

ORIGINAL ARTICLE

Extreme gradient boosting driven intelligent system for tomato leaf disease identification

Vijayanandh Rajamanickam^{1*} , Deepak Ramsubhag¹ , Jayaraj Jayaraman² ¹ Department of Computing and Information Technology, The University of the West Indies, St Augustine, 330912, St Augustine, Trinidad and Tobago² Department of Life Sciences, The University of the West Indies, St Augustine, 330912, St Augustine, Trinidad and Tobago

Vol. 66, No. 3: 398–408, 2026

DOI: 10.24425/jppr.2026.3684

Received: May 23, 2025

Accepted: September 10, 2025

Online publication: September 18, 2026

*Corresponding address:

Vijayanandh.Rajamanickam@sta.uwi.edu

Responsible Editor:

Assunta Bertaccini

Abstract

Tomato (*Solanum lycopersicum*) is an important vegetable crop which is susceptible to multiple diseases that can impact yield and produce quality. The current research aimed to classify major tomato diseases in the Caribbean region through a developed intelligent system. The system was based on six key diseases including bacterial spot, early blight, late blight, septoria leaf spot, yellow leaf curl virus, and mosaic virus. To improve the identification accuracy, all images were first resized and then segmented using the GrabCut method to isolate the leaf regions. The segmented images were subsequently converted to Hue, Saturation, and Value (HSV) color space and grayscale before feature extraction was performed. A total of 96 features were extracted, including color histograms, Local Binary Pattern (LBP), and Gray-Level Co-Occurrence Matrix (GLCM). The Extreme Gradient Boosting (XGBoost) algorithm was employed for disease identification. This study used a dataset of 11,979 images from the PlantVillage collection, including both healthy and diseased samples. The developed model achieved a disease identification accuracy of 97.20% while remaining computationally efficient.

Keywords: color features, Gray-Level Co-occurrence Matrix, identification, local binary, texture features, XGBoost

Introduction

The tomato cropping season varies by region worldwide but generally occurs in the spring and summer. However, in Trinidad and Tobago and the wider Caribbean region, tomatoes can be grown year-round. It is a mixed cropping, interplanted with other crops in the home garden (<https://agriculture.gov.tt/wp-content/uploads/2017/11/Tomato-Factsheet-2013.pdf>). Tomato cultivation is challenged by different diseases and pests which cause serious damage to plant yield and product quality. The most common method of visual aided disease identification is often inaccurate due to the redundancy of symptoms. Human error was a significant fact with respect to the need for human input in the process. Although laboratory-based disease identification is more accurate, it is time consuming and may result in delays in identification, and application of disease control measures. Hence, farmers

should be aware and proactive in early detection of disease in order to put forth effective control methods and to maximize plant yield and quality.

The advancement of digital tools for disease detection has been significantly driven by key developments in image preprocessing, segmentation, feature extraction, and identification. Image preprocessing plays a crucial role in enhancing image quality and improving the efficiency of subsequent analyses. Among the widely adopted segmentation techniques, clustering (Ramesh *et al.* 2020; Jamjoom *et al.* 2023; Rusliyawati *et al.* 2024) and thresholding (Vini and Rathika 2023; Ramesh Babu *et al.* 2024) are the most commonly used methods. Feature extraction techniques focus on identifying key attributes such as color, edge, and texture information. Some of the prevalent feature extraction approaches include Convolutional Neural

Networks (CNNs) (Sapkal and Kulkarni 2018; Kadam *et al.* 2024), Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT). The primary advantage of feature extraction is its ability to reduce computational complexity while ensuring that machine learning models can effectively classify the extracted features (Chakole *et al.* 2022; Patil and Burkpalli 2022).

In agriculture, new technologies (Golka *et al.* 2025) have been introduced to help farmers protect their fields and increase their productivity. Researchers have developed numerous web-based (Czembor *et al.* 2025) and mobile-based (Rani *et al.* 2023) applications to assist in this effort. For plant foliar disease identification, many researchers have focused on analyzing the color and texture features of leaves. Some of these contributions are highlighted here. Traditional approaches emphasized segmentation and handcrafted features, such as the modified factorization-based active contour with color, GLCM, and LBP features proposed by Patil and Burkpalli (2022), fuzzy logic with 36 parameters and GLCM features (Tutygin *et al.* 2019), and the CIELAB-based preprocessing with GLCM and *Radial Basis Function Neural Network* (RBFNN) identification (Rusliyawati *et al.* 2024), achieved accuracies between 91–95%. Comparative studies show that deep learning outperforms these classical approaches. Sapkal and Kulkarni (2018) demonstrated that CNNs surpassed GLCM-based Backpropagation Neural Network (BPNN) (93.85%), while Suvarna *et al.* (2023) found CNNs more accurate than K-Means, Support Vector Machine (SVM), Radial Basis Function (RBF), Multilayer Perceptron (MLP), and BPNN algorithms produced less than 90% accuracy. Hybrid methods that blend clustering and deep learning utilized by Ramesh *et al.* (2020) improved the lesion localization and identification across multiple diseases. In terms of architectures, Ji and Du (2023) showed that a basic CNN (90%) performed comparably to heavier models like VGG16 (89%) and Inceptionv3 (87%), with MobileNet (91%) proving most suitable for mobile deployment. Similarly, Zhou *et al.* (2021) combined residual and dense networks to reduce parameters while achieving 95% accuracy. Yue *et al.* (2024) introduced a modified YOLOv7-GCA with GhostNetV2, cascading fusion, and Convolutional Block Attention Module (CBAM) which improved robustness while reducing model size for mobile use. Optimization approaches have also been explored for foliar disease identifications. For instance, Nagpal and Goel (2024) applied Dynamic Accelerated Memory-Based PSO (DAMPPO) to tune hyperparameters and achieved 95% accuracy with XGBoost for maize and cotton, while Chakole *et al.* (2022) integrated low-cost sensors with a hybrid analog-digital algorithm to improve detection across fungal,

bacterial, and viral diseases. On the deployment side, Kadam *et al.* (2024) developed a web-based application by Docker, POSTMAN API and TensorFlow serving server for tomato and potato leaf diseases, underscoring scalability, while other studies emphasized lightweight CNNs and MobileNet for offline, mobile deployment. Overall, the literature shows that CNNs and compact architectures such as MobileNet, GhostNetV2, and hybrid residual–dense models consistently provide the best trade-off between accuracy (around 95%), efficiency, and scalability. Detection-based and hybrid optimization methods further enhance robustness, making these approaches the most promising for real-time, and field-ready plant disease identification.

Much existing research on plant leaf disease identification has focused on neural networks due to their higher accuracy. However, these models require large amounts of data, which are computationally expensive. Also, there is a lack of interpretability and complexity in tuning. Hence, the proposed method enhanced plant disease detection by focusing on efficient feature extraction, precise segmentation of diseased areas, and improved identification accuracy. By leveraging advanced texture and color features, it can ensure better differentiation between healthy and diseased regions. Accurate segmentation minimizes noise, leading to more reliable feature extraction. This, in turn, enhances the performance of identification models, resulting in higher accuracy and robustness in disease identification.

The following were the objectives of the current study: to develop an intelligent system for identifying emerging bacterial and viral diseases affecting tomato crops in the Caribbean region; to enhance image processing techniques by applying preprocessing methods; to extract the meaningful features from tomato leaf images; to implement the XGBoost algorithm as a machine learning model for disease identification; and to evaluate the performance of the developed model.

Materials and Methods

The proposed system for disease identification consisted of preprocessing, segmentation, feature extraction, and identification. The working flow of the proposed method is shown in Figure 1.

Preprocessing

Preprocessing is crucial in image processing as it enhances image quality and facilitates more consistent feature extraction for improved analysis. In this

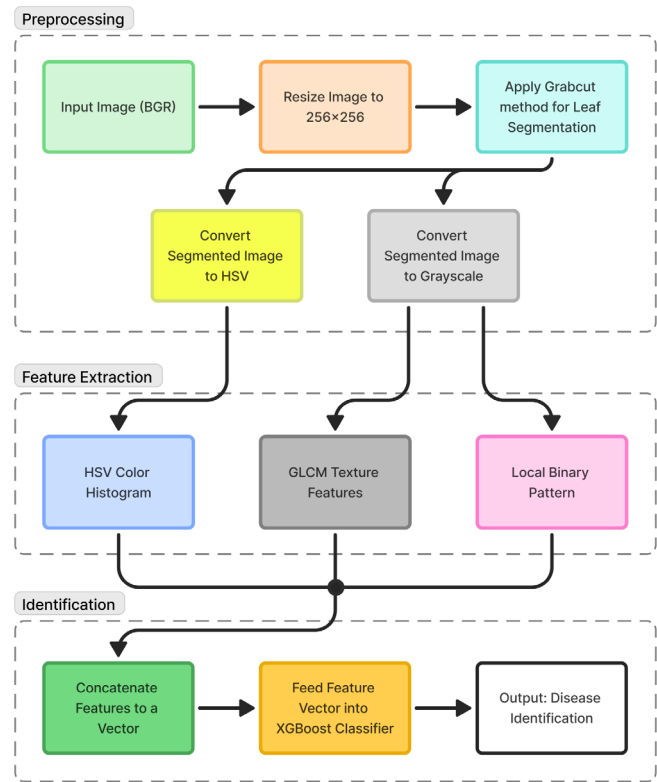


Fig. 1. Proposed method workflow

research, all images in the dataset were resized to 256×256 pixels prior to segmentation.

Segmentation

After resizing, segmentation of leaves from back-grounds was performed using the GrabCut method (Rother *et al.* 2004). GrabCut is an iterative image segmentation technique that models foreground and background color distributions with Gaussian Mixture Models and refines the partition using a graph-cut optimization procedure, allowing for efficient and relatively accurate extraction of the object of interest. This method uses two parameters: the rectangular margin, which determines how far inward from the image

borders the initial segmentation rectangle is drawn to fully capture the object, and the number of iterations, which defines how many refinement cycles the algo-rithm performs to enhance segmentation accuracy. The study was conducted using a 10-pixel margin and five iterations. Figure 2 highlights segmented images from each class.

Feature extractions

This process was divided into two parts:

Color features

The segmented image was first converted from Blue, Green and Red (BGR) to Hue, Saturation and Value

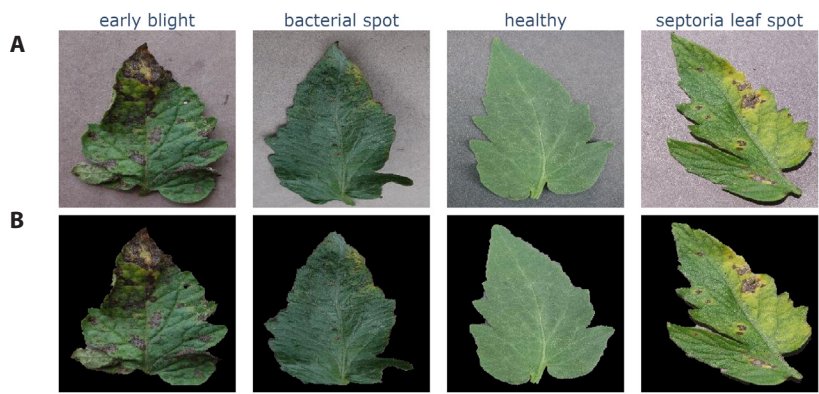


Fig. 2. Segmentation results (A) Original images (B) Segmented images

(HSV), a color space that separates an image into three components: Hue, Saturation, and Value. This separation is useful for capturing perceptual color differences. In this process, each channel is divided into a fixed number of bins. The Hue channel, which spans from 0 to 180, is divided into four intervals: approximately 0–45°, 45–90°, 90–135°, and 135–180°. Similarly, the Saturation and Value channels, each spanning from 0 to 256, are partitioned into four bins of 64 intensity levels each (0–64, 64–128, 128–192, and 192–256). As a result, a 3D histogram is created with $4 \times 4 \times 4 = 64$ unique bins, where each bin records the normalized frequency of pixels that fall within a specific combination of hue, saturation, and value ranges. This quantization process ultimately produces a 64-dimensional feature vector that effectively summarizes the color distribution within the region of interest. Table 1 groups the 64 HSV features by Hue bin. Each row represents 16 features, corresponding to every combination of the four Saturation and four Value bins for that specific Hue range.

Pattern and texture features

The segmented image is then converted to grayscale to focus on intensity variations and the following features are extracted:

- Local Binary Patterns (LBP):

LBP is applied to capture local texture patterns in the grayscale image (Ojala *et al.* 2002). With 24 sampling

points around each pixel and a radius of eight, each pixel's neighborhood is analyzed by comparing the pixel value to its neighbors. This comparison creates a binary pattern that encodes local structural information. The method then maps these patterns into a compact representation by focusing on uniform patterns, which are those with at most two transitions. As a result, the LBP process generates a histogram with 26 bins, each representing the frequency of a specific uniform pattern (with one bin reserved for non-uniform patterns). This 26-bin histogram provides a concise summary of the local texture features present in the image. Table 2 maps the LBP feature indexes to their corresponding pattern types.

- Gray Level Co-occurrence Matrix (GLCM)

After converting the image to grayscale, it is first quantized into eight discrete intensity levels to reduce complexity and emphasize the dominant gray-level variations. A Gray-Level Co-occurrence Matrix (GLCM) is then computed to capture the frequency at which pairs of pixel intensities occur in specified spatial relationships (Haralick *et al.* 1973; Rajamanickam *et al.* 2025). This matrix encodes how often different combinations of gray levels appear next to each other, providing a basis for a rich analysis of texture.

From the GLCM, six key statistical measures are extracted to characterize the image's texture and spatial relationships which are summarized in Table 3.

Table 1. Grouped HSV color histogram

Feature index range	Hue range	Saturation bins	Value bins	Description
0–15	0–45°	4 bins	4 bins	features for pixels with Hue in the first interval (0–45°)
		0–64	0–64	
		64–128	64–128	
		128–192	128–192	
16–31	45–90°	192–256	192–256	features for pixels with Hue in the second interval (45–90°)
		4 bins	4 bins	
		0–64	0–64	
		64–128	64–128	
32–47	90–135°	128–192	128–192	features for pixels with Hue in the third interval (90–135°)
		192–256	192–256	
		4 bins	4 bins	
		0–64	0–64	
48–63	135–180°	64–128	64–128	features for pixels with Hue in the fourth interval (135–180°)
		128–192	128–192	
		192–256	192–256	
		4 bins	4 bins	

Table 2. Local Binary Pattern feature index mapping and pattern types

Feature index	Pattern type	Description
0 to 24	uniform patterns	each bin corresponds to one of the 25 distinct uniform patterns, capturing common local texture structures with at most two transitions
25	non-uniform patterns	aggregates all patterns that do not conform to the uniform pattern definition, representing complex local textures

Table 3. Gray-Level Co-Occurrence Matrix statistical measures for texture analysis

Feature Index	GLCM Feature	Description
0	contrast	measures local intensity variations, emphasizing differences between neighboring pixel values
1	dissimilarity	quantifies the absolute differences between neighboring pixel intensities
2	homogeneity	indicates the concentration of GLCM values near the diagonal, reflecting similarity among adjacent pixels
3	energy	represents textural uniformity by summing the squared GLCM elements; higher values denote more uniform textures
4	correlation	assesses the linear dependency between pixel intensities, offering insights into the texture's structure
5	ASM	angular second moment, a measure of overall uniformity, closely related to the energy metric

Table 4. Summary of extracted features

Feature group	# of features	Feature names	Value range (typical)	Description
HSV color histogram	64	HSV bins 1-64	each bin $\in [0,1]$ Σ bins = 1	– aptures the color distribution in HSV space – Uses 4 bins per channel. $4 \times 4 \times 4$ – normalized so each bin value is between 0 and 1 – all bins sum to 1
Local Binary Patterns histogram	26	LBP bins 0-25	each bin $\in [0,1]$ Σ bins = 1	– extracted from the grayscale image – uses uniform LBP with 24 sampling points + 2 additional bins – encodes local texture patterns, resulting in a normalized histogram of 26 bins
Gray Level Co-Occurrence Matrix features	6	contrast, dissimilarity, homogeneity, energy, correlation, Angular Second Moment (ASM)	contrast: [0,49] dissimilarity: [0, 7] homogeneity: [0, 1] energy: [0, 1] correlation: [-1, 1] ASM: [0, 1]	– computed from a Gray-Level Co-occurrence Matrix (GLCM) with 8 intensity levels – measures second-order texture properties in the grayscale image – typical ranges assume normalized GLCM and 8 intensity levels

Total number of features: 64 (Color) + 26 (LBP) + 6 (GLCM) = 96 features.

Table 4 provides a comprehensive overview of each feature group along with the number of features,

specific feature names, typical value ranges, and detailed descriptions.

Identification

For disease identification, XGBoost was utilized. It is a scalable gradient boosting framework that builds sequential ensembles of decision trees to enhance performance (Chen and Guestrin 2016).

Results

The dataset used in this study consisted of 11,979 images of tomato leaves originally from the PlantVillage dataset (Arun Pandian and Geetharamani 2019). The dataset consists of detached leaf images captured under controlled, laboratory-like conditions with plain backgrounds. This research focused on major diseases occurring in the Caribbean region which consisted of seven specific classes: bacterial spot (2127), early blight (1000), late blight (1909), septoria leaf spot (1771), yellow leaf curl virus (3209), mosaic virus (373) and healthy controls (1591).

Using a stratified 80-20 split to preserve class distributions, the dataset was partitioned into training (80%) and test (20%) sets. The XGBoost model was trained on the training set (9583 images), while the test set (2396 images) was reserved for performance

evaluation. Hyperparameter tuning was performed using a comprehensive grid search with three-fold stratified cross-validation, which identified the optimal parameter combination. The results of the final model on the test set are highlighted in Tables 5 and 6.

Mosaic virus (Class 0) achieved a flawless 100% accuracy, while the other classes also maintained accuracy above 90% (Table 5). Precision, recall, and F1-scores showed near-perfect metrics for most categories, resulting in an overall F1-score of 0.97 (Table 6). These consistently high values highlight the robustness of the model in differentiating subtle variations among the diseases, even when classes share visual similarities. Notably, the perfect identification of mosaic virus highlights the model's ability to pinpoint unique characteristics associated with a specific disease. Overall, the model attained a 97.20% accuracy across 2396 samples, demonstrating its strong potential for reliable, large-scale screening and identification of plant diseases, which could significantly aid in timely interventions and effective crop management.

Figure 3 presents the multiclass Receiver Operating Characteristic (ROC) curves for plant disease identification, generated using a one-vs-rest approach. Each curve illustrates how effectively the model distinguished one class from all others at various threshold settings, providing insight into its overall discriminatory power. Notably, each class achieved an AUC of

Table 5. Identification results

Disease	Class	Total samples	Correct predictions	Incorrect predictions	Accuracy [%]
Mosaic virus	0	75	75	0	100.00
Yellow leaf curl virus	1	642	637	5	99.22
Bacterial spot	2	425	416	9	97.88
Early blight	3	200	180	20	90.00
Healthy	4	318	315	3	99.06
Late blight	5	382	361	21	94.50
Septoria leaf spot	6	354	345	9	97.46
Overall	–	2396	2329	67	97.20

Table 6. Per-class performance metrics

Disease	Class	Total samples	Precision	Recall	F1-score
Mosaic virus	0	75	0.99	1.00	0.99
Yellow leaf curl virus	1	642	0.99	0.99	0.99
Bacterial spot	2	425	0.97	0.98	0.98
Early blight	3	200	0.93	0.90	0.92
Healthy	4	318	1.00	0.99	0.99
Late blight	5	382	0.95	0.95	0.95
Septoria leaf spot	6	354	0.97	0.97	0.97
Overall	–	2396	0.97	0.97	0.97

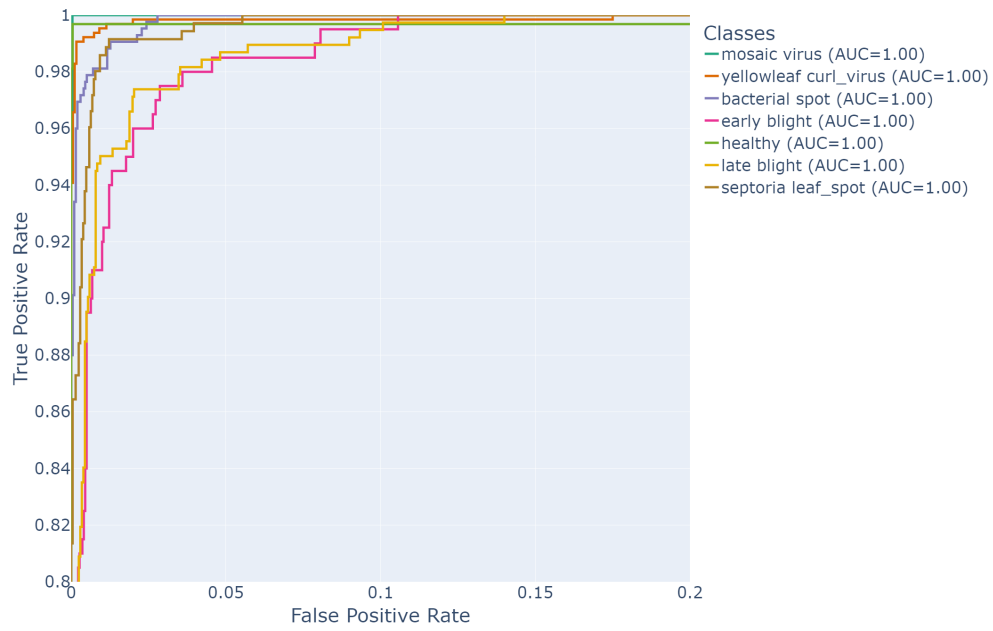


Fig. 3. Multiclass Receiver Operating Characteristic (ROC) curves for disease identification

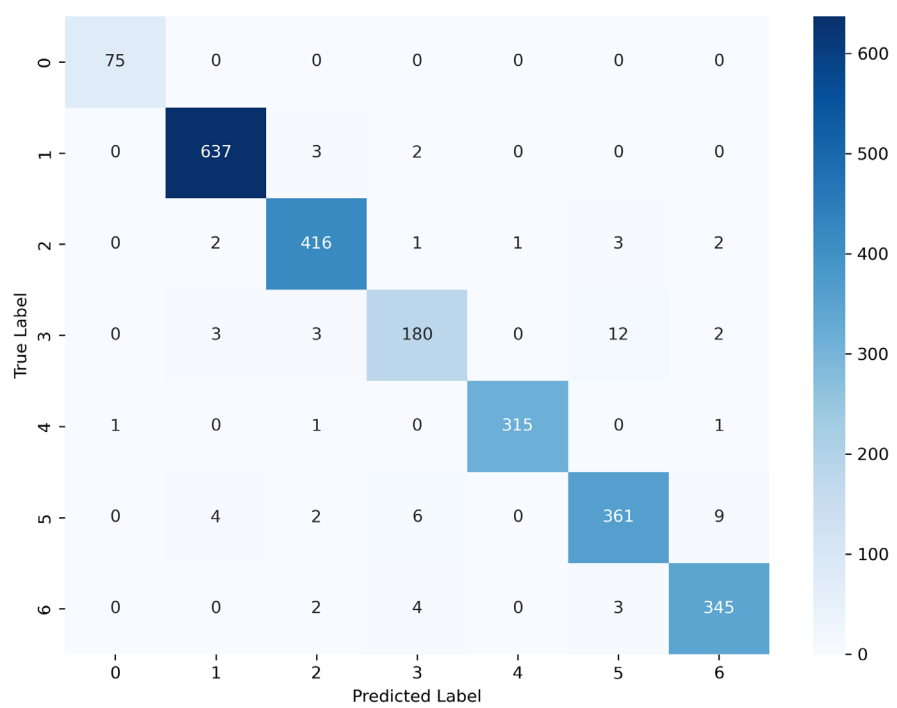


Fig. 4. Confusion matrix for tomato leaf disease identification

1.0, indicating that the model perfectly separated each disease category from the rest.

Figure 4 presents the confusion matrix for the model's predictions on the test dataset, providing a detailed representation of the distribution of correct and incorrect identifications across the respective classes.

Discussion

Accurate and efficient identification of plant diseases is crucial for early detection and effective management. Traditional methods often face challenges related to

noise, poor segmentation, and ineffective feature extraction, leading to suboptimal identification performance. To address these issues, this study proposed a method incorporating proper preprocessing, precise segmentation, and advanced feature extraction techniques. By refining these processes, the model enhanced disease detection accuracy while maintaining computational efficiency. The following discussion and Table 7 evaluate the effectiveness of this approach in comparison to existing models and highlights its advantages in improving identification performance.

Jamjoom *et al.* (2023) utilized GLCM and LBP features with SVM classifiers to classify four distinct types of plant diseases caused by *Phytophthora infestans*, *Fusarium graminearum*, *Puccinia graminis* and tomato yellow leaf curl virus, achieving an identical accuracy of 97.20%. In contrast, Yadav and Tewari (2025) proposed a two-stage model for tomato disease identification. In the first stage, tomato leaves were isolated using a Faster Region-Based Convolutional Neural Network (FRCNN) enhanced with Feature Pyramid Networks (FPN) to improve detection accuracy. In the second stage, a customized conformer-based model was employed for feature extraction from the segmented leaves. The integration of these two stages, referred to as CONF-RCNN, resulted in an identification accuracy of up to 90% in identifying tomato diseases. Similarly, Ji and Du (2023) recommended the suitability of MobileNet and CNN for mobile devices due to its lightweight structure, fast functioning ability and adaptability. These models also attained an accuracy of 90%. Using another CNN-based model, Wu *et al.* (2020) proposed GoogLeNet which incorporated data augmentation using generative adversarial networks for leaf disease recognition. This model achieved an average identification accuracy of 94.33% on a dataset of tomato leaf disease and consisting of 1500 images across 5 classes.

The findings of this study highlight the effectiveness of the XGBoost model in identifying six types of tomato diseases along with healthy leaves. This identification was achieved through the integration of feature extraction techniques, including HSV color histogram, LBP, and GLCM. The model's performance was assessed using key evaluation metrics such as Precision, Recall, F1-Score, and accuracy, providing a comprehensive analysis of its identification capabilities.

XGBoost's strong performance for identification can be attributed to several key features, namely it is a scalable and efficient tree boosting framework that implements gradient boosting algorithms for both identification and regression tasks by sequentially building an ensemble of weak decision trees, each trained to correct the errors of its predecessors. Additionally, it integrates L1 and L2 regularization to mitigate overfitting and enhance model generalization,

employs a weighted quantile sketch to efficiently handle sparse data, and supports parallel processing to boost computational efficiency. These combined capabilities contribute to its robust predictive performance.

Even though a comprehensive set of 96 features was extracted to capture detailed color and texture characteristics, the overall process remains computationally efficient. The feature extraction techniques employed included quantized HSV color histogram, LBP, and GLCM. These were chosen because they directly reflect the visual symptoms most commonly associated with plant disease. HSV captures leaf color changes such as yellowing, chlorosis, or necrosis, which are primary indicators of infection. LBP and GLCM, complement this by encoding structural and textural irregularities that often appear on the leaf surface. These methods were preferred over more complex alternatives such as Scale-Invariant Feature Transform (SIFT) or Histogram of Oriented Gradients (HOG), which can generate several hundred to several thousand descriptors per image. While such approaches provide highly detailed representations, they increase computational cost and may include redundant information not directly related to disease symptoms. By contrast, the 96 features used here were compact and interpretable, contributing to the strong overall identification performance achieved by the model. Moreover, XGBoost's optimized gradient boosting framework ensured that the classifier is trained very quickly despite the extensive feature set. This synergy between effective feature extraction and efficient model training allows rapid and robust disease identification.

Traditional machine learning models like XGBoost can be more advantageous than neural networks and convolutional neural networks in image processing when the available data is limited or when faster training times and easier model interpretability are required. While CNNs excel at automatically extracting spatial features from images (LeCun, Bengio, & Hinton 2015), XGBoost can offer competitive performance by working with preprocessed features and demands lower computational resources, making it a practical choice in scenarios where deep learning solutions may be excessive.

Each baseline was compared with XGBoost using McNemar's paired test on the same test images, as seen in Table 8. The effect column Δ shows how many additional correct decisions XGBoost made compared to the baseline on items where the models disagreed. The XGB share of discordant percentage shows how often those disagreements favored XGBoost. Across all baselines XGBoost won decisively with p values far below 0.05. The closest competitor was Random Forest with $b = 18$ and $c = 67$ which meant that XGBoost made 49 extra correct decisions and 79 percent of disagreements favored XGBoost. Larger advantages were seen

Table 7. Comparison results with reported methods

Methods(s)	Crop(s)	Accuracy [%]	Reference
Basic CNN	Tomato	90	Ji and Du (2023)
CONF-RCNN	Tomato	90	Yadav and Tewari (2025)
Radial Basis Function Neural Network (RBFNN)	Pepper	91.67	Rusliyawati <i>et al.</i> (2024)
GoogLeNet - Augmentation method by generative adversarial networks	Tomato	94.33	Wu <i>et al.</i> (2020)
SVM	Tomato yellow leaf curl and three other diseases	97.20	Jamjoom <i>et al.</i> (2023)
XGBoost	Tomato	97.20	Current method

Table 8. McNemar's paired test on the shared test set

Baseline	b [base-line only correct]	c [XGB only correct]	$\Delta = c - b$	XGB share of discordant [%]	χ^2	p-value
Dummy (most frequent)	5	1692	1687	99.7%	1675.071	0
GaussianNB	11	718	707	98.5%	683.726	1.034e-150
DecisionTree	20	338	318	94.4%	280.696	5.297e-63
Logistic regression	13	303	290	95.9%	264.307	1.976e-59
LinearSVC	12	268	256	95.7%	232.232	1.943e-52
KNN(k=5)	23	241	218	91.3%	178.367	1.101e-40
RBF-SVC	16	200	184	92.6%	155.042	1.371e-35
Random forest(200)	18	67	49	78.8%	27.106	1.926e-07

for kNN, SVMs, logistic regression, and Naive Bayes. These results indicate that XGBoost achieved higher accuracy and more unique correct predictions on the very cases where baselines tend to fail. Overall, the consistent paired gains point to a robust and practical meaningful advantage for XGBoost on this dataset and support the choice as the primary model for identification.

The relative importance of each feature group, including the HSV histogram, LBP, and GLCM was also examined. XGBoost calculated feature importance by evaluating how much each feature reduced the loss function, often by measuring improvements in the splitting criterion (such as gain or cover) across all trees in the model. After obtaining these individual scores, the 96 features were grouped into three segments and their importances were summed. The analysis revealed that the first group, HSV Color Histogram (features 0–63), accounted for approximately 76.9% of the total importance. The second group, LBP Histogram (features 64–89), contributed about 18.9% and the third

group, GLCM (features 90–95), comprised roughly 4.2% of the overall importance. This indicates that the HSV color-based features played a dominant role in the identification process because they captured a wide range of perceptual color information. In contrast, the texture-based features from LBP and GLCM added extra discriminative power, though their contribution was relatively smaller. Such disparities may be partly explained by the larger number of features in the HSV segment, allowing it to capture a broader range of relevant information.

Conclusions

This study highlighted the critical importance of early detection for common tomato diseases in the Caribbean. A highly accurate identification model was introduced that leverages efficient, low-resource feature extraction and employed XGBoost for robust predictions. This model achieved an overall accuracy of

97.20% in distinguishing healthy tomatoes from six disease classes. This approach required significantly less computational expense than leading neural network architectures commonly used for similar work. Future research will be needed to focus on expanding the scope of the model to multiple crops and validating its performance under real-world field conditions. Additionally, deployment pathways will need to be explored, including integration into mobile applications for farmers, training programs to promote adoption and correct usage, and pilot field-testing across diverse agricultural environments. These steps will help ensure both the scalability and practical utility of the model for supporting agricultural disease management in resource-constrained regions.

Acknowledgements

We thank the Campus Research and Publication Fund Committee, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago for funding this study.

References

- Arun Pandian J., Geetharamani G. 2019. Data for: Identification of plant leaf diseases using a 9-layer deep convolutional neural network [Data set]. Mendeley Data, V1. DOI: <https://doi.org/10.17632/tywbtsjrjv.1>.
- Chakole S. S., Khera S., Ukani N. A. 2022. Analysis of Plant Leaf Disease Based on Sensor and Machine Learning Technique, p. 1-4. In: Proceedings of the IEEE International Conference on Current Development in Engineering and Technology (CCET). Bhopal, India. DOI: 10.1109/CCET56606.2022.10080352.
- Chen T., Guestrin C. 2016. XGBoost: A scalable tree boosting system. p.785-794. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery. DOI: <https://doi.org/10.1145/2939672.2939785>.
- Czembor E., Tratwal A., Pukacki J., Krystek M., Czembor J.H. 2025. Managing fungal pathogens of field crops in sustainable agriculture and AgroVariety internet application as a case study. *Journal of Plant Protection Research* 65 (1): 1–26. DOI: <https://doi.org/10.24425/jppr.2025.153820>.
- Golka W., Szechynska-Hebda M., Kowalska J., Golka A., Grolak R., Jakubowska M., Mateusz G., Bomberski A. 2025. Selective protection of cereals using artificial neural networks. *Journal of Plant Protection Research* 65 (2): 164–173. DOI: <https://doi.org/10.24425/jppr.2025.155053>.
- Haralick R.M., Shanmugam K., Dinstein I. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* 3 (6): 610–621. DOI: <https://doi.org/10.1109/TSMC.1973.4309314>.
- Jamjoom M., Elhadad A., Abulkasim H., Abbas S. 2023. Plant leaf diseases classification using improved k-means clustering and SVM algorithm for segmentation. *Computers, Materials & Continua* 76 (1): 367–382. DOI: <https://doi.org/10.32604/cmc.2023.037310>.
- Ji Z., Du M. 2023. Study on Tomato Disease Classification based on Leaf Image Recognition based on Deep Learning Technology, *International Journal of Advanced Computer Science and Applications (IJACSA)* 14 (4). DOI: <http://dx.doi.org/10.14569/IJACSA.2023.0140463>.
- Kadam P., Bidwe R., Thorat H., Shelke S., Mahajan K., Yewale P., Hiremani V. 2024. Web-based Application for Plant Leaf Disease Detection. *Journal of Electrical Systems* 20 (1s): 73–83. DOI: <https://doi.org/10.52783/jes.753>.
- LeCun Y., Bengio Y., Hinton G. 2015. Deep learning. *Nature* 521 (7553): 436–444. DOI: <https://doi.org/10.1038/nature14539>.
- Nagpal J., Goel L. 2024. DAMPSO: Dynamic Accelerated Memory-Based PSO for Hyperparameter Tuning of Plant Disease Classifiers. In: Proceedings of Fifth Doctoral Symposium on Computational Intelligence (DoSCI 2024). Lecture Notes in Networks and Systems 1086. Springer, Singapore. DOI: https://doi.org/10.1007/978-981-97-6036-7_25.
- Ojala T., Pietikäinen M., Mäenpää T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 971–987. DOI: <https://doi.org/10.1109/TPAMI.2002.1017623>.
- Patil B., Burkpalli V. 2022. Cotton Leaf Disease Classification by Combining Color and Texture Feature-based Approach. p. 1783-1789. In: Proceedings of the 6th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India. DOI: 10.1109/ICICCS53718.2022.9788405.
- Rajamanickam V., Ramsubhag A., Jayaraman J. 2025. Sweet pepper foliar diseases quantification and identification using an image analysis tool. *Journal of Plant Protection Research* 65 (1): 89–99. DOI: <https://doi.org/10.24425/jppr.2025.153821>.
- Ramesh G., Albert W., Ramu G. 2020. Detection of Plant Diseases by analyzing the Texture of Leaf using ANN Classifier. *International Journal of Advanced Science and Technology* 29 (8): 1656–1664. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/12580>.
- Ramesh Babu P., Srikrishna A., Gera V.R. 2024. Diagnosis of tomato leaf disease using OTSU multi-threshold image segmentation-based chimp optimization algorithm and LeNet-5 classifier. *Journal of Plant Diseases and Protection* 131: 2221–2236. DOI: <https://doi.org/10.1007/s41348-024-00953-7>.
- Rani S., Das K., Aminuzzaman F., Ayim B. Y., Borodynko-Filas N. 2023. Harnessing the future: cutting-edge technologies for plant disease control. *Journal of Plant Protection Research* 63 (4): 387–398. DOI: <https://doi.org/10.24425/jppr.2023.147829>.
- Rother C., Kolmogorov V., Blake A. 2004. GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics* 23 (3): 309–314. DOI: <https://doi.org/10.1145/1015706.1015720>.
- Rusliyawati R., Karnadi K., Tanniewa A.M., Widyawati A.C., Jusman Y., Borman R.I. 2024. Detection of Pepper Leaf Diseases Through Image Analysis Using Radial Basis Function Neural Networks. *BIO Web Conference* 144 01005. DOI: 10.1051/bioconf/202414401005.
- Sapkal T.A., Kulkarni V.U. 2018. Comparative study of leaf disease diagnosis system using texture features and deep learning features. *International Journal of Applied Engineering Research* 13 (19): 14334–14340.
- Suvarna T.S.L., Sai N. N., Rao S.D., Gowtham G., Anoosha C. 2023. Enhancing Crop Health by Identifying Leaf Disease with High Accuracy Using CNN. p. 403-408. In: Proceedings of the Third International Conference on Secure Cyber Computing and Communication (ICSCCC). Jalandhar, India. DOI: 10.1109/ICSCCC58608.2023.10176621.
- Tutygin V.S., Al-Vindi B.X. M.A., Leliuhin, D.O. 2019. The use of an extended set of key texture features haralick in the diagnosis of plant diseases on leaf images. *Vibroengineering PROCEDIA* 25: 122–127. DOI: <https://doi.org/10.21595/vp.2019.20830>.

- Vini S.L., Rathika P. 2023. Thresholding based tomato leaf disease classification. p. 143–147. In: Proceedings of the 4th International Conference on Signal Processing and Communication (ICSPC). Coimbatore, India. DOI: 10.1109/ICSPC57692.2023.10125865.
- Wu Q., Chen Y., Meng J. 2020. DCGAN-Based Data Augmentation for Tomato Leaf Disease Identification. *IEEE Access* 8: 98716–98728. DOI: 10.1109/ACCESS.2020.2997001.
- Yadav S., Tewari A.S. 2025. CONF-RCNN: a conformer and faster region-based convolutional neural network model for multi-label classification of tomato leaves disease in real field environment. *Journal of Plant Diseases and Protection* 132: 61. DOI: <https://doi.org/10.1007/s41348-024-01057-y>.
- Yue X., Li H., Song Q., Zeng F., Zheng J., Ding Z., Kang G., Cai Y., Lin Y., Xu X., et al. 2024. YOLOv7-GCA: A lightweight and high-performance model for pepper disease detection. *Agronomy* 14 (3): 618. DOI: <https://doi.org/10.3390/agronomy14030618>.
- Zhou C., Zhou S., Xing J., Song J. 2021. Tomato leaf disease identification by restructured deep residual dense network. *IEEE Access* 9: 28822–28831. DOI: <https://doi.org/10.1109/ACCESS.2021.3058947>.