Journal of Plant Protection Research eISSN 1899-007X

ORIGINAL ARTICLE

Sweet pepper foliar diseases quantification and identification using an image analysis tool

Vijayanandh Rajamanickam[1]*, Adesh Ramsubhag[2], Jayaraj Jayaraman[2]

[1] Department of Computing and Information Technology, Faculty of Science and Technology, The University of the West Indies St. Augustine Campus, Trinidad and Tobago
[2] Department of Life Science, Faculty of Science and Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago

Vol. 65, No. 1: 89–99, 2025

DOI: 10.24425/jppr.2025.153821

Received: April 25, 2024
Accepted: July 25, 2024
Online publication: March 25, 2025

*Corresponding address:
Vijayanandh.Rajamanickam@sta.uwi.edu

Responsible Editor:
Piotr Iwaniuk

Abstract

Quantification and assessment of disease symptoms are important elements of plant disease management systems and are required to assist with making decisions on the choice of protective agents to be applied to crops or for screening plant genotypes for the development of resistant varieties. Traditional methods of identifying and quantifying disease severity are cumbersome, involving visual assessment tools or scales, and rating of plants at a point in time. Visual assessment is prone to human bias and error, thereby reducing the efficiency and accuracy of this method. In this study, we developed a smartphone camera-based image recording, processing, and assessment tool for measurement of symptoms of early and late blight, and bacterial leaf spot diseases in sweet pepper caused by *Alternaria solani, Phytophthora infestans,* and *Xanthomonas campestris* pv. *vesicatoria,* respectively. Sweet pepper or bell pepper is a major vegetable crop grown in the Caribbean region, but production is severely affected by plant diseases, most important of which include foliar infections by fungi and bacteria that cause major losses in fruit yield. This research utilized smartphone captured images of leaf specimens for severity measurement and classification of diseases. The steps involved were color space conversions, detection of leaf area by Otsu's method, and thresholding for foliar diseased area detection and quantification. Gray-Level Co-occurrence Matrix (GLCM) extracted the texture features from the diseased area of leaves. These features are trained and classified by various machine learning classifiers including trees, rule-based and Bayes models. Application of decision trees and rule-based classifier models achieved 98% accuracy individually, while Bayes model achieved 86% accuracy. The image input into the above classifier models resulted in fast and accurate identification of the diseases by matching the features of trained images of disease symptoms. This method could work well for leaves collected from field-grown plants as well as from inoculated greenhouse plants.

Keywords: foliar disease, GLCM, machine learning, Otsu, thresholding

Introduction

Peppers (*Capsicum* spp.) are the third most important solanaceous crop of the world, and in the Caribbean region, sweet pepper (*Capsicum-annuum*) is consumed as a staple vegetable that is used in many West Indian dishes (https://agriculture.gov.tt/wp-content/uploads/2017/11/how-to-grow-sweet-peppers.pdf). It can be grown throughout the year, giving the advantage of being available as multiple varieties and F1 hybrids which are specific for seasons, soil types, and consumer preferences. Sweet pepper is a rich source of carotenoids, vitamins, protein, fiber, and calcium. It is often preferred by many market growers and home gardeners for its commercial value and easiness of cultivation (http://files-do-not-link.udc.edu/docs/causes/online/Pepper%2010.pdf). Even though the environmental conditions are favorable to sweet pepper cultivation in the Caribbean region, the damage caused by foliar diseases, especially early and late blight, and bacterial leaf spot, are the most serious in nature, and affect the crop at every growth stage contributing to

significant yield losses (Naegele *et al.* 2017; Chinnadurai *et al.* 2018).

Plants are subjected to damage caused by pathogens and pests throughout a crop's life stages. Infections by pathogens lead to disease epidemics at varying levels of severity and among the most devastating are those caused by fungal and bacterial pathogens. With the advent of global climate change the intensity of diseases and stress factors are increasing, causing serious losses in crop yield and quality, and ultimately posing a threat to food security and safety (Ramsubhag and Jayaraman 2018). Therefore, in the current era, it is more important than ever to develop and implement effective plant health management systems, to protect crops from harmful biological pathogens, pests, and environmental stresses (Shafique *et al.* 2016; Rani *et al.* 2023).

Quantification of foliar diseases to obtain data on the occurrence, severity and progression of diseases is an essential component of plant disease management systems. Such quantification is required in many phases of the crop cycle, particularly for disease assessment and surveillance, selection of suitable protection agents to apply to crops, and resistance breeding (Bock *et al.* 2022). The traditional methods of disease estimation include direct or indirect approaches, with measurements typically including nominal or descriptive, ordinal, interval or category, and ratio scales (Alheeti *et al.* 2021). Disease measurements using traditional methods are normally done by naked-eye observations utilizing the scales for quantification. There are several limitations involved in this process including the need for technical expertise, and intensive labor is required. To complete the process is time consuming, and there is subjectivity in interpreting results, as well as a lack of precision and accuracy of the data collected. Therefore, the method is error-prone, with a high chance of generating incorrect estimates, which compromises the quality of data, and could contribute to incorrect conclusions (Shi *et al.* 2023). To overcome these issues, image analysis tools could be employed that are machine interfaced and less biased in data measurement and analysis. Digital tools can be designed to make the process semi-automatic or automatic, which speeds up the process of data collection, maintains consistency and accuracy of estimates, and significantly minimizes the cost of the operation. Furthermore, the developed tools can be easily applied by researchers and even farmers for early detection and precise disease prediction and measurement (Li *et al.* 2020).

The application of digital tools helps in early alert and deployment of appropriate disease management practices at the right point in time to improve the crop yield and produce quality. Machine-based assessment of infection and measurement of responses to stress factors can contribute to significantly improving the effectiveness and efficiency of plant health management systems. The development of machine-based assessment methods using image analysis will enable accurate and faster measurement of host responses as well as in real-time and could also be easily automated.

Many researchers have focused on developing various machine learning models for plant disease detection and classification (Das *et al.* 2022). Machine learning models utilize information such as color, shape, texture, and deep learning features. Preprocessing and segmentation contribute to this process. Texture features of plant samples and deep learning models (Saleem *et al.* 2019; Zeng *et al.* 2021; Nyarko *et al.* 2023) resulted in better performance outcomes than conventional handcrafted features (Hassan *et al.* 2022; Francisco *et al.* 2023). Chakole *et al.* (2022) investigated a low-cost sensor and a hybrid machine-learning algorithm for detecting any plant fungal, bacterial, and viral diseases and reported improved detection accuracy using a hybrid algorithm that combined both analog and digital machine learning. Sapkal and Kulkarini (2018) compared and discussed two feature extraction techniques: Gray Level Covariance Matrix and deep learning model. The classification was carried out using a Backpropagation Neural Network (BPNN). The accuracy achieved was 93.85%, which is significantly higher than the texture feature extraction technique. Ramesh *et al.* (2020) utilized deep learning methods using K-Means clustering for detecting and classifying diseases including Alternaria spot, anthracnose, bacterial blight, and *Cercospora*.

The two methods for detecting plant leaf diseases proposed by Singh *et al.* (2022) include CNN to extract features, a Bayesian optimized Support Vector Machine (SVM) classifier and preprocessing before extracting texture and color features using histograms of oriented gradients (HoG) and GLCM. To simplify the process of classification, image segmentation plays an important role in the image analysis (Singh and Misra 2017). Otsu threshold is a widely used technique for image segmentation. It identifies a global threshold for segmenting an image into two categories namely, the foreground and the background. Every pixel value in an image is compared to a threshold value. If the value of the pixel is greater than the threshold, then it is classified as foreground, otherwise it is classified as background pixel (Yogeshwari and Thailambal 2020).

This research aimed to develop a digital tool which would accurately identify and quantify the severity of pepper foliar diseases, thus enabling effective application of disease management tools. The research was completed using ImageJ, which is a free, platform-independent program for performing scientific image analysis.

Materials and Methods

Crop plants

Symptomatic sweet pepper plant leaf samples were collected from Orange Grove and Macoya, Trinidad from seven fields during wet and dry seasons (2022–2023). Totally, 200 samples were collected from identified plants in all the fields and crop stages. The sweet pepper plants were grown according to the production manual published by the Ministry of Agriculture, Land and Fisheries, Trinidad and Tobago (https://agriculture.gov.tt/category/publications/manuals/). The collected leaf samples were transferred to sandwich bags and kept in an ice box and brought to the lab immediately for picturing. Pictures available in the PlantVillage repository (https://www.kaggle.com/datasets/emmarex/plantdisease) were also used for comparison.

Image capture and analysis

The plant leaf was captured by an 8.0 MP Smartphone (Samsung Galaxy A03 Core). The images of leaves were captured in the field as well as in the lab. In the lab, pictures were taken with and without an inbuilt flash. The captured images were given as input to the proposed application developed in the ImageJ program for the quantification, classification, and identification of disease.

The efficiency of the proposed digital tool for disease severity measurement was compared parallelly with the results obtained from the APS Assess 2.2: Image analysis software for plant disease quantification (https://my.apsnet.org/APSStore/Product-Detail.aspx?WebsiteKey=2661527A-8D44-496C-A730-8CFEB6239BE7&iProductCode=43696). Different resolution images were tested, and validated for challenge of shadow.

Proposed pipeline

This research elaborated on three major contributions namely quantification, classification, and identification of foliar diseases of sweet pepper as follows.

Quantification

Quantification of severity was further divided into multiple sub-steps in a sequence. The captured image was passed as input to the program and then converted to two color spaces namely Grayscale and Hue-Saturation-Brightness (HSB). The grayscale image simplifies algorithms and eliminates the complexities related to computational requirements due to the spatial dimensions. The grayscale image was applied to Otsu's method (Otsu 1979) which yields the binary leaf image classified as foreground (leaf area) and background.

The HSB image is considered Hue (H) and Brightness (B) threshold values to detect the diseased areas. The following range was obtained based on the trials made on both Smartphone captured and PlantVillage repository images.

$$0 \leq H \leq 44 \quad \text{and} \quad 24 \leq B \leq 242.$$

The pixel count can be used for diseased area quantification which was calculated from the pixel count measured from the Otsu's leaf area and the pixel count of the diseased area from the threshold. The noises of the segmented image were eliminated by dilation of morphological operations since it can be used to remove small noise points or to fill small holes in images.

Detailed analysis of the quantification was carried out by comparisons with APS Assess 2.2 automatic assessment and the traditional visual-based disease intensity/severity index scale proposed by Ali *et al.* (2019). Also, the challenges in image resolutions by comparing images of detached leaves captured in the field and laboratory are discussed. The background of the leaf may not be uniquely bright due to the field environment. Furthermore, it varies with the conditions including field, in-house, lab, etc. and with or without a flash. The leaf may also have a mixture of disease symptoms. All these parameters were taken into consideration while photographing the samples.

Classification

Classification of diseases can be performed by extracting the texture features from the isolated diseased area. The GLCM features were extracted by applying features including: Energy, Entropy, Contrast, Correlation, Homogeneity, Prominence, Shade, Variance, Angular Second Moment (ASM), Inverse Difference Moment (IDM) and Inertia (Priyanka and Kumar 2020). The statistics can be divided into first, second, and higher orders, but higher-order statistics are not practical to implement due to their computational complexity.

- **Energy:** It distinguishes homogeneous regions from non-homogeneous regions (Priyanka and Kumar 2020). The energy feature gauges the repetition of pixel pairs and assesses the disorder in texture within an image (Mall *et al.* 2019).
- **Entropy**: Entropy evaluates the randomness in the image. As a result, a homogeneous image will have a lower entropy value (Priyanka and Kumar 2020). Entropy is closely but inversely related to energy, as images with more gray levels have high entropy (Mall *et al.* 2019).
- **Contrast:** Contrast assesses the intensity differences between a pixel and its neighbors in the entire image (Mall *et al.* 2019).

- **Correlation:** Correlation measures the linear dependencies of gray tones in an image. It indicates how a pixel is related to its neighbors (Priyanka and Kumar 2020). The correlation value ranges from –1 to 1 for negative and positive correlation in images, and for a constant image, the value is infinite (Mall *et al.* 2019).
- **Homogeneity:** Homogeneity refers to the similarity of pixels present in an image. The GLCM matrix of a homogeneous image results in a value of 1. It is very low when the image texture needs minimal adjustments (Mall *et al.* 2019).
- **Prominence:** A higher prominence value of image indicates more irregularity, while a lower value indicates a peak, centered around the average (Mall *et al.* 2019).
- **Shade:** A high value of shade indicates a greater level of asymmetry in an image, while a low value implies a more uniform distribution of gray levels (Mall *et al.* 2019).
- **Variance:** The variance evaluates the spread of the gray level sum distribution in an image. As the gray level values deviate from their average, the variance of the image increases (Mall *et al.* 2019).
- **Angular Second Moment** (**ASM**): The ASM is an indicator of the uniformity of an image. If a scene is homogeneous, it will have only a limited number of gray levels, resulting in a GLCM with only a few, but relatively high values of P(i, j), and thus, the sum of squares will be high (Kulmaganbetov *et al.* 2022).
- **Inverse Difference Movement** (**IDM**): IDM is also influenced by the homogeneity of an image. Because of the weighting factor $[1+ (i - j)^2]^{-1}$ IDM will get small contributions from nonhomogeneous areas $(i \neq j)$ (Kulmaganbetov *et al.* 2022).
- **Inertia**: The inertia of an image gives more importance to each value in the GLCM matrix, with a focus on areas with high contrast.

The extracted features were trained by machine learning algorithms and classified by using various classifiers. Three different classifiers, namely decision trees, rule-based and Bayes models were applied. Decision tree models, including J48, RandomForest, RandomTree and HoeffdingTree, and the rule-based model called DecisionTable, and the Bayes' theorem model called NaiveBayes were applied for classification (Sarker *et al.* 2019). The classifier models' performance was compared with Receiver Operating Characteristic (ROC) curves that depict how the number of correctly classified positive instances varies with the number of incorrectly classified negative instances. If the class distribution is skewed, ROC curves provide an overly optimistic view of an algorithm's performance (Davis and Goadrich 2006).

Identification

One of the core objectives of this study was to identify the type of disease in the plant specimen. The process of identification can be performed by the input image which went through the step of isolating the diseased area, followed by GLCM features – extraction. Based on the above results, a filtered classifier was built to match the GLCM features with the existing trained feature sets. From these matchings, the disease type of the given input image was predicted. The prediction results were highly efficient in terms of accuracy and fastness of time. The workflow of the proposed quantification of leaf disease severity and identification of leaf disease is presented in Figure 1.

Experimental results

Quantitative and qualitative image analysis

The quantification, classification, and identification of foliar diseases in sweet pepper were performed as per the steps indicated in Figure 1. The isolation of leaf area plays an important role that eliminates the healthy parts of a plant image. It also helps to measure the leaf area in terms of pixels, which include both healthy and diseased areas of an image. The sample output obtained from Otsu's method is shown in Figure 2. Hue and brightness values were used to isolate the diseased area from the HSB-converted image.

Figure 2A shows the original image, which was converted to grayscale and then applied to Otsu's method. This method produced the foreground image as black which refers to the leaf and the background as white. The result of Otsu's method is shown in Figure 2B.

The workflow of the proposed quantification method is shown from A to G in Figure 5 (on page 95). The result of Otsu's method and HSB diseased threshold values were used to isolate the disease area of the image. The final processed image (Fig. 5F) shows the output of the proposed quantification method. The total quantified leaf area of the processed image contained 19 2238 pixels of which the diseased area was 33 276 pixels. The calculated severity measurement was 17.31%.

Comparison of methods

The proposed method was tested with both Smartphone captured images and the PlantVillage dataset. The robustness of the proposed method in terms of quantification was compared with the results obtained from the APS Assess 2.2. Table 1 shows the comparison of the results of the proposed method Vs APS Assess tool and visual assessment. APS Assess 2.2 used the automatic panel tool for segmentation of diseased areas. Multiple leaf samples were analyzed using the developed tool. By comparing the results from the visual assessment by grading method (Ali *et al.* 2019) and APS Assess 2.2,

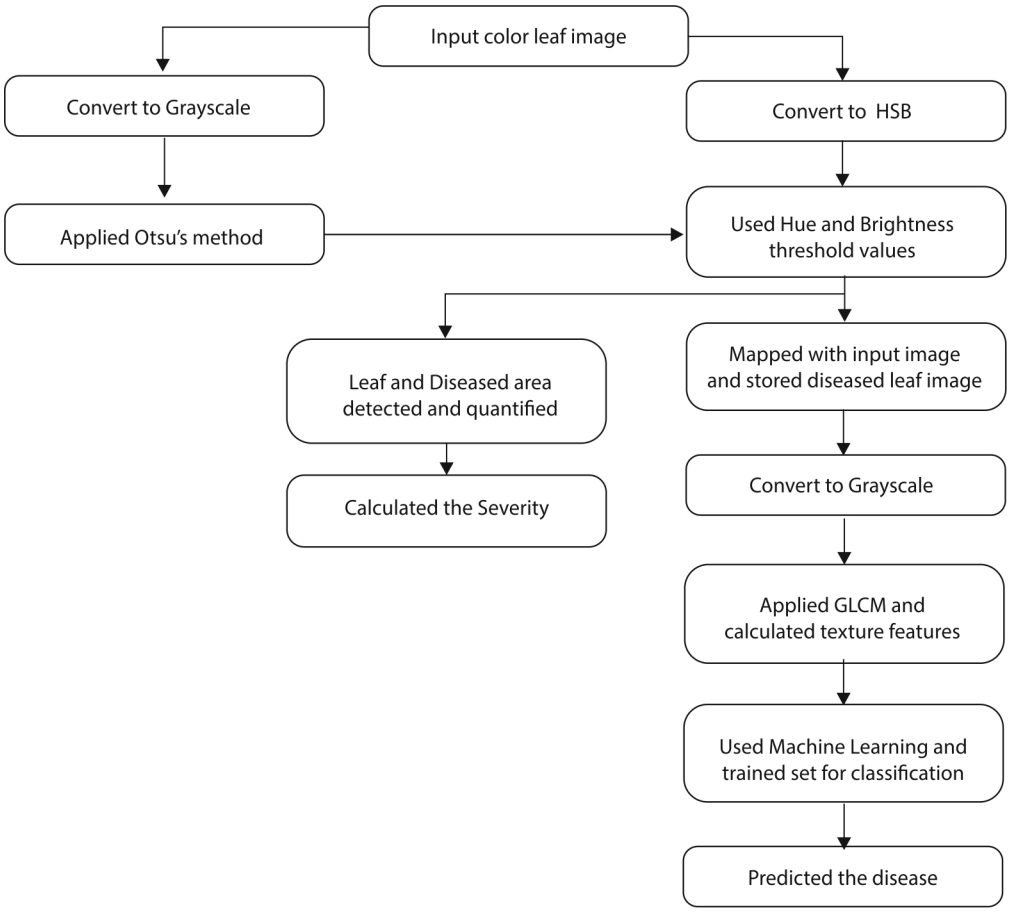


Fig. 1. Workflow of the proposed method

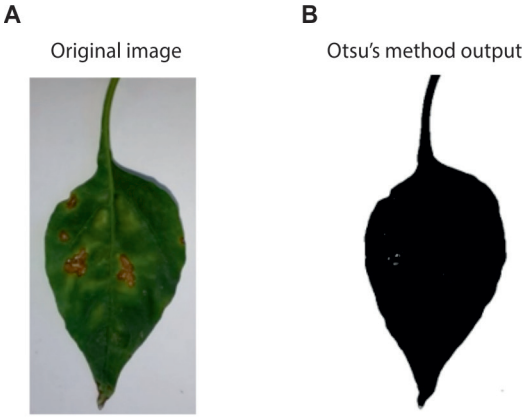


Fig. 2. Sample output obtained by Otsu's method

it is evident that the developed method was efficient and reliable.

Otsu's method simplifies the complexity of the problem of background noise. Also, it helps to detect the leaf area in a more precise manner (Fig. 2). This further demonstrates the consistency of results obtained under various resolutions of image capture.

With flash

To evaluate the precision and applicability of the proposed method, the images having different backgrounds were captured with a smartphone and with a flash. Since the outline of the leaf was detected, there were no challenges in leaf disease segmentation.

Without flash

Compared to flash-captured images, without flash-captured images produced more accuracy of detection.

Challenges

The selection of threshold values for background and foreground was based on the resolution of the image by Otsu's method. The challenge in the image below (Fig. 3) is the thin shadow around the leaf margin. These challenges were applicable to both images captured with smartphones and those from the PlantVillage repository dataset.

After isolating the diseased area, 11 feature sets namely Energy, Entropy, Contrast, Correlation, Homogeneity, Prominence, Shade, Variance, Angular Second Moment (ASM), Inverse Difference Moment (IDM) and Inertia were extracted. Table 2 shows values of the extracted features for different diseases of the leaf sample.

Table 1. Comparison of the results of the proposed method and APS Assess 2.2 and visual assessment

Original image	Proposed method		APS Assess 2.2 automatic assessment and visual assessment	
	diseased area	severity	APS Assess diseased area (red color)	severity
		Area: 64,839 pixels Diseased: 3494 pixels Severity: 5.39%		Area: 65,433 pixels Diseased: 4780 pixels Severity: 7.31% Manual grading Severity: 8.00%

Fig. 3. Original image (A) and diseased area segmented image with shadow (B)

In the training phase, features of 118 leaf samples were extracted and fed to the machine learning algorithm. Three categories of machine learning classifiers, namely decision trees, rule-based and Bayes model were applied. The same training datasets were utilized for classification employing different classifier models (decision trees and rule-based, and Bayes' theorem model). The classification accuracy was higher (98%) in decision trees and rule-based models. For evaluating these models, we employed 10-fold cross-validation technique on the trained disease datasets. The 10-fold cross validation technique splits the disease data into 10 sets of size N/10 of which it utilizes nine sets for training and one set for testing. The cross--validation procedure was repeated 10 times to arrive at a mean prediction result. Table 3 presents the accuracy of disease classification by machine learning classifiers.

The classifier algorithm's performance was compared with ROC curves. Figure 4 shows the comparison ROC curves of the proposed classifiers for late blight class. The detailed accuracy of these results by each class and confusion matrix were provided in supplementary Tables S1–S6, and S1A–S6A, respectively.

Based on the trained disease dataset, the experimental results (Table 3) show that J48 and DecisionTable correctly classified instances at a level of 98.31% which is higher than the other classifier-based models. The correctly classified instances for other models namely RandomForest, RandomTree and HoeffdingTree were at a level of 95.76, 91.53 and 87.29%, respectively, which is higher than the Bayes' theorem (NaiveBayes) model (86.44%).

The final contribution of this research was the accurate identification of the disease. Figure 5 depicts the workflow of the prediction of disease. The given input

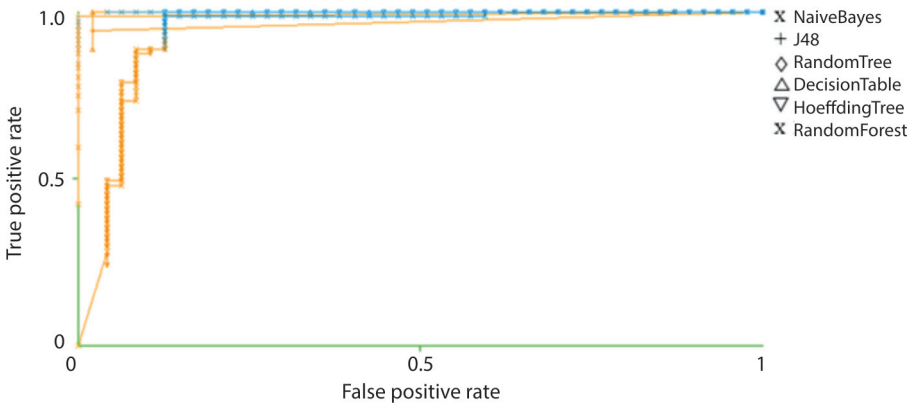
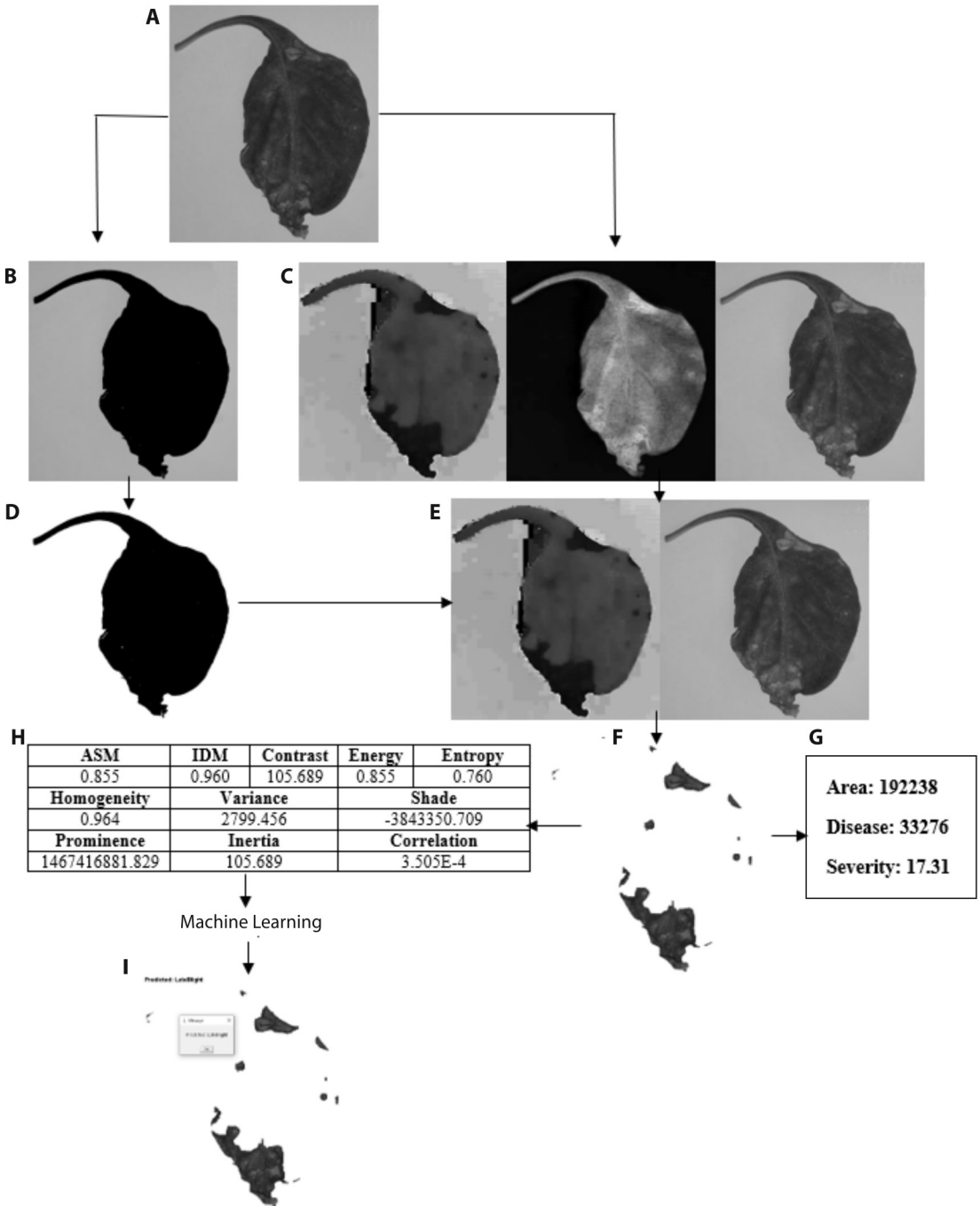


Fig. 4. Comparison Receiver Operating Characteristic (ROC) curves of the proposed classifiers for late blight class



ASM	IDM	Contrast	Energy	Entropy
0.855	0.960	105.689	0.855	0.760
Homogeneity	**Variance**		**Shade**	
0.964	2799.456		-3843350.709	
Prominence	**Inertia**		**Correlation**	
1467416881.829	105.689		3.505E-4	

Fig. 5. Workflow of the proposed disease identification with a sample image

Table 2. Values of the extracted features for different diseases of a leaf sample

Disease name	Energy	Entropy	Contrast	Correlation	Homo-geneity	Prominence	Shade	Variance	ASM	IDM	Inertia
Late blight	0.945	0.306	35.094	0.000842	0.989	770,247,800.1	–18,855,050.034	1169.553	0.945	0.988	35.094
Early blight	0.835	0.832	85.924	0.000304	0.964	1,686,769,285.0	–4,407,830.514	3245.158	0.835	0.960	85.924
Bacterial spot	0.98	0.125	43.781	0.002	0.995	269,879,838.2	–648,838.651	409.731	0.980	0.994	43.781

ASM – Angular Second Moment, IDM – Inverse Difference Moment

Table 3. Accuracy of classification by machine learning classifiers

Features	Classifiers	Accuracy [%]
GLCM	J48	98.31
	RandomForest	95.76
	RandomTree	91.53
	HoeffdingTree	87.29
	DecisionTable	98.31
	NaiveBayes	86.44

GLCM – Gray-Level Co-occurence Matrix

image went through the quantification and feature extraction of the diseased area. The filtered classifier matches the features with the existing trained feature sets for prediction. From this matching, the diseased type of the input image was predicted. The type of predicted disease was displayed on the disease-segmented image which was accurate.

Discussion

The software tool ImageJ was employed in the current study. Adaptability and extensibility of ImageJ are the primary strengths of the software (Schroeder *et al.* 2021). The implemented software is platform-independent and runnable via online or as a downloadable application which makes it simple and user-friendly. The proposed method produces results quickly and the results have a high degree of accuracy.

Digital tools can play a very important role in foliar disease quantification, classification, and identification. These have the potential to simplify the process and overcome the challenges including visual bias in the measurement by the human expert, time constraints, and repetitive tasking, etc. (Hasanaliyeva *et al.* 2022). Most studies referenced in this investigation placed less emphasis on measuring the severity of the disease. Hence, the current experimental study focused on addressing the above gaps and challenges.

Thresholding improves the robustness of the segmentation results and simplifies the efficient procedure (Akay *et al.* 2022). Hence, the proposed method employed Otsu's method and Thresholding. The Otsu's method performs well by separating the leaf from the background area of the image. These results help to measure the precise area of the leaf in terms of pixels. The total leaf area was obtained from Otsu's method of analysis using the threshold values of Hue and Brightness, the diseased area was isolated. The disease severity was measured from the number of pixels representing the disease symptoms. The severity measurements were mainly in alignment with the visual assessment scale and APS Assess 2.2 tool.

The major challenge of this proposed method was the leaf's shadow. The image captured through a smartphone may have a shadow, particularly at the edges, which depends on the background. The same kind of challenge was reported by Barbedo (2016). A darker background image does not have a shadow whereas an image with a light background will tend to have a shadow. For image analysis, an image with shadow has the challenge of differentiating the foreground from the background. The dynamic selection of foreground and background thresholds reduces the challenges.

The precision and applicability of the proposed method were tested using images with different lighting conditions and backgrounds. From a variety of leaf images, the diseased areas were isolated and GLCM 11 features were extracted and trained in the machine learning algorithms of trees, rules, and Bayes classifier model. The trained dataset was cross-validated by 10-fold cross-validation. By employing this software tool, the results on decision trees and rule-based classifiers achieved 98.31% accuracy, whereas NaiveBayes's classifier achieved 86.44% accuracy. Previous studies (Puspha Annabel *et al.* 2019; Anjna *et al.* 2020; Applalanaidu and Kumaravelan 2021; Chakole *et al.* 2022) on the quantification and classification/identification of plant diseases achieved similar levels of accuracy in decision tree classifiers.

The efficiency of the proposed method was evaluated through various classifier models. Each classifier model employed various criteria for evaluation. The decision tree classifiers J48, RandomForest, RandomTree and HoeffdingTree provided classification

accuracy of 98.31, 95.76, 91.53 and 87.29%, respectively. Bayes model of NaiveBayes provided classification accuracy of 86.44% while the Rule-based model DecisionTable provided 98.31% accuracy of classification. The true positive and false positive rates and confusion matrix of all the classifiers were provided in the supplementary Tables: S1, S1A, S2, S2A, S3, S3A, S4, S4A, S5, S5A, S6 and S6A. In the decision tree, every branch was built using a hierarchy technique which can be viewed as an "if-else" expression. By dividing the dataset into subsets based on the most important features, the subsets formed branches called decision tree leaves (Kotsiantis 2007).

The random forest algorithm generates a collection of decision trees and is a common type of ensemble method which aggregates results from multiple predictors. Random forest additionally utilizes a bagging technique that allows each tree to be trained on a random sampling of the original dataset and takes the majority vote from trees. Based on the investigation and the experimental results, Bezabh *et al.* (2023) recommended hybrid Machine Learning technique classifiers such as decision trees, random forests, and SVM. In this work, GLCM was employed instead of SVM. In contrast, the random tree considers a given number of random features at each node, performing no pruning. Hoeffding tree is the incremental decision tree algorithm and offers to create splits based on information gain or the Gini index. Predictions of the leaves of the tree can be made by either majority class or naïve Bayes models (Witten *et al.* 2016). DecisionTable builds a decision table classifier. It evaluates feature subsets using best-first search and can perform cross-validation. Based on the same set of features, an option determines the class for each instance that is not covered by a decision table entry using the nearest-neighbor method rather than the table's global majority (Witten *et al.* 2016). NaiveBayes is based on Bayes' theorem and an approach to calculate conditional probability based on prior knowledge, and the naive assumption that each feature is independent to each other. The biggest advantage of Naive Bayes is that, while most machine learning algorithms rely on large amounts of training data, it performs relatively well even when the training data size is small. Gaussian Naive Bayes is a type of Naive Bayes classifier that follows the normal distribution (Kotsiantis 2007).

All the above discussed classifier models were tested and achieved results with high levels of accuracy, which make them suitable to assist with the identification of diseases. The identification of the disease of the input image was performed with the same quantification and features extracted from the diseased area. These features matched with the existing trained set and filtered classifier were applied to find the exact disease of the input image. The proposed method performed well in the accurate identification of disease; furthermore, the result was obtained very quickly.

This research has outperformed the classification results of J48 decision tree (98.31%) and rule-based classifier (98.31%). Comparable to the current study results, the RandomForest classifier by Local Binary Pattern (LBP) and Visual Geometry Group network (VGG-16) features obtained 99.75% accuracy for the bacterial spot disease in bell pepper as reported by Bhagat *et al.* (2023). CNN for measurement of bell pepper leaf bacterial spots by VGG16 and VGG19 obtained 97 and 96%, respectively (Das 2023). However, the above studies did not achieve multiple measurements including disease severity, classification, and identification of diseases at a point in time. However, this has been achieved utilizing the developed software tool in the current study.

This proposed digital tool helps the user/farmer to quickly obtain outputs of the given leaf image input of the leaf area, diseased area (in pixels), diseased severity (in percentage), GLCM texture features of the diseased area, and disease type. Visual outputs of an isolated diseased leaf area and the type of disease were displayed in the alert message which will be highly beneficial to farmers for disease identification as well as for undertaking precise management methods.

Conclusions

Sweet pepper foliar blight disease intensity measurements could benefit from digitally captured plant images. The method was tested on a variety of Smartphone and PlantVillage datasets. The digital tool's flexibility and platform independence enables users to quickly assess the disease severity and predict the type of disease. This would further alert farmers to take the necessary precautionary measures for disease management. Beyond this, the developed software tool in ImageJ can be very useful for plant breeders and pathologists who regularly use physical scoring for estimating the disease severity in a crop/line or for measurement of disease resistance reactions in vivo.

Acknowledgements

We wish to thank the Campus Research and Publication Fund Committee, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago for funding this study and the farmers of Macoya, Orange Grove and Aranguez for their assistance in the collection of plant samples.

References

Akay R., Saleh R.A.A., Farea S.M.O., Kanaan M. 2022. Multilevel thresholding segmentation of color plant disease images using metaheuristic optimization algorithms. Neural Computing and Applications 34: 1161–1179. DOI: https://doi.org/10.1007/s00521-021-06437-1

Alheeti A.A. M., Farhan M.A., Al-Saad L.A., Theer R.M. 2021. IOP Conference Series: Earth and Environmental Science 761 012030. DOI: 10.1088/1755-1315/761/1/012030

Ali O., Ramsubhag A., Jayaraj J. 2019. Biostimulatory activities of *Ascophyllum nodosum* extract in tomato and sweet pepper crops in a tropical environment. PLOS ONE 14 (5): e0216710. DOI: https://doi.org/10.1371/journal.pone.0216710

Anjna, Sood M., Singh P.K. 2020. Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis. Procedia Computer Science 167: 1056–1065. DOI: https://doi.org/10.1016/j.procs.2020.03.404

Applalanaidu M.V., Kumaravelan G. 2021. A review of machine learning approaches in plant leaf disease detection and classification. p. 716–724. In: Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). 4–6 February 2021, Tirunelveli, India. DOI: 10.1109/ICICV50876.2021.9388488

Barbedo J.G.A. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering 144: 52–60. DOI: https://doi.org/10.1016/j.biosystemseng.2016.01.017

Bezabh Y.A., Salau A.O., Abuhayi B.M., Mussa A.A., Ayalew A.M. 2023. CPD-CCNN: classification of pepper disease using a concatenation of convolutional neural network models. Scientific Reports 13: 15581. DOI: https://doi.org/10.1038/s41598-023-42843-2

Bhagat M., Kumar D., Kumar S. 2023. Bell pepper leaf disease classification with LBP and VGG-16 based fused features and RF classifier. International Journal of Information Technology 15: 465–475. DOI: https://doi.org/10.1007/s41870-022-01136-z

Bock C.H., Chiang K.S., Del Ponte E.M. 2022. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. Tropical Plant Pathology 47: 25–42. DOI: https://doi.org/10.1007/s40858-021-00439-z

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). 23–24 December 2022, Bhopal, India. DOI: 10.1109/CCET56606.2022.10080352

Chinnadurai C., Ramkissoon A., Rajendran R., Ramsubhag A., Jayaraj J. 2018. Disease status of tomatoes in the Southern Caribbean. Tropical Agriculture 95: 29–35.

Das P.K. 2023. Leaf disease classification in bell pepper plant using VGGNet. Journal of Innovative Image Processing 5 (1): 36–46. DOI: 10.36548/jiip.2023.1.003

Das S., Pattanayak S., Behera P.R. 2022. Application of machine learning: a recent advancement in plant diseases detection. Journal of Plant Protection Research 62 (2): 122–135. DOI: https://doi.org/10.24425/jppr.2022.141360

Davis J., Goadrich M. 2006. The relationship between Precision-Recall and ROC curves. p. 233–240. In: Proceedings of the 23rd International Conference on Machine Learning (ICML '06). Association for Computing Machinery, New York, NY, USA. DOI: https://doi.org/10.1145/1143844.1143874

Francisco M., Ribeiro F., Metrôlho J., Dionísio R. 2023. Algorithms and models for automatic detection and classification of diseases and pests in agricultural crops: A systematic review. Applied Sciences 13 (8): 4720. DOI: https://doi.org/10.3390/app13084720

Hasanaliyeva G., Si Ammour M., Yaseen T., Rossi V., Caffi T. 2022. Innovations in disease detection and forecasting: A digital roadmap for sustainable management of fruit and foliar disease. Agronomy 12: 1707. DOI: https://doi.org/10.3390/agronomy12071707

Hassan S.M., Amitab K., Jasinski M., Leonowicz Z., Jasinska E., Novak T., Maji A.K. 2022. A survey on different plant diseases detection using machine learning techniques. Electronics 11: 2641. DOI: https://doi.org/10.3390/electronics11172641

Kotsiantis S.B. 2007. Supervised machine learning: A review of classification techniques. Informatica 31: 249–268.

Kulmaganbetov M., Bevan R.J., Anantrasirichai N., Achim A., Erchova I., White N., Albon J., Morgan J.E. 2022. Textural feature analysis of optical coherence tomography phantoms. Electronics 11 (4): 669. DOI: https://doi.org/10.3390/electronics11040669

Li C., Adhikari R., Yao Y., Miller A.G., Kalbaugh K., Li D., Nemali K. 2020. Measuring plant growth characteristics using smartphone based image analysis technique in controlled environment agriculture. Computers and Electronics in Agriculture 168: 105123. DOI: https://doi.org/10.1016/j.compag.2019.105123

Mall P.K., Singh P.K., Yadav D. 2019. GLCM based feature extraction and medical x-ray image classification using machine learning techniques. p. 1–6. In: Proceedings of the IEEE Conference on Information and Communication Technology. 6–8 December 2019, Allahabad, India. DOI: 10.1109/CICT48419.2019.9066263

Naegele R.P., Granke L.L., Fry J., Hill T.A., Ashrafi H., Van Deynze A., Hausbeck M.K. 2017. Disease resistance to multiple fungal and oomycete pathogens evaluated using a recombinant inbred line population in pepper. Phytopathology 107 (12): 1522–1531. DOI: https://doi.org/10.1094/PHYTO-02-17-0040-R

Nyarko B.N.E., Bin W., Jinzhi Z., Odoom J. 2023. Tomato fruit disease detection based on improved single shot detection algorithm. Journal of Plant Protection Research 63 (4): 405–417. DOI: https://doi.org/10.24425/jppr.2023.146877

Otsu N. 1979. A threshold selection method from gray-level histograms. p. 62–66. In: Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics 9 (1). DOI: 10.1109/TSMC.1979.4310076

Priyanka, Kumar D. 2020. Feature extraction and selection of kidney ultrasound images using GLCM and PCA. Procedia Computer Science 167: 1722–1731. DOI: https://doi.org/10.1016/j.procs.2020.03.382

Puspha Annabel L.S., Annapoorani T., Deepalakshmi P. 2019. Machine learning for plant leaf disease detection and classification – A review. p. 0538–0542. In: Proceedings of the International Conference on Communication and Signal Processing (ICCSP). 4–6 April 2019, Chennai, India. DOI: 10.1109/ICCSP.2019.8698004

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.

Ramsubhag A., Jayaraj J. 2018. Review paper rapid molecular methods for plant disease diagnosis. Tropical Agriculture 95: 166–177.

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

Saleem M.H., Potgieter J., Arif K.M. 2019. Plant disease detection and classification by deep learning. Plants 8 (11): 468. DOI: https://doi.org/10.3390/plants8110468

Sapkal A.T., Kulkarni U.V. 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.

Sarker I.H., Kayes A.S.M., Watters P. 2019. Effectiveness analysis of machine learning classification models for predicting

personalized context-aware smartphone usage. Journal of Big Data 6: 57. DOI: https://doi.org/10.1186/s40537-019-0219-y

Schroeder A.B., Dobson E.T.A., Rueden C.T., Tomancak P., Jug F., Eliceiri K.W. 2021. The ImageJ ecosystem: Open-source software for image visualization, processing, and analysis. Protein Science 30: 234–249. DOI: https://doi.org/10.1002/pro.3993

Shafique H., Viqar S., Ehteshamul-Haque S., Athar M. 2016. Management of soil-borne diseases of organic vegetables. Journal of Plant Protection Research 56 (3): 221–230. DOI: 10.1515/jppr-2016-0043

Shi T., Liu Y., Zheng X., Hu K., Huang H., Liu H., Huang H. 2023. Recent advances in plant disease severity assessment using convolutional neural networks. Scientific Reports 13: 2336. DOI: https://doi.org/10.1038/s41598-023-29230-7

Singh A.K., Sreenivasu S., Mahalaxmi U.S., Sharma H., Patil D.D., Asenso E. 2022. Hybrid feature-based disease detection in plant leaf using convolutional neural network, Bayesian Optimized SVM, and random forest classifier. Journal of Food Quality. 2845320: 16. DOI: https://doi.org/10.1155/2022/2845320

Singh V., Misra A.K. 2017. Detection of plant leaf diseases using image segmentation and soft computing techniques, Information Processing in Agriculture 4 (1): 41–49. DOI: https://doi.org/10.1016/j.inpa.2016.10.005

Witten I.H., Frank E., Hall M. A., Pal C.J. 2016. Data Mining, Practical Machine Learning Tools and Techniques. 4th ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Yogeshwari M., Thailambal G. 2020. Automatic Segmentation of Plant Leaf Disease Using Improved Fast Fuzzy C Means Clustering and Adaptive Otsu Thresholding (IFFCM-AO) Algorithm. European Journal of Molecular & Clinical Medicine 7 (3): 5447–5462.

Zeng Y., Zhao Y., Yu Y., Tang Y., Tang Y. 2021. IOP Conference Series: Earth and Environmental Science 792 012001. DOI: https//doi.org/10.1088/1755-1315/792/1/012001

www

https://agriculture.gov.tt/wp-content/uploads/2017/11/how-to-grow-sweet-peppers.pdf

http://files-do-not-link.udc.edu/docs/causes/online/Pepper%2010.pdf

https://agriculture.gov.tt/category/publications/manuals/

https://www.kaggle.com/datasets/emmarex/plantdisease

https://my.apsnet.org/APSStore/Product-Detail.aspx?WebsiteKey=2661527A-8D44-496C-A730-8CFEB6239BE7&iProductCode=43696

Journal of Plant Protection Research eISSN 1899-007X

ORIGINAL ARTICLE

Sweet pepper foliar diseases quantification and identification using an image analysis tool

Vijayanandh Rajamanickam[1]*, Adesh Ramsubhag[2], Jayaraj Jayaraman[2]

[1] Department of Computing and Information Technology, Faculty of Science and Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago
[2] Department of Life Science, Faculty of Science and Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago

Vol. 65, No. 1: 100–102, 2025

DOI: 10.24425/jppr.2025.153821

Received: April 25, 2024
Accepted: July 25, 2024
Online publication: March 25, 2025

*Corresponding address:
Vijayanandh.Rajamanickam@sta.uwi.edu

Responsible Editor:
Piotr Iwaniuk

SUPPLEMENTARY MATERIAL

The authors are fully responsible for both the content and the formal aspects of the supplementary material. No editorial adjustments were made.

Table S1. True positive (TP) and false positive (FP) rate of J48 classifiers

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	0.986	0.000	1.000	0.986	0.993	0.983	0.993	0.994	Late blight
	0.968	0.011	0.968	0.968	0.968	0.956	0.978	0.945	Early blight
	1.000	0.010	0.941	1.000	0.970	0.965	0.995	0.941	Bacterial spot
Weighted Avg.	0.983	0.004	0.984	0.983	0.983	0.973	0.989	0.974	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S1A. Confusion matrix for J48

a	b	c	Classified as
70	1	0	a = Late blight
0	30	1	b = Early blight
0	0	16	c = Bacterial spot

Table S2. True positive (TP) and false positive (FP) rate of RandomForest tree classifier

	TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	0.972	0.000	1.000	0.972	0.986	0.965	1.000	1.000	Late blight
	0.968	0.046	0.882	0.968	0.923	0.896	0.995	0.988	Early blight
	0.875	0.010	0.933	0.875	0.903	0.889	0.996	0.976	Bacterial spot
Weighted Avg.	0.958	0.013	0.960	0.958	0.958	0.937	0.998	0.993	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S2A. Confusion matrix for RandomForest

a	b	c	Classified as
69	2	0	a = Late blight
0	30	1	b = Early blight
0	2	14	c = Bacterial spot

Table S3. True positive (TP) and false positive (FP) rate of RandomTree classifier

	TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	0.944	0.021	0.985	0.944	0.964	0.914	0.961	0.964	Late blight
	0.839	0.046	0.867	0.839	0.852	0.801	0.896	0.769	Early blight
	0.938	0.049	0.750	0.938	0.833	0.811	0.944	0.712	Bacterial spot
Weighted Avg.	0.915	0.032	0.922	0.915	0.917	0.870	0.942	0.878	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S3A. Confusion matrix for RandomTree

a	b	c	Classified as
67	3	1	a = Late blight
1	26	4	b = Early blight
0	1	15	c = Bacterial spot

Table S4. True positive (TP) and false positive (FP) rate of HoeffdingTree classifier

	TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	0.901	0.128	0.914	0.901	0.908	0.771	0.942	0.923	Late blight
	0.871	0.103	0.750	0.871	0.806	0.734	0.926	0.926	Early blight
	0.750	0.000	1.000	0.750	0.857	0.850	0.975	0.931	Bacterial spot
Weighted Avg.	0.873	0.104	0.883	0.873	0.874	0.772	0.942	0.925	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S4A. Confusion Matrix for HoeffdingTree

a	b	c	Classified as
64	7	0	a = Late blight
4	27	0	b = Early blight
2	2	12	c = Bacterial spot

Table S5. True positive (TP) and false positive (FP) rate of NaiveBayes

	TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	0.887	0.128	0.913	0.887	0.900	0.755	0.935	0.922	Late blight
	0.871	0.115	0.730	0.871	0.794	0.717	0.921	0.867	Early blight
	0.750	0.000	1.000	0.750	0.857	0.850	0.972	0.923	Bacterial spot
Weighted Avg.	0.864	0.107	0.877	0.864	0.866	0.758	0.936	0.908	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S5A. Confusion matrix for NaiveBayes

a	b	c	Classified as
63	8	0	a = Late blight
4	27	0	b = Early blight
2	2	12	c = Bacterial spot

Table S6. True positive (TP) and false positive (FP) rate of DecisionTable classifiers

	TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area	Class
	1.000	0.021	0.986	1.000	0.993	0.982	0.988	0.985	Late blight
	0.968	0.011	0.968	0.968	0.968	0.956	0.973	0.941	Early blight
	0.938	0.000	1.000	0.938	0.968	0.964	0.990	0.959	Bacterial spot
Weighted Avg.	0.983	0.016	0.983	0.983	0.983	0.973	0.984	0.970	

MCC – Matthews Correlation Coefficient; ROC – Receiver Operating Characteristic; PRC – Precision-Recall Curve

Table S6A. Confusion matrix for DecisionTable

a	b	c	Classified as
71	0	0	a = Late blight
1	30	0	b = Early blight
0	1	15	c = Bacterial spot