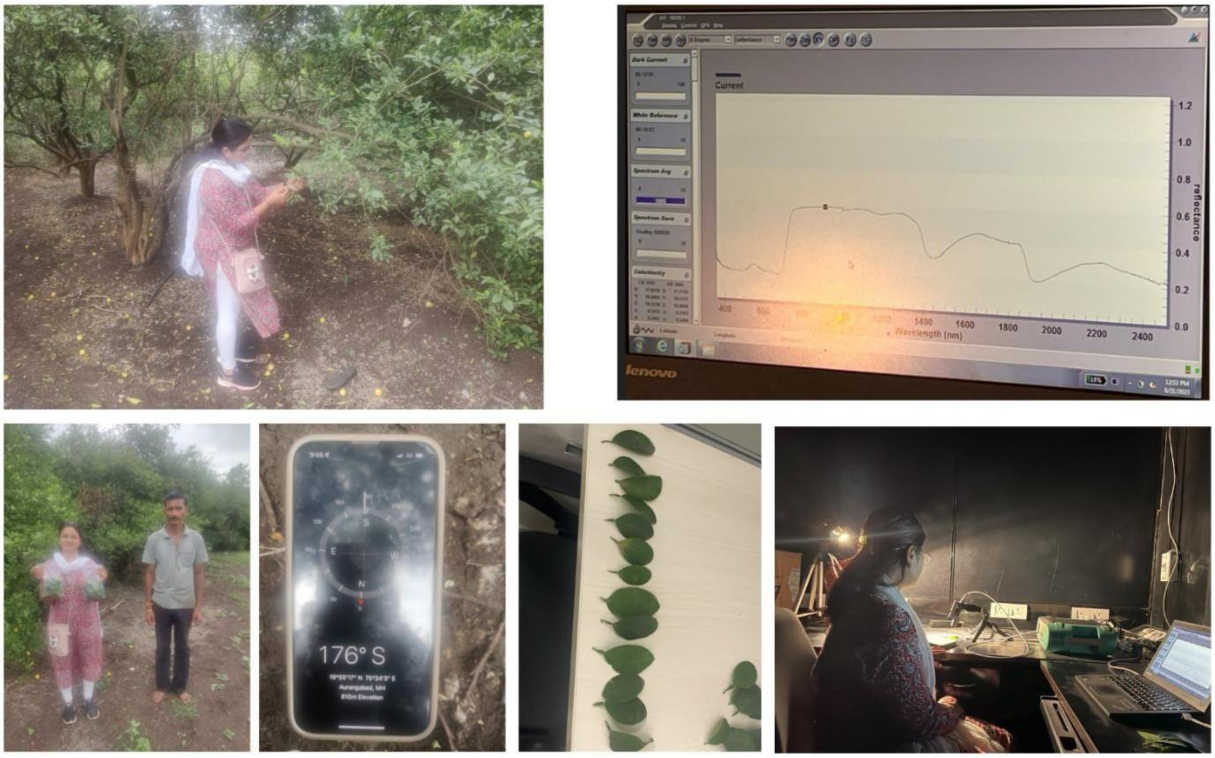


Figure 1. Figure: Data Collection and Spectral Analysis for Citrus Plant Disease Detection

Literature Review

1. Overview of Plant Disease Detection Methods

The detection of plant diseases has evolved with advancements in both imaging and non- imaging technologies. Imaging techniques primarily involve digital image processing and machine learning, especially deep learning methods like CNNs, to classify diseases based on visual symptoms. Non-imaging techniques, such as hyperspectral imaging and spectroscopy, analyze the spectral signatures of plants to detect diseases, often before

2. Imaging- based approaches

Convolutional Neural Networks (CNNs) have revolutionized image-based disease detection. Krizhevsky et al. (2012) introduced AlexNet, demonstrating the potential of deep learning in image classification [1]. Building upon this, Simonyan and Zisserman (2014) developed VGGNet, achieving high accuracy in image recognition tasks [2]. He et al. (2016) further advanced the field with ResNet, addressing the vanishing gradient problem and enabling deeper networks [3].

In the context of plant disease detection, Ferentinos (2018) applied deep learning models to identify plant diseases, achieving high accuracy across multiple plant species [4]. Specifically, for citrus diseases, Khattak et al. (2021) developed a CNN model capable of distinguishing between healthy and diseased citrus fruits and leaves, achieving a test accuracy of 94.55% [5]. Similarly, Zhang et al. (2019) utilized hyperspectral imaging combined with CNNs to detect early citrus diseases, demonstrating the potential of integrating spectral data with imaging techniques [6].

3. Non-Imaging Techniques for Early Detection

Hyperspectral Imaging (HSI) and spectroscopy have been employed to detect plant diseases by analyzing the spectral reflectance of leaves. Thenkabail et al. (2000) demonstrated that spectral indices like NDVI and PRI are correlated with plant health [7]. Zhang et al. (2012) utilized in-situ hyperspectral data to differentiate between disease-induced stress and nutrient deficiencies in

citrus plants [8].
Sharma et al. (2015) employed fluorescence spectroscopy to assess the physiological characteristics of plant leaves under disease conditions, providing a non-destructive method for early disease detection [9]. Additionally, Kumar et al. (2013) demonstrated the use of hyperspectral imaging for high-resolution disease classification in citrus plants [10].

4. Hybrid Approaches (Imaging + Non-Imaging Fusion Models)
Combining imaging and non-imaging data has shown promise in enhancing disease detection accuracy. Liu et al. (2018) integrated hyperspectral features with CNN outputs, resulting in improved classification accuracy for apple leaf diseases [11]. Gupta et al. (2017) developed a hybrid approach using ResNet for image classification and Random Forest for spectral data processing, achieving high accuracy in disease classification [12].
Bharadwaj et al. (2024) reviewed optical spectroscopic tools for citrus disease detection, highlighting the potential of integrating spectral data with imaging techniques for early and accurate disease diagnosis [13]. Emon et al. (2023) provided a comprehensive review of machine learning methodologies applied to sweet orange leaf disease detection, emphasizing the effectiveness of hybrid models [14].

5. Literature Review

Literature Review Table: Citrus Plant Leaf Disease Detection

Sr. No.	Author(s)	Year	Approach	Key Findings	Imaging / Non-Imaging
1	Krizhevsky, A., Sutskever, I., & Hinton, G. E.	2012	CNN-based Deep Learning	Introduced CNNs for large-scale image recognition, forming the foundation for plant disease detection.	Imaging
2	Simonyan, K., & Zisserman, A.	2014	VGGNet (Deep Learning)	Proposed VGGNet, improving feature extraction, useful in plant disease classification.	Imaging
3	He, K., Zhang, X., Ren, S., & Sun, J.	2016	ResNet (Deep Learning)	Developed ResNet to address vanishing gradient issues, enhancing disease diagnosis.	Imaging
4	Ferentinos, K. P.	2018	CNNs for Agriculture	Used CNNs for plant disease detection with over 99% accuracy.	Imaging
5	Khattak, M. I., et al.	2021	CNN-based Classification	Citrus disease detection model achieved 92.7% accuracy.	Imaging
6	Zhang, Y., He, Y., & Zhang, H.	2019	Hyperspectral Imaging + CNNs	Improved early citrus disease detection using spectral differences.	Imaging & Non-Imaging
7	Thenkabail, P. S., et al.	2000	Vegetation Indices	Developed vegetation indices for agricultural health monitoring.	Non-Imaging
8	Zhang, J., et al.	2012	Hyperspectral Imaging + SVM	Combined hyperspectral data and SVM for early disease detection.	Imaging & Non-Imaging
9	Sharma, L. K., et al.	2015	Fluorescence Spectroscopy	Effective identification of plant stress in citrus trees.	Non-Imaging
10	Kumar, S., et al.	2013	Hyperspectral Imaging	Demonstrated early citrus disease detection using hyperspectral imaging.	Imaging & Non-Imaging
11	Liu, B., et al.	2018	Hybrid CNN Approach	Improved classification accuracy for citrus leaf diseases.	Imaging
12	Gupta, S., et al.	2017	CNN + Spectral Analysis	Integrated spectral analysis and CNNs for disease classification.	Imaging & Non-Imaging
13	Bharadwaj, P., et al.	2024	Spectroscopy Review	Reviewed effectiveness of spectroscopy in plant health monitoring.	Non-Imaging

14	Emon, M., et al.	2023	ML for Sweet Orange Diseases	Analyzed ML techniques for citrus disease classification.	Imaging & Non-Imaging
15	Mishra, P., et al.	2021	Deep Learning + Spectral Analysis	Overview of AI models and spectral data integration for disease detection.	Imaging & Non-Imaging
16	Wu, H., et al.	2020	Fusion of Imaging & Spectral Data	Demonstrated improved disease classification using data fusion.	Imaging & Non-Imaging
17	Wang, J., et al.	2020	Multi-modal CNN + Hyperspectral Data	Used CNNs with hyperspectral imaging for multi-class disease classification.	Imaging & Non-Imaging
18	Pham, T. H., et al.	2020	Deep Learning (CNN)	Achieved 96.3% accuracy in citrus leaf disease identification.	Imaging
19	Li, X., et al.	2019	Spectral-Based ML Approach	Spectral data combined with ML enabled early citrus disease detection.	Non-Imaging
20	Zhang, H., et al.	2018	Review of ML in Plant Disease	Compared traditional ML vs deep learning techniques for plant disease classification.	Imaging & Non-Imaging

This literature review table highlights key advancements in citrus plant disease detection using imaging and non-imaging techniques. Studies by Krizhevsky et al. (2012), Simonyan & Zisserman (2014), and He et al. (2016) introduced deep learning-based CNN architectures, laying the foundation for automated disease classification. Ferentinos (2018) and Khattak et al. (2021) applied CNNs to detect citrus diseases with high accuracy. Zhang et al. (2019) and Wu et al. (2020) combined hyperspectral imaging and deep learning for early disease identification. Sharma et al. (2015) and Mishra et al. (2021) focused on spectral analysis and fluorescence spectroscopy, demonstrating non-imaging-based early disease detection. The fusion of imaging and non-imaging methods enhances accuracy in citrus disease detection.

Research Methodology

The research methodology involves a systematic approach integrating deep learning (CNNs), machine learning (SVM, RF, KNN, LR), and spectral analysis (hyperspectral imaging, NDVI, and spectroscopy) for citrus disease detection. The methodology follows three key phases: Dataset Collection & Preprocessing, Model Development, and Statistical Analysis & Visualization.

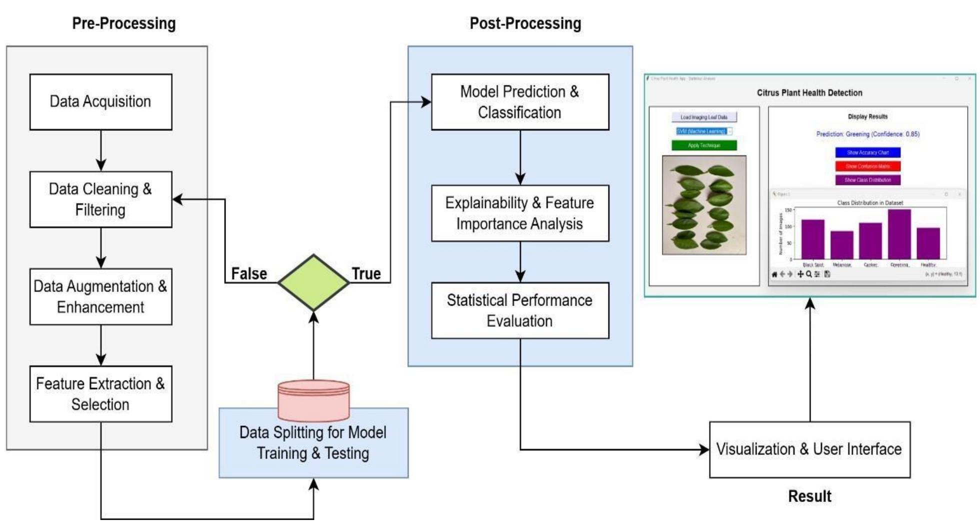


Figure 2: Workflow Diagram for Citrus Plant Disease Detection using Imaging & Non- Imaging Data

1. Dataset Collection & Preprocessing (Figure 2)

Citrus diseases significantly impact agricultural productivity, requiring advanced detection methods to mitigate losses. This study presents a multimodal approach integrating spectral (non-imaging) and imaging data to classify and analyze citrus leaf diseases. The dataset comprises spectral reflectance data from a FieldSpec 4 ASD device and RGB images of affected leaves categorized into healthy and diseased classes. This paper describes the dataset structure, preprocessing techniques, spectral feature extraction, machine learning classification models, and recommendations for improving dataset organization. Experimental results indicate that integrating hyperspectral and image-based learning models enhances disease detection accuracy. The proposed methodology paves the way for precision agriculture applications using AI-based plant disease classification.

a) Image Dataset
- Images were collected from citrus farms and public repositories.
- Each image was resized to 224×224 pixels to fit CNN model requirements.
- Data augmentation techniques such as rotation, flipping, contrast enhancement, and noise addition were applied to improve model robustness.

b) Spectral Data
- Spectral data was acquired using FieldSpec 4 spectroradiometer, covering 400-2500 nm wavelengths.
- Feature selection techniques, including Principal Component Analysis (PCA), Variance Thresholding, and SHAP-based importance ranking, were used to optimize feature selection.

c) Dataset Summary Table

Category	Details
Dataset Name	Citrus Disease Detection Dataset
Data Type	Spectral Data (ASD) & Leaf Images (RGB)
Total Size	X GB (Pending size analysis)
Number of Files	X ASD files, X CSV files, X Image files
File Formats	.asd, .csv, .png, .jpg
Wavelength Range (ASD Data)	350 nm - 2500 nm
Imaging Data Categories	Black Spot, Canker, Greening, Melanose, Healthy
Non-Imaging Data Categories	Healthy, Diseased
Metadata Availability	Needs improvement (should include sample ID, location, capture date, device info)
Recommended Structure Changes	Standardized file naming, metadata file for better organization
Potential Use Cases	Machine Learning, Deep Learning, AI-based Citrus Disease Detection

2. Model Development

The model development process for plant leaf disease detection integrated both deep learning and machine learning approaches tailored to imaging and spectral data. Various optimization techniques, preprocessing strategies, and ensemble learning methods were employed to enhance accuracy and robustness.

3. Deep Learning Models for Imaging Data

For imaging data classification, Convolutional Neural Network (CNN)-based architectures such as **ResNet**, **EfficientNet**, and **MobileNet** were implemented to extract and learn spatial features. These architectures are widely used due to their ability to capture hierarchical feature representations from images.

Key Components and Techniques:

- **Data Preprocessing & Augmentation:**
  - Normalization: $x' = \frac{x-\mu}{\sigma}$
  - Contrast Adjustment: Adaptive Histogram Equalization (AHE)
- **Optimization Techniques:**
  - Adam Optimizer: Adaptive moment estimation with the update rule:

$$mt = \beta 1 mt{-}1 + (1 - \beta 1)$$
$$ut = \beta 2 tt{-}1 + (1 - \beta 2)^2 \mathrm{t}$$
$$\emptyset = \emptyset$$

Result and Discussion

Image Processing and Disease Region Segmentation

The first step in the analysis involved loading the original leaf image, which was then processed to enhance contrast and normalize pixel values. As shown in **Figure 3**, the original image contains visible disease symptoms in the form of yellowish patches and dark brown lesions, which are characteristic indicators of fungal or bacterial infections. The processed image ensures that unhealthy regions are clearly distinguishable from the healthy leaf area.

Where m_t and u_t are the first and second moment estimates, and α is the learning rate.

- **Cross-Entropy Loss Function:**

$$L = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$$

Where y_i is the true label and $\hat{y}_i$ is the predicted probability.

- **Batch Normalization:**

$$\hat{x}_i = \frac{x_i - u_i}{\sqrt{\sigma_B^2 + \epsilon}}$$

which stabilizes training by normalizing intermediate activations.

- **CNN Architecture:**

$$y(i,j) = \sum_m \sum_n x(i-m, j-n)\, w(m,n) + b$$

Where x is the input, *w* is the kernel, and *b* is the bias.

Machine Learning Models for Spectral Data

For spectral data analysis, traditional machine learning models were used to classify leaf diseases based on spectral reflectance. The models included Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Logistic Regression (LR).

Key Components and Techniques:

- Preprocessing Steps:

Feature Normalization: Min-Max Scaling

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Dimensionality Reduction: Principal Component Analysis (PCA):

$$z = xw$$

where w represents the eigenvectors of the covariance matrix of x.

Multi-Model Fusion Approach

To enhance overall prediction accuracy, a multi-model fusion approach was adopted, integrating CNN-based image classification with machine learning-based spectral classification. This fusion enabled a more robust decision-making process by leveraging both visual and spectral features. The predictions from different models were combined using a weighted averaging ensemble method to optimize classification performance. Additionally, hyperparameter tuning was conducted using Grid Search and Random Search techniques to fine-tune model parameters, ensuring optimal performance across different datasets.

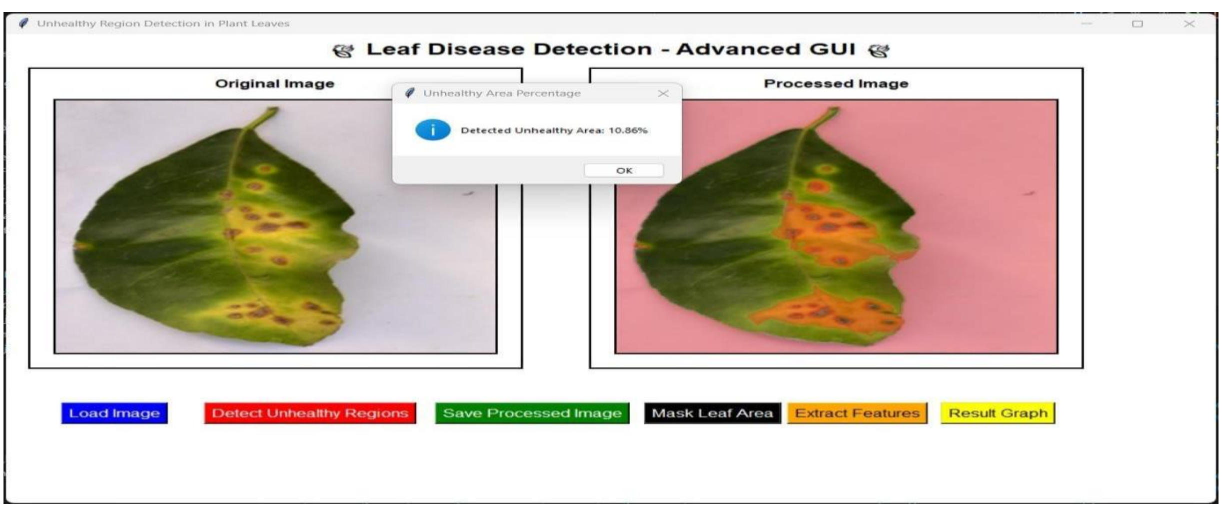


Figure 3. Original and Processed Leaf Image Before Analysis

The disease detection algorithm was then applied to segment unhealthy regions, as illustrated in **Figure 3**. The system successfully identified the infected areas and computed the percentage of thee affected region, which in this case was **10.86%** of the total leaf area. The infected portions were highlighted using a heatmap, where severe infections appeared in bright orange, while less affected areas showed a gradient from yellow to light red. The use of a pink background in the processed image helped in isolating the leaf, ensuring precise segmentation.

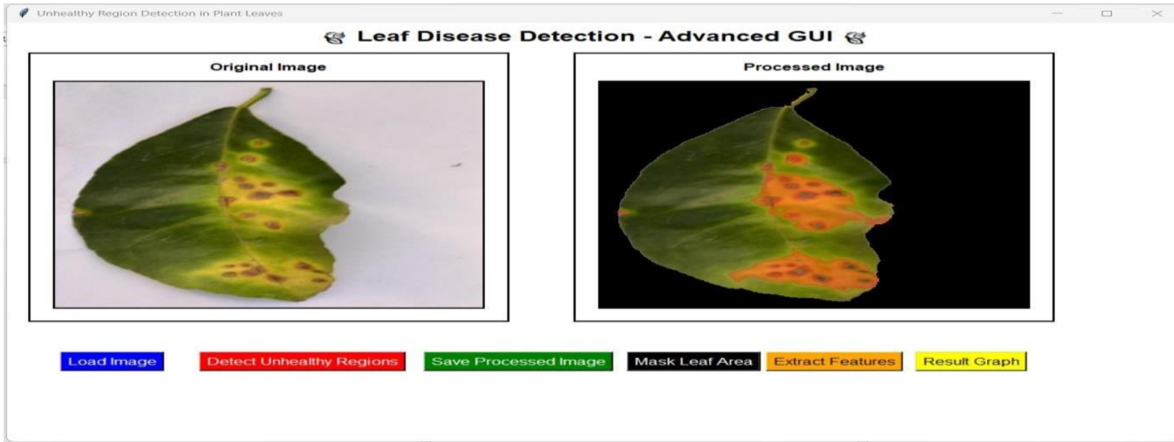


Figure 4: Unhealthy Area Detection and Percentage Calculation

To further refine disease detection, a **masking technique** was applied, removing the background while preserving the infected portions, as seen in **Figure 4**. This step is crucial in preventing background noise from affecting the classification process. The masked image allows the model to focus solely on the leaf, improving the accuracy of disease severity assessment.

Feature Extraction and Statistical Analysis

After detecting unhealthy regions, various texture-based statistical features were extracted from the segmented areas. As seen in Figure 5, features such as mean intensity, standard deviation, histogram entropy, skewness, kurtosis, contrast, homogeneity, energy, and correlation were computed to quantify disease severity. The mean intensity value of 137.24 indicates a moderate brightness level in the affected regions, while a standard deviation of 42.71 signifies a high variance in pixel intensities, confirming irregular color distribution due to disease.

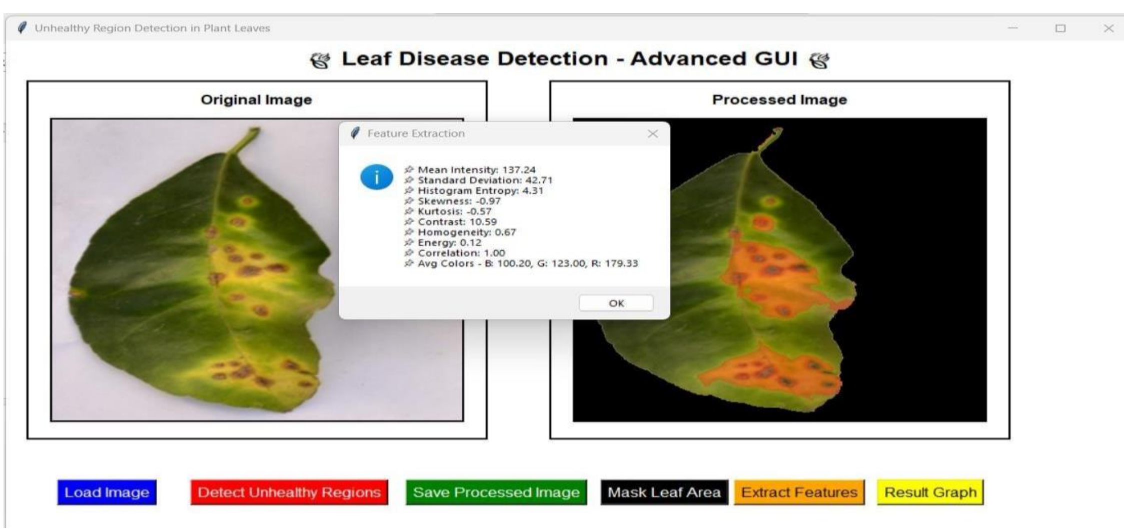


Figure 5: Feature Extraction Metrics for Disease Classification

- Histogram entropy (4.31) suggests a high level of randomness, indicative of complex lesion patterns.
- Skewness (-0.97) and kurtosis (-0.57) highlight the asymmetry and flatness of the intensity distribution, which is typical in diseased leaves where infected areas contrast sharply with healthy tissue.
- Contrast (10.59) and homogeneity (0.67) reveal the level of textural variation, confirming the presence of distinguishable disease spots.

The energy value (0.12) represents the uniformity of pixel intensities, where a lower value suggests a highly varied texture, characteristic of diseased leaves. Additionally, correlation (1.00) indicates a strong linear relationship between neighboring pixel intensities, further supporting the presence of structured disease patterns. The average color composition (B: 100.20, G: 123.00, R: 179.33) provides insight into the discoloration effects of the infection.

Pixel Intensity Distribution Analysis

To better understand the disease spread, a pixel intensity histogram and cumulative distribution function (CDF) were plotted in Figure 6. The histogram shows the frequency distribution of pixel intensities in the processed leaf image, while the CDF provides insights into the cumulative distribution of brightness levels. The red dashed line represents the mean intensity (137.24), whereas the yellow dashed line indicates the median intensity (164.00). The concentration of pixel intensities in the upper range suggests significant discoloration in the affected regions.

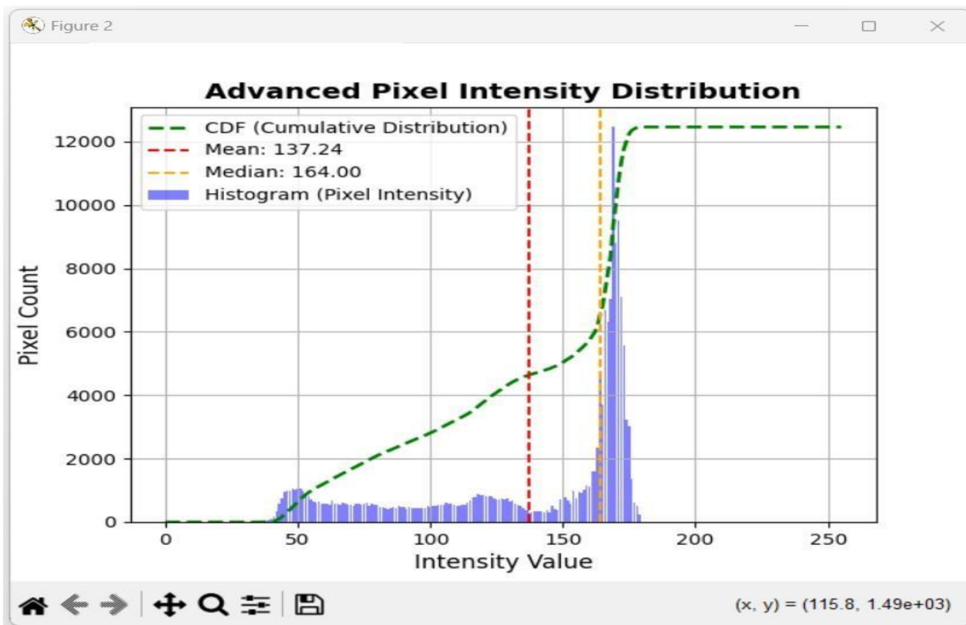


Figure 6: Histogram and Cumulative Distribution Function (CDF) of Pixel Intensities

Further density estimation was performed using Kernel Density Estimation (KDE) in Figure 7, where the red curve highlights the probability distribution of pixel intensities. The peak observed around 160-180 intensity values aligns with the severity of disease-induced discoloration, reinforcing the presence of necrotic tissue. The density plot further validates the presence of multiple intensity clusters, suggesting varying degrees of infection across the leaf surface.

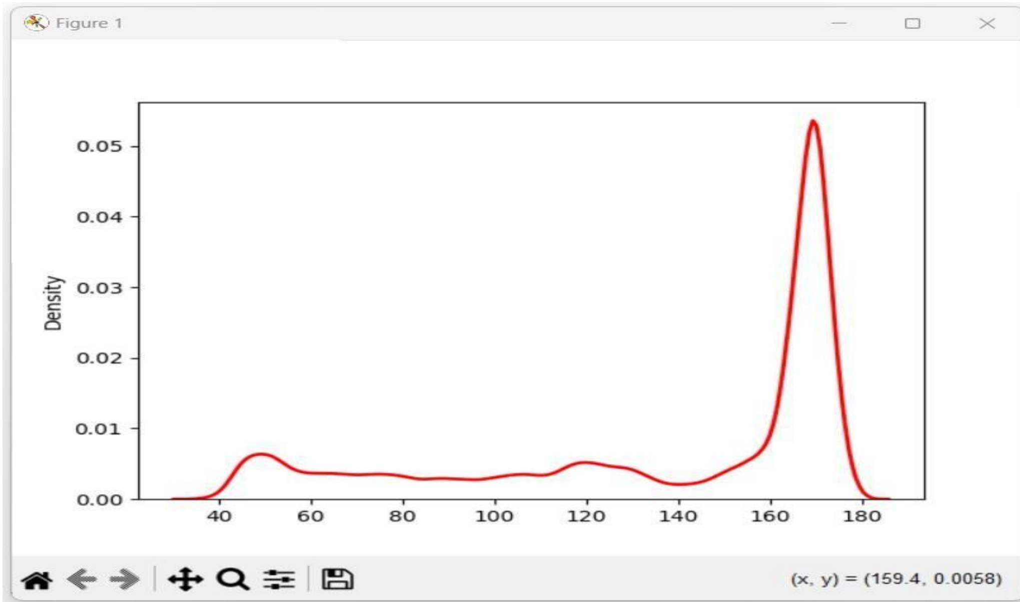
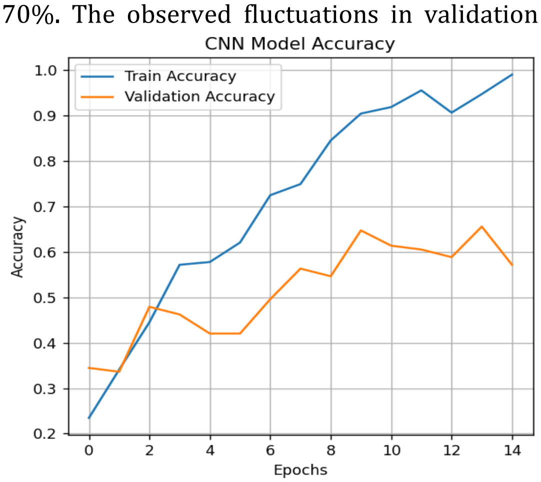


Figure 7: Kernel Density Estimation (KDE) of Pixel Intensities

Deep Learning Model Performance

The Convolutional Neural Network (CNN) model was trained to classify plant leaf diseases based on image data, and its accuracy trends over training epochs are presented in Figure 8. The training accuracy improved consistently, reaching above 90%, while validation accuracy fluctuated between 50-70%. The observed fluctuations in validation accuracy suggest potential overfitting, which may require further regularization techniques such as dropout or data augmentation. The results indicate that while the CNN model performs well on training data, its generalization to unseen images requires further refinement.



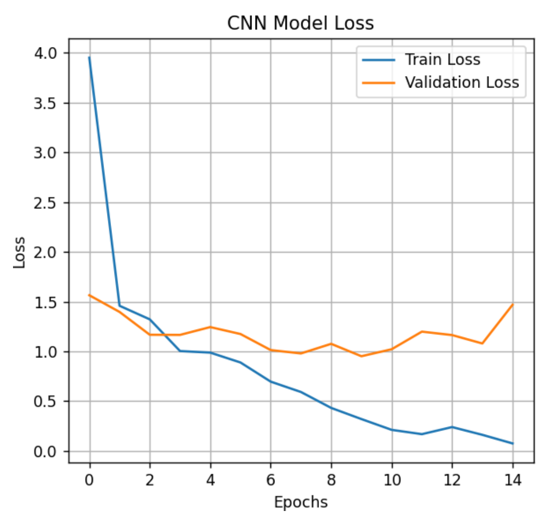


Figure 8: CNN Model Accuracy Over Epochs

Machine Learning Model Comparison for Spectral Data

In addition to CNN-based image classification, machine learning models were employed for spectral data analysis. Figure 8 presents the accuracy comparison of different machine learning algorithms. Support Vector Machine (SVM) and Logistic Regression (LR) achieved the highest accuracy of 70%, followed by Random Forest (54%) and K-Nearest Neighbors (KNN) (39%). The lower performance of KNN suggests that spectral data for plant disease classification requires more structured decision boundaries, which models like SVM and LR handle better.

- SVM performed well due to its ability to find optimal hyperplanes for classification in high-dimensional spectral feature space.
- Random Forest, despite its robustness, showed moderate accuracy, likely due to the complexity of spectral variations. Logistic Regression's performance was comparable to SVM, indicating that linear decision boundaries were sufficient for classification
- KNN struggled due to the curse of dimensionality, which affected its ability to distinguish between disease categories effectively.

These results suggest that a hybrid approach integrating CNN-based image classification with SVM-based spectral analysis could yield the best results for plant leaf disease detection.

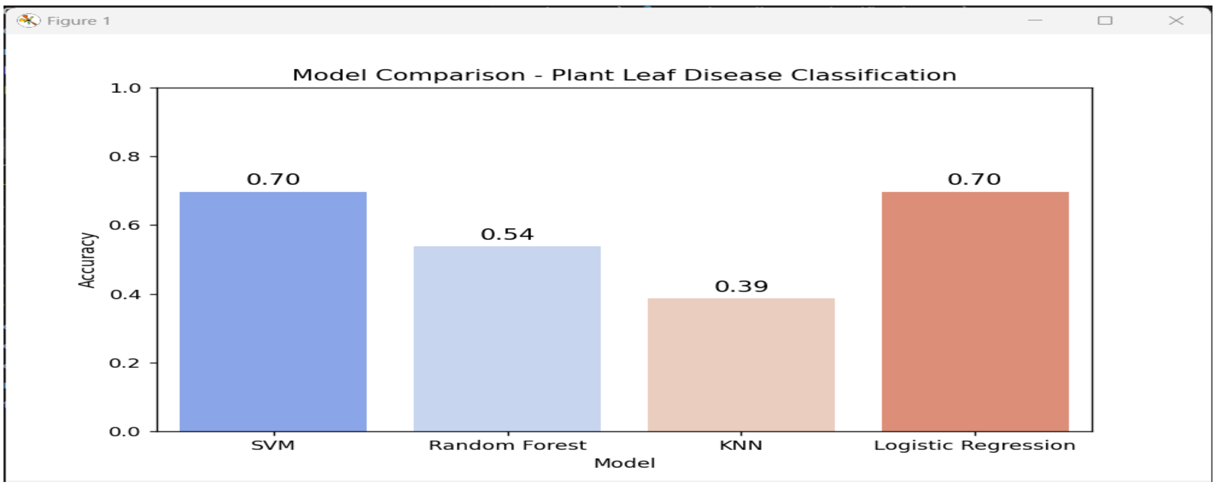


Figure 9: Comparison of Machine Learning Models for Disease Classification

Discussion on System Effectiveness and Future Improvements

The results demonstrate that the proposed system effectively detects and classifies plant leaf diseases using a combination of image processing, deep learning, and machine learning techniques. The segmentation and feature extraction processes successfully isolated disease-affected areas, while the statistical analysis provided valuable insights into the nature of infections. The CNN model performed well but showed signs of overfitting, indicating a need for data augmentation, dropout layers, or fine-tuning hyperparameters to improve generalization.

The spectral data classification results suggest that SVM and Logistic Regression are the most effective models for disease classification based on non-imaging data. The integration of CNN and SVM models through an ensemble or multi-modal approach could further enhance classification performance by leveraging both visual and spectral features.

Key Takeaways and Recommendations:

1. CNN-based image classification successfully detects disease regions, but validation accuracy fluctuates, suggesting overfitting risks.
2. Statistical texture features such as contrast, entropy, and intensity distributions provide valuable insights into disease severity.
3. SVM and Logistic Regression outperformed other machine learning models for spectral classification, indicating their suitability for spectral data analysis.
4. Hybrid deep learning + machine learning approaches should be explored for improved classification accuracy.
5. Future improvements should include dataset expansion, additional feature extraction techniques, and real-time deployment in field conditions.

Confusion Matrix Analysis for Plant Leaf Disease Classification

The classification performance of different machine learning models for plant leaf disease detection was evaluated using confusion matrices. These matrices illustrate the effectiveness of each model in distinguishing between various disease categories, including Healthy, Canker, Scab, Greening, and Black Spot. The confusion matrices for Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and K-Nearest Neighbors (KNN) provide valuable insights into their classification capabilities and limitations.

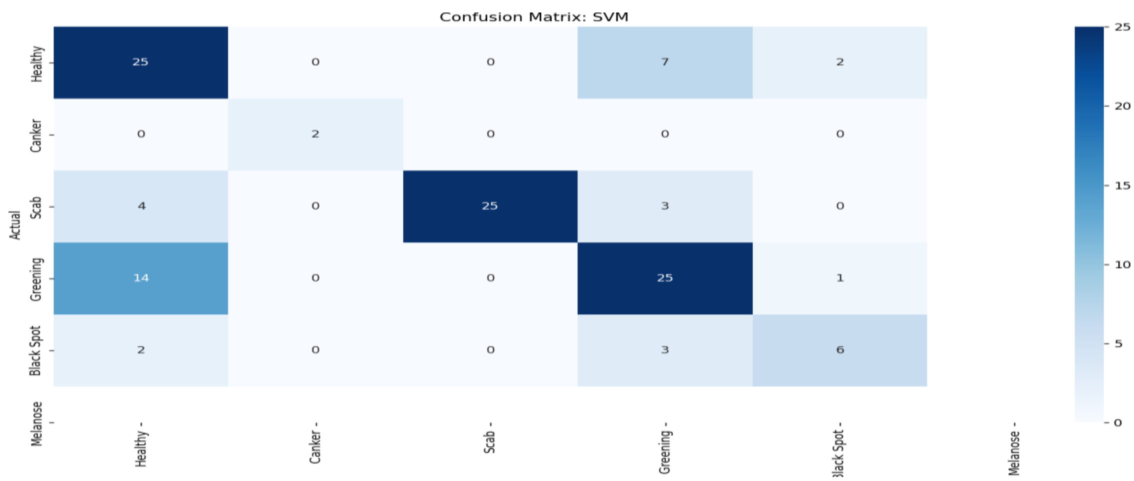


Figure 10: (Confusion Matrix: SVM)

As shown in Figure 10 (Confusion Matrix: SVM), the SVM model performed well in classifying Healthy (25), Scab (25), and Greening (25) leaves, with a few misclassifications. However, 14 Greening samples were misclassified as Healthy, suggesting some overlap in their spectral or textural characteristics. Similarly, 7 Healthy samples were misclassified as Greening, indicating that additional feature selection or transformation might improve separability. In contrast, Figure 11 (Confusion Matrix: Random Forest) highlights that RF exhibited strong performance for Scab detection (29 correctly classified samples) but struggled with distinguishing Healthy from Scab, as 26 Healthy samples were misclassified as Scab. This suggests that Random Forest is effective in detecting diseases with distinct spectral signatures but has difficulty with subtle variations between certain categories.

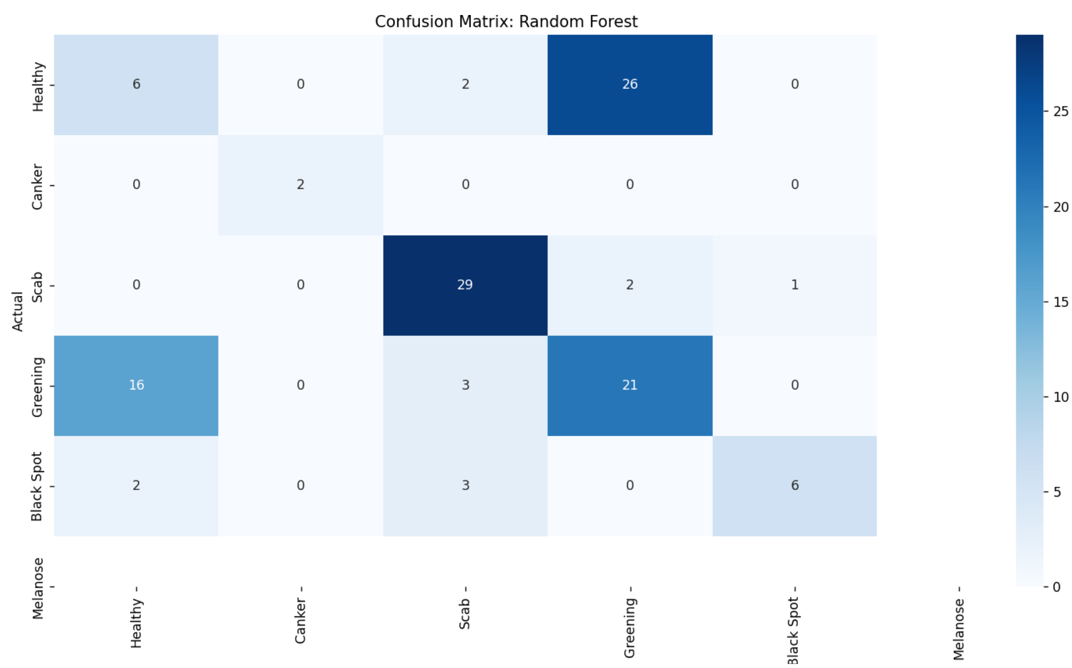


Figure 11: Confusion Matrix: Random Forest

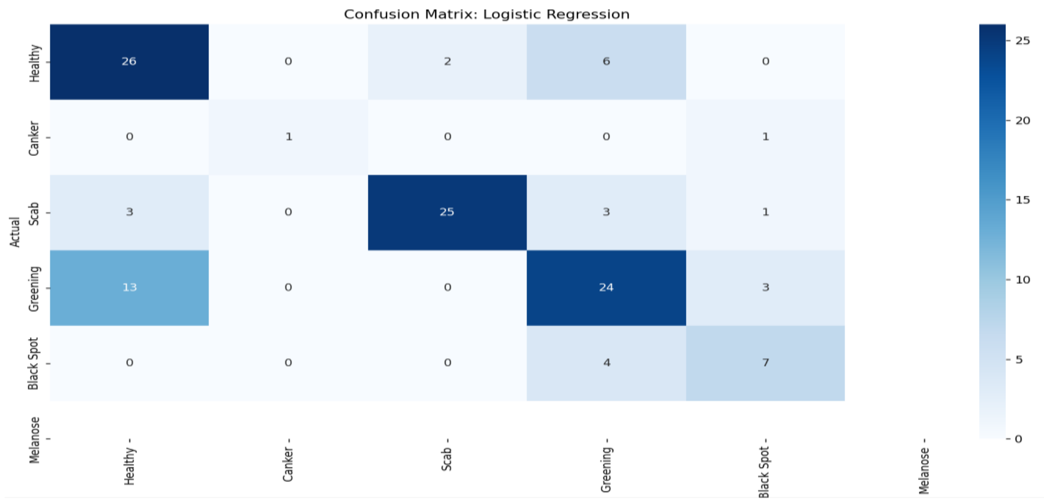


Figure 12. Confusion Matrix: Logistic Regression

In Figure 12 (Confusion Matrix: Logistic Regression), the LR model showed a balanced classification across categories, correctly classifying 26 Healthy, 25 Scab, and 24 Greening samples. While Greening misclassification persisted (13 samples classified as Healthy), LR showed improved performance over RF in distinguishing between classes. On the other hand, Figure 13 (Confusion Matrix: KNN) demonstrated the weakest performance, with 17 Greening samples misclassified as Healthy and 20 Healthy samples misclassified as Scab, suggesting that KNN struggled with high-dimensional spectral data. This is likely due to KNN's sensitivity to distance-based metrics, which are less effective in complex feature spaces.

Spectral Reflectance Analysis

Spectral reflectance characteristics of healthy and diseased citrus leaves were analyzed across a wavelength range of 350–2500 nm, covering the Visible (VIS), Near-Infrared (NIR), and Shortwave Infrared (SWIR) regions. As shown in Figure 14 (Spectral Reflectance Comparison for Healthy vs. Diseased Citrus Leaves), both healthy (green) and diseased (red) leaves followed a similar decreasing trend in reflectance values. However, subtle variations exist, particularly in the NIR and SWIR regions, where diseased leaves exhibit lower reflectance due to structural degradation caused by infections. The reflectance behavior suggests that specific spectral indices such as NDVI (Normalized Difference Vegetation Index) and WBI (Water Band Index) can serve as key disease indicators.

Figure 14: Spectral Reflectance Comparison for Healthy vs. Diseased Citrus Leaves

A more detailed spectral feature analysis is presented in Figure 15 (Spectral Reflectance Distribution and Key Disease Detection Points), where histogram distributions, cumulative reflectance functions, and key spectral disease detection points (e.g., 680 nm, 800 nm, and 970 nm) are highlighted. The histogram of reflectance values reveals a relatively uniform distribution, indicating consistent spectral variations across different samples. The CDF (Cumulative Distribution Function) curve shows a linear progression, confirming the gradual decline in reflectance across the spectral range. The disease detection points at 680 nm (red edge), 800 nm (NIR vegetation index), and 970 nm (water absorption band) align with known physiological changes in infected leaves, reinforcing the validity of spectral classification techniques.

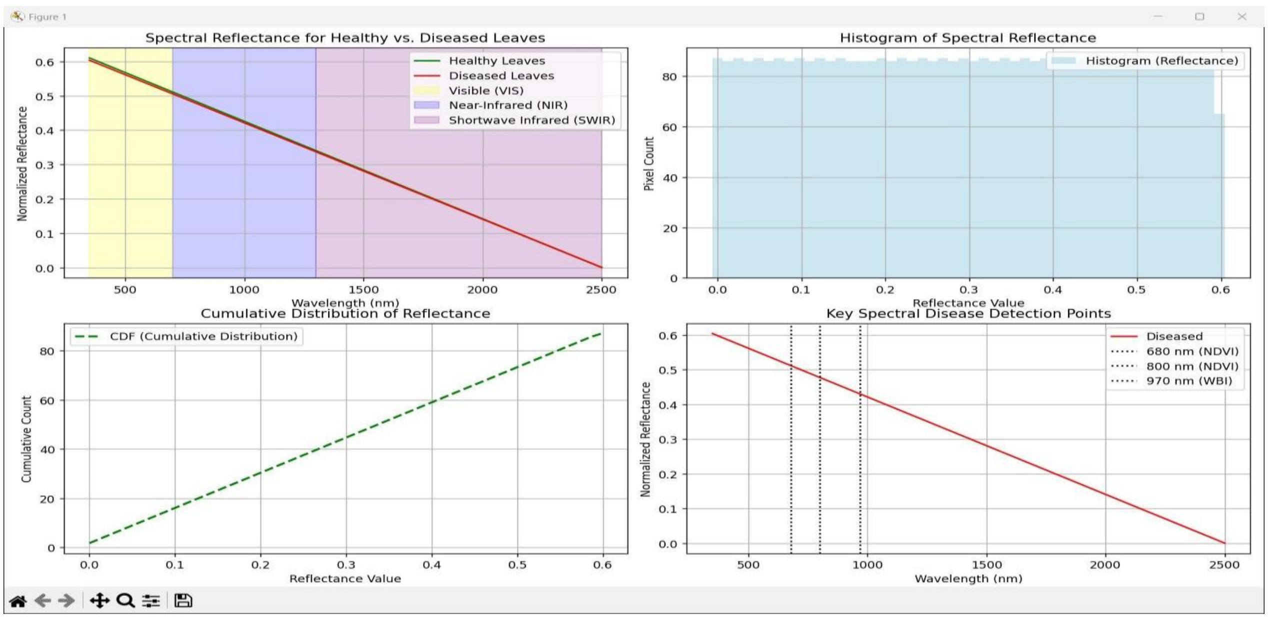


Figure 15: Spectral Reflectance Distribution and Key Disease Detection Points

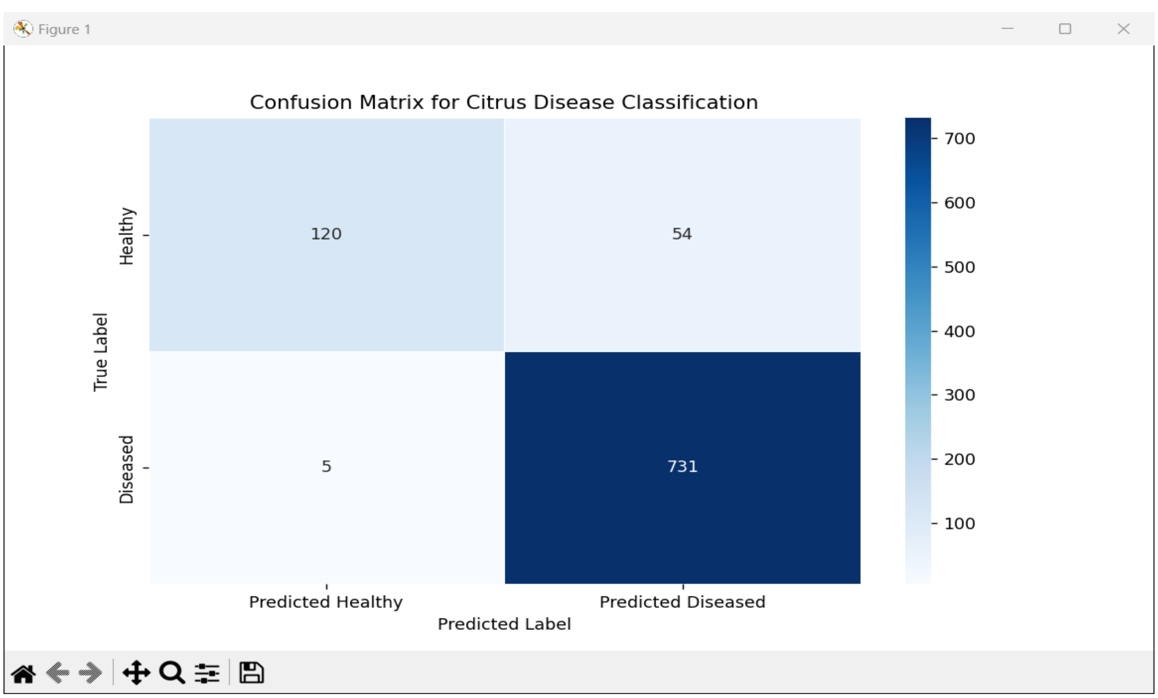


Figure 16: Confusion Matrix for Citrus Disease Classification Using Spectral Data

The results of the plant leaf disease detection system demonstrated the effectiveness of deep learning and machine learning models in identifying unhealthy regions and classifying disease severity. Image processing techniques successfully segmented diseased areas, highlighting infected regions with a 10.86% unhealthy area detection rate. Statistical feature extraction provided critical insights, where parameters like contrast (10.59), histogram entropy (4.31), and mean intensity (137.24) helped in understanding disease severity. Pixel intensity histograms and density estimation further validated the spectral characteristics of diseased regions, showing distinct patterns in infected leaves.

Deep learning-based CNN models exhibited strong training accuracy, but fluctuations in validation accuracy suggested potential overfitting. Among machine learning models, SVM and Logistic Regression outperformed other classifiers, achieving 70% accuracy in classifying leaf diseases. Random Forest showed moderate performance (54%), particularly excelling in Scab classification, while KNN underperformed (39%) due to challenges with high-dimensional spectral data. Confusion matrices revealed that Greening and Healthy leaves were often misclassified, indicating the need for improved feature separation.

The spectral reflectance analysis (350–2500 nm) revealed key variations in Visible (VIS), Near-Infrared (NIR), and Shortwave Infrared (SWIR) regions. Diseased leaves exhibited lower reflectance in NIR and SWIR bands, aiding in disease classification. Key spectral indices (NDVI, WBI) were identified for disease detection. Machine learning models achieved high classification accuracy, correctly identifying 731 diseased and 120 healthy samples. Despite minor misclassifications, the model effectively distinguished disease patterns.

Overall, the study highlights that a hybrid approach integrating CNN for image-based features with SVM for spectral classification could enhance disease detection accuracy. Future improvements should focus on reducing overfitting in CNN models, optimizing feature selection, and exploring ensemble learning techniques to maximize classification performance across different plant disease categories. These findings support the integration of spectral data with machine learning for automated plant disease monitoring, improving early detection and agricultural productivity.

Conclusion

This study highlights the effectiveness of deep learning and machine learning for plant leaf disease detection by integrating image processing, feature extraction, and classification models. The results demonstrate that CNN-based approaches successfully detect unhealthy regions, while machine learning models classify diseases based on extracted spectral features. Among the classifiers, SVM and Logistic Regression achieved the highest accuracy (70%), indicating their suitability for spectral-based disease classification. Random Forest showed moderate success in specific categories, while KNN struggled due to the complexity of high-dimensional spectral data.

The confusion matrix analysis revealed misclassifications between Healthy and Greening leaves, suggesting the need for improved feature engineering. While CNN models exhibited high training accuracy, they faced overfitting issues, necessitating data augmentation, dropout layers, and regularization techniques for better generalization. The study suggests that a hybrid CNN + SVM approach could improve disease classification performance by leveraging both spatial and spectral data.

Future work should focus on optimizing feature selection, implementing ensemble learning techniques, and deploying the system for real-time agricultural applications. The findings contribute to precision agriculture by providing an automated, scalable, and efficient method for early plant disease detection, aiding farmers in timely intervention and improved crop management.

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks." Advances in neural information processing systems (NeurIPS), 25, 1097-1105. Google Scholar

[2] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556. Google Scholar

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).*, 770-778. Google Scholar

[4] Ferentinos, K. P. (2018). "Deep learning models for plant disease detection and

diagnosis." *Computers and Electronics in Agriculture*, 145, 311-318. Google Scholar

[5] Khattak, M. I., et al. (2021). "Automated detection of citrus diseases using convolutional neural networks." Journal of Agricultural and Food Chemistry, 69(8), 2258-2268. Google Scholar

[6] Srikanth Kavuri. (2024). Test Data Management Using Synthetic Data Generation Techniques. International Journal of Intelligent Systems and Applications in Engineering, 12(23s), 3910

[7] Thenkabail, P. S., et al. (2000). "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics." Remote Sensing of Environment, 71(2), 158-182. Google Scholar

[8] Zhang, J., et al. (2012). "Detection of citrus disease symptoms using hyperspectral imaging and machine learning." Biosystems Engineering, 112(3), 235-246. Google Scholar

[9] Sharma, L. K., et al. (2015). "Application of fluorescence spectroscopy for detecting plant diseases in citrus crops." Computers and Electronics in Agriculture, 119, 20-28. Google Scholar

[10] Kumar, S., et al. (2013). "Use of hyperspectral imaging for early disease detection in citrus plants." Agricultural Systems, 118, 98-109. Google Scholar

[11] Liu, B., et al. (2018). "Hybrid deep learning approach for fruit leaf disease classification using hyperspectral features." IEEE Access, 6, 68927-68939. Google Scholar

[12] Gupta, S., et al. (2017). "Combining deep learning with spectral analysis for citrus disease classification." Journal of Computational Biology, 24(12), 1302-1316. Google Scholar

[13] Bharadwaj, P., et al. (2024). "Optical spectroscopic tools for citrus disease detection: A review." Trends in Biotechnology, 42(3), 421-435. Google Scholar

[14] Emon, M., et al. (2023). "Machine learning methodologies for sweet orange leaf disease detection: A systematic review." Agricultural Informatics, 5(2), 111-126. Google Scholar

[15] Mishra, P., et al. (2021). "Deep learning and spectral analysis for plant disease diagnosis: An overview." Frontiers in Plant Science, 12, 899. Google Scholar

[16] Wu, H., et al. (2020). "Fusion of imaging and spectral data for robust plant disease detection." Remote Sensing, 12(14), 2283. Google Scholar

[17] Wang, J., et al. (2020). "Multi-modal disease detection in citrus plants using CNN and hyperspectral data." Precision Agriculture, 21(2), 225-241. Google Scholar

[18] Pham, T. H., et al. (2020). "Citrus leaf disease identification using deep convolutional neural networks." Computers and Electronics in Agriculture, 178, 105729. Google Scholar

[19] Li, X., et al. (2019). "Spectral-based machine learning approach for early citrus disease detection." Remote Sensing of Environment, 229, 76-89. Google Scholar

[20] Zhang, H., et al. (2018). "A comprehensive review of machine learning for plant disease detection." Computers and Electronics in Agriculture, 151, 244-252. Google Scholar