Plant Leaf Disease Recognition Using Histogram Based Gradient Boosting Classifier

Chapter · February 2021

DOI: 10.1007/978-3-030-68154-8_47

2 authors:

Syed Mohammod Minhaz Hossain
Premier University

Deb Kaushik
Chittagong University of Engineering & Technology

Plant Leaf Disease Recognition using Histogram Based Gradient Boosting Classifier

Syed Md. Minhaz Hossain[⋆†,1] and Kaushik Deb[⋆,2]

[⋆]Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh
[†]Premier University, Chattogram-4000, Bangladesh
[1]minhazpuccse@gmail.com,
[2]debkaushik99@cuet.ac.bd

Abstract. Plant leaf disease (PLD) recognition's current techniques lack proper segmentation and locating similar disorders due to overlapping features in different plants. For this reason, we propose a framework to overcome the challenges of tracing Region of Interest(ROI) under different image backgrounds, uneven orientations, and illuminations. Initially, modified Adaptive Centroid Based Segmentation (ACS) is applied to find K's optimal value from PLDs and then detect ROIs accurately, irrespective of the background. Later, features are extracted using a modified Histogram Based Local Ternary Pattern (HLTP) that outperforms for PLDs with uneven illumination and orientation, capitalizing on linear interpolation and statistical threshold in neighbors. Finally, Histogram-based gradient boosting is utilized to reduce biasness for similar features while detecting disorders. The proposed framework recognizes twelve PLDs having an overall accuracy of 99.34% while achieves 98.51% accuracy for PLDs with more than one symptom, for instance, fungal and bacterial symptoms.

Keywords: Plant Leaf Disease Recognition, Modified Adaptive Centroid-based Segmentation, Histogram-based Local Ternary Pattern, Histogram-based Gradient Boosting Classifier

1 Introduction

Diagnosis and detection of various plant diseases through leaves' symptoms are complicated for farmers and agronomists. It creates complexity due to various symptoms in the same plant and similar symptoms in different plant diseases. This complicated task may cause misleading to conclude the status of plants and their proper treatments. Automated plant diagnosis using the mobile application through the real field's capturing image helps the agronomist and farmers make better decisions on plant health monitoring. Due to the growth of the Graphical Processing Unit(GPU) embedded processors, Machine Learning, Artificial Intelligence makes it possible to incorporates new models and methods to detect the appropriate ROIs and hence, identify the plant diseases correctly. However, memory space(number of parameters) is still in consideration for mobile-based PLD recognition.

Machine learning techniques mainly investigate localizing the ROIs, feature extraction, and classification, such as in [9, 11]. Limitations of learning-based techniques are: **a.** lack of sensitivity to proper segmentation in different image backgrounds and under different capture conditions and **b.** failing to trace similar symptoms in different plant disorders.

The recent trend of Convolutional Neural Network (CNN) performs complex patterns using a large number of data.The state-of-the-art architecture of convolutional neural network (CNN) such as, VGG in [5, 13],GoogleNet in [8], ResNet50, ResNet101, ResNet152, Inception V4 in [13], Student-teacher CNN in [4], AlexNet in [3, 5, 8] and DenseNet in [13] are applied in recognizing PLDs.Though CNN achieves better results, tuning the parameters depends on the CNN architecture, to an extend. Furthermore, space(memory) limitation, especially in handheld devices, to support such a high volume of network parameters is not considered. Last but not least, when exposed to a new dataset, CNN fails to generalize, and its accuracy drops down drastically [5, 8].

Our primary emphasis is to modify the K means clustering to overcome the limitations of lack of sensitivity to proper segmentation in [10] and remove the noises, including unwanted objects beside the plant leaf or leaves. The modified ACS suggested here to find optimal K such that it can cause a.segment of the appropriate disease symptoms in different background images and uneven illuminations and b. identify the disorders having similar symptoms. This work also employs modified HLTP to alleviate the limitation of the traditional local ternary pattern (LTP), which outperforms the uneven illumination and orientation counterparts. Detecting ROIs from complex backgrounds and extracting histogram features under various health states generalizes better when exposed to the unspecified dataset. As memory space is a significant factor for mobile devices, we propose a PLD framework to recognize PLDS using histogram-based gradient boosting classifier instead of CNN. It improves PLDs' recognition rate than various machine learning algorithms for histogram features and reduces the memory cost compared to CNN.

The remaining paper is demonstrated as follows. Section 2 depicts the literature review including the related works; proposed framework for recognizing plant leaf diseases is described in Section 3; experiments, performance evaluation and observations are presented in Section 4; and lastly, conclusion of this paper is illustrated in Section 5.

2 Related Work

Plant/crop-related machine learning-based works are categorized into PLD recognition, prediction production of crop-based on weather parameters, and post-harvest monitoring of grains in [1]. A study has been conducted to predict the co-relations between the weather parameters(temperature, rainfall, evaporation, and humidity) and crop production in [2]. For this, the authors in [2] design a fuzzy rule-based system using the Takagi Sugeno-Kang approach. Besides, the machine learning and image processing based PLD recognition framework have several parts; the localization of symptoms of the disease (region of interest), feature extraction, and classification. Before localization, image enhancement technique is used in [11]. However, it is not always mandatory to improve the intensity of plant leaf images. Plant image intensities are changing under different capture conditions and uneven illumination. Two conditions are used based on statistical features to trace the changing pattern of plant images. It makes robust PLD detection and avoids unnecessary image enhancement.

GrabCut algorithm in [9], the Genetic Algorithm in [11], k-means clustering in [10] has been used to get proper disease region in leaf image. Besides, in [10], a couple of limitations of lack of sensitivity to proper segmentation in K-means clustering due to improper initialization of K and localizing multiple disorders in a PLD image. In [11], there are some misclassifications between the two leaf spot conditions because of similar features. Our modified ACS overcomes the limitations of lack of sensitivity of segmentation using the auto initialization of K from the plant leaf images. Also, it makes the segmentation effective under different critical environments and in different backgrounds.

The texture feature has been extracted by histogram-based local binary pattern (LBP) in [9], by color co-occurrence matrix (local homogeneity, contrast, cluster shade, energy, and cluster prominence)in [11]. In [9], histogram-based local binary pattern extracts the better feature under different orientations and uneven illuminations. We use a feature extraction method HLTP using linear interpolation and dynamic threshold. The neighbors found using interpolation make the feature extraction method sensitive to orientations. The variation of the gray level of neighbors makes it invariant to illumination in recognizing PLD.

Moreover, multiple classifiers have been used to recognize the correct PLD in various works. One Class Support Vector Machine (OCSVM) is used in [9], and SVM is used in [11] for recognizing PLD. Further, Minimum Distance Criterion (MDC) is used in [11].Though in all of the works, better accuracy is achieved, there still is a lack of proof in recognizing better in case of similar symptoms in different disorders.

Also, there are many works for recognizing various plant diseases using CNN. PLD recognition frameworks using the CNN model still have some limitations. These limitations have an impact on the performance of CNN models. Some of the works are restricted to plain backgrounds, e.g., [5, 8, 13] and inconsistent with image capturing conditions of not doing data augmentation in [7]. Finally, sometimes, plant leaf diseases have a generalization problem in an independent dataset [5, 8].

Using the ensemble learning classifiers, we can reduce the biasness of classifiers and improve accuracy than machine learning. Also, we can reduce the parameters than the state-of-the-art CNN PLD recognition models. Though random forest takes less time to build trees, gradient boosting classifiers are better in the benchmark result. Especially for histogram features, histogram-based gradient boosting classifiers perform well in considering memory cost and recognition rate than gradient boosting classifier.

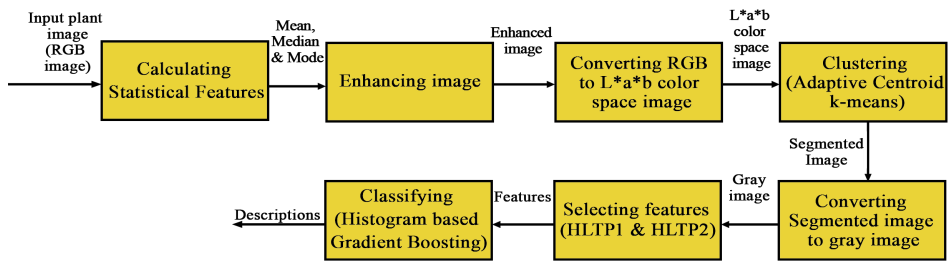

Fig. 1. The proposed framework for recognizing plant leaf disease.

We can conclude that auto initialization in this framework's segmentation phase overcomes lacking sensitivity to proper segmentation in [10] using modified Adaptive Centroid Based Segmentation (ACS). The automatic initialization of K defined using ACS can effectively detect changes in image characteristics for different orientations and illuminations and improve generalization. This paper also explores histogram-based local ternary patterns (HLTP) to alleviate the limitation of the traditional local ternary pattern (LTP), outperforms in the uneven illumination and orientation. Finally, histogram-based gradient boosting classifier is used to classify PLD because of the classification phenomena of histogram over features. This classifier is more suitable than CNN in considering restricted memory devices like mobile. Besides, histogram-based features make this framework useful to recognize the health status of newly-added plant images, increasing the generalization. So, accuracy never falls in the newly added diverse plant image, and this phenomenon overcomes the limitation of drastic fall in the validation of CNN with new plant leaf images in [5, 8].

3 Proposed framework for recognizing plant leaf diseases

In this section, the proposed framework is demonstrated in detail. Initially, the disease recognition framework optionally enhances the plant leafs' RGB image, and then modified adaptive centroid-based segmentation(ACS) is applied to trace the ROIs. After that, features selection from the grayscale image is executed using a histogram-based local ternary pattern. At last, the plant leaf disease is classified using a histogram-based gradient boosting classifier. The proposed PLD recognition framework has been exhibited in Fig. 1.

3.1 Dataset

In the experiment, 403 images of size 256x256 pixels comprising eight different plants, such as rice, corn, potato, pepper, grape, apple, mango, and cherry, and twelve diseases are used to train the proposed framework. The images are collected from the PlantVillage dataset[1] except rice disease images. Rice disease images are gathered from the Rice diseases image dataset in Kaggle [2], the International Rice Research Institute (IRRI)[3] and Bangladesh Rice Research Institute (BRRI)[4].

We vary the image backgrounds among natural, plain, and complex to trace a disease properly in different backgrounds. Our framework includes six fungal diseases, two bacterial diseases, two diseases having both fungal and bacterial symptoms, one viral disease, and another one from a different category. Further, the framework considers various symptoms, such as small, massive, isolated, and spread. Twelve samples of eight plants are represented, considering different symptoms and image backgrounds, as shown in Fig. 2. For generalization, 235 independent (excluding the training dataset) images from twelve different classes are used during the test phase. Complete information regarding the plant leaf disease dataset is described in Table 1.

Table 1. Dataset description of recognizing plant leaf disease.

Health-wise condition	**Plant Type**	**Disease Samples**	**# of training images**	**# of test images**	**# of training images (Health-wise)**	**# of test images (Health-wise)**
Fungal	Rice	Blast	54	30	208	134
	Potato	Early-blight	42	39		
		Late-blight	21	10		
	Corn	Northern-blight	50	30		
	Mango	Sooty-mould	19	12		
	Cherry	Powdery-mildew	22	13		
Bacterial	Rice	Bacterial leaf-blight	65	30	115	60
	Pepper	Bacterial-spot	50	30		
Fungal/Bacterial	Rice	Sheath-rot	20	10	35	17
	Apple	Black-rot	15	7		
Virus	Rice	Tungro	10	5	10	5
Miscellenous	Grape	Black-measles	35	19	35	19
Total Images			**403**	**235**	**403**	**235**

3.2 Enhancing image

If images are not captured precisely due to hostile conditions, image enhancement is needed to increase the PLD image quality. The enhancement is optional as it depends on the magnitude of degradation. Two enhancement conditions have been used here using statistical features such as mean(μ), median(x'), and mode(M_0) of a plant leaf image. The first condition for image enhancement is devised as in Eq. 1.

$$\mu < x' < M_0 \tag{1}$$

[1] https://www.kaggle.com/emmarex/plantdisease

[2] https://www.kaggle.com/minhhuy2810/rice-diseases-image-dataset

[3] https://www.irri.org/

[4] http://www.brri.gov.bd/

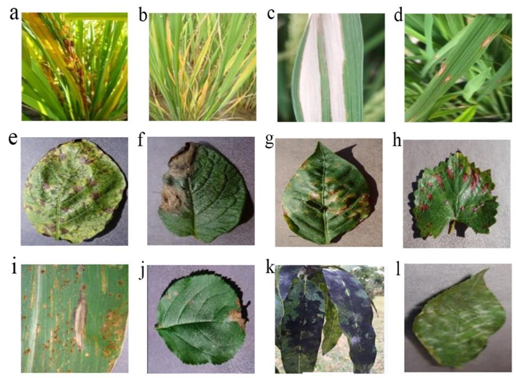

Fig. 2. Samples of plant leaf disease images under numerous health conditions in various backgrounds and having different symptoms: (a) Rice Sheath-rot(natural background, spread symptoms), (b) Rice Tungro(natural background, spread symptoms), (c) Rice Bacterial leaf-blight(complex background, spread symptoms), (d) Rice blast (complex background, isolated, small symptoms), (e) Potato Early-blight(plain background, isolated small symptoms), (f) Potato Late-blight(plain background, isolated small symptoms), (g) Pepper Bacterial-spot(plain background, small symptoms), (h) Grape Black-measles(plain background, small symptoms), (i) Corn Northern Leaf-blight(plain background, spread, spot symptoms), (j) Apple Black-rot(plain background, small symptoms), (k) Mango Sooty-mould(natural background, spread symptoms) and (l) Cherry Powdery-mildew(natural background, small symptoms).

According to Eq. 1, the image enhancement condition performs effectively in tracing ROIs with the identical background color as shown in Fig. 3($a_1 - c_2$). The second statistical condition for image enhancement is formulated as in Eq. 2). The second statistical condition is effective when there is a shadow of the leaf image on the background, as shown in Fig. 3($a_2 - c_4$). Otherwise, the leaf image is directly converted to the L*a*b color space image without enhancement.

$$\mu < x' > M_0 \tag{2}$$

3.3 Clustering by adaptive centroid based segmentation

The modified adaptive centroid-based segmentation (ACS) has been applied once the PLD image quality has been enhanced. In the beginning, the RGB (PLD) image space is converted to L*a*b color space for better perceptual linearity in differentiating colors. Conversion from RGB space to L*a*b color space significantly increases K-means clustering performance, especially when narrow distinguishes among symptoms colors in different plant leaf disorders. Differentiating among the color intensities having identical ROI color and background is non-trivial. Another challenge is distinguishing the basic color of ROIs in the same sunlight shade and shadowing the background. To overcome these challenges, we perform L*a*b color conversion before segmentation. In Fig. 3(c_2, c_4), improvements in segmentation is shown comparing with Fig. 3(c_1, c_3) having extra noise in the PLD RGB image. Our modified ACS focuses on initializing optimal K, automatically from the leaf image, to eliminate the limitation of lacking sensitivity of K in [10]. In traditional K-means, euclidean distance between each point and centroid has been calculated to check whether the point is in the same cluster. In the modified ACS, data points are investigated for eligibility by using a statistical threshold. After that, we calculate the distance between these eligible points and centroids, thus, comparatively reducing the effort to form clusters and restrict misclustering of data points. The statistical threshold (ST) value has been calculated by Eq. 3.

$$ST = \sqrt{\sum_{i=1}^{N}((X_i - C)^2)/N} \tag{3}$$

Where, X_i, C, and N stand for data points, the centroid of data points, and the total number of data points.

The automatic initialization of K defined using ACS can effectively detect image characteristics for different orientations and illuminations. ACS, also,increases the scalability of the proposed segmentation technique as shown in Fig. 3(c_2, c_4) and Fig. 3(c_1, c_3). A few examples under different circumstances, such as in the same colored reflection on ROIs, in the presence of the shadow behind the ROIs, overlapped blur images, the orientation of leaf images such as shrunk ROI and rotation, are as shown in Fig. 4($b_1 - b_5$).

3.4 Selecting features using HLTP

Once the PLD image's ROIs has been traced, the RGB segments are converted to grayscale images. Then HLTP has been applied to extract the features of leaf disease. We perform two approaches of feature extraction; namely HLTP-1 (8 pixels with radius 1) and HLTP-2 (8 pixels with radius 2). Firstly, four neighboring points are determined using Eq. 7 – Eq. 10. Other four points have been calculated by using linear interpolation coefficient for 45° in both HLTPs formulated using Eq. 11 – Eq. 14.

$$a = r - \sqrt{r} \tag{4}$$

$$b = 1 - a \tag{5}$$

$$f(n + a) = a * f(n + 1) + b * f(n) \tag{6}$$

$$d_0 = A(r_0, c_0 - r) - I \tag{7}$$

$$d_2 = A(r_0, c_0 + r) - I \tag{8}$$

$$d_4 = A(r_0 - r, c_0) - I \tag{9}$$

$$d_6 = A(r_0 + r, c_0) - I \tag{10}$$

$$d_1 = a * A(r_0 + r - 1, c_0 - r + 1) + b * A(r_0 + r, c_0 - r) - I \tag{11}$$

$$d_3 = a * A(r_0 + r - 1, c_0 + r - 1) + b * A(r_0 + r, c_0 + r) - I \tag{12}$$

$$d_5 = a * A(r_0 - r + 1, c_0 + r - 1) + b * A(r_0 - r, c_0 + r) - I \tag{13}$$

$$d_7 = a * A(r_0 - r + 1, c_0 - r + 1) + b * A(r_0 - r, c_0 - r) - I \tag{14}$$

Where, a and b are interpolation coefficients, and r is the radius. $A(r_0, c_0)$ stands for the matrix of PLD gray image I considering each neighbor of position (r_0, c_0). In Eq. 6, f(n+a) is the unknown pixel, f(n), and f(n+1) are two known pixels. Unknown pixels, as shown in Eq. 11 – Eq. 14, are formulated by Eq. 6 using Eq. 7 – Eq. 10. In Eq. 7 – Eq. 14, $d_0, d_1, d_2, d_3, d_4, d_5, d_6$, and d_7 are all neighboring pixels' derivatives. These derivatives are then put into 1 x 8 vector, d. 1 x 8 vector for each pixel P_i; where, i= 0,1,2,3,...., i.e total (m x n) x 8 matrix is found; where, m is he width of the plant leaf disease image and n is the height of plant leaf images.Then, mean threshold(MT) for each pixel P_i is determined using the surrounding eight pixels of this pixel. Then we get two values; one contains the lower pattern values and another contains the upper pattern values formulated in [12]. From that using histogram, we get two vectors of 1 x 256; one from lower values and another from upper values.

Traditional LTP has the limitation of uneven illumination and orientation in leaf image. In our modified HLTP, the mean threshold(MT) in [12] has been considered instead of a fixed threshold to overcome LTP's drawback. It handles the variation of the gray level of neighbors and makes it invariant to illumination. Using linear interpolation in determining directives helps increase the ability to extract features from different oriented plant leaf images. It outperforms, as shown in Fig. 3($d_2 - e_2$) and Fig. 3($d_4 - e_4$) compared to traditional LTP, as shown in Fig. 3($d_1 - e_1$) and Fig. 3($d_3 - e_3$). Our modified HLTP functions effectively in the same colored reflection on ROIs, in the shadow behind the ROIs, overlapped blur images, and the orientation of leaf images such as shrunk ROI and rotation, as shown in Fig. 4($c_1 - f_5$).

3.5 Classifying using Histogram-based gradient boosting classifier

Finally, a histogram-based gradient boosting classifier in [6] is used to recognize PLD. Feature vectors developed in HLTP-1 and HLTP-2 have been applied to a histogram-based gradient boosting classifier. Histogram-based gradient boosting classifier is used due to its benchmark accuracy using histogram features and computational cost compared to gradient boosting classifier.Unlike the gradient boosting classifier, in a histogram-based gradient boosting classifier, optimum splitting feature points are found by feature histogram. So, computational complexity reduces due to the histogram data structure. Moreover, it takes memory cost of $O(\#features * \#data * 1byte)$.

In histogram-based gradient boosting classifier, for every feature, we build the Histogram using 255 bins. Then gradient and hessian are calculated based on the loss. As we classify 12 PLDs, we use categorical cross-entropy. Trees are expanded based on the information gain from every feature. Information gain is evaluated using the gradient and hessian of each feature. The maximum depth for each is considered as 20. Each leaf includes a minimum of 30 samples of PLD images, and each tree has 30 leaf nodes. As, histogram-based boosting classifier (inspired by LightGBM)in [6], adds each best split tree level-wise, a new gradient and hessian are calculated to predict the next one. The boosting process are examined up to maximum iterations of 10 to 1000 and are learned with a learning rate from 0.1 to 1. However, our classification method gets a minimum loss function using a learning rate of 0.2 and a maximum iteration of 100. The best-tuned parameters used to train the histogram-based gradient boosting is represented in Table 2.

Table 2. Parameters used in histogram based gradient boosting classifier for plant leaf disease recognition.

Parameters	Value(s)
Loss function	Categorical cross-entropy
Max iterations	100
Minimum samples in leaf node	30
Maximum leaf nodes	30
Max depth	20
Max bins	255
Learning rate	0.2

4 Results and observations

In this section, the results of our experiments for recognizing plant leaf diseases are presented.

Environment The experiments for recognizing plant leaf disease are executed on Intel(R) Core i5 7200U 2.5 GHz with 4GB RAM. The proposed framework is implemented in Python with packages sklearn and MATLAB.

Dataset for training and test In this experiment, 403 images of eight plants of size 256 x 256 pixels, are used to train and 235 PLD images are used to test for twelve classes from different sources.The statistics of different PLD train and test images is shown in Table 3.

Effect of image enhancement conditions From Fig. 3, it is observed that without image enhancement, there are some noises in segmentation and also further have its impact on feature extraction in critical cases. Two image enhancement conditions of PLD images have been performed effectively in ROIs with the same color background, due to higher mode as shown in Fig. 3(c_2) and in a shadow of the leaf image on the background due to its higher median than other two statistical values, as shown in Fig. 3(c_4).

Effect of modified Adaptive Centroid Based Clustering The automatic initialization of K defined using ACS can effectively detect image characteristics for different orientations and illuminations. ACS, also,increases the scalability of our modified segmentation technique as shown in Fig. 3(c_2, c_4) and Fig. 3(c_1, c_3). In various critical circumstances, such as same-colored reflection on ROIs, when

Table 3. Dataset description according to the sources.

Source condition	Plant Type	Disease Samples	# of training images	# of test images	# of training images (Source-wise)	# of test images (Source-wise)
PlantVillage	Pepper	Bacterial-spot	50	30	254	160
	Potato	Early-blight	42	39		
		Late-blight	21	10		
	Corn	Northern-blight	50	30		
	Mango	Sooty-mould	19	12		
	Apple	Black-rot	15	7		
	Cherry	Powdery-mildew	22	13		
	Grape	Black-measles	35	19		
Kaggle	Rice	Blast	54	30	119	60
		Bacterial leaf-blight	65	30		
IRRI/BRRI	Rice	Sheath-rot	20	10	30	15
		Tungro	10	5		
Total Images			**403**	**235**	**403**	**235**

background and ROIs have the same color, ROIs in the natural background with shrunk, rotated, and overlapped blur images, modified ACS outperforms, as shown in Fig. 4($b_1 - b_5$).

Table 4. Comparison among the experiments using traditional K-means clustering, Local ternary pattern, modified adaptive centroid-based segmentation and modified histogram-based local ternary pattern.

Frameworks	**Accuracy**	**F1-score**
Traditional K-means clustering+ Local ternary pattern	90%	88%
Traditional K-means clustering+ HLTP	92.76%	89.4%
Modified ACS+ Local ternary pattern	94.89%	90.4%
Our PLD framework(Modified ACS+ HLTP)	99.34%	94.10%

Effect of our HLTP on feature extraction One thousand twenty-four(1024) histogram features (512 features of each HLTP-1 and HLTP-2) are extracted using HLTP. The dynamic mean threshold handles the variation of neighbors' gray level and makes it invariant to illumination. Linear interpolation in determining directives helps increase the ability to extract features from different oriented plant leaf images. It outperforms, as shown in Fig. 3($d_2 - e_2$) and Fig. 3($d_4 - e_4$) compared to traditional LTP, as shown in Fig. 3($d_1 - e_1$) and Fig. 3($d_3 - e_3$). HLTP functions effectively in the same colored reflection on ROIs, in the shadow behind the ROIs, overlapped blur images, and the orientation of leaf images such as shrunk ROI and rotation, as shown in Fig. 4($c_1 - f_5$). From Table 4, it is observed that our proposed PLD recognition using HLTP comparatively achieves better accuracy of 99.34% and F1-score of 94.10%.

Effect of Histogram-based gradient boosting classifier A total of 1024 features are applied to the histogram-based gradient boosting classifier. Histogram-based gradient boosting classifier reduces computational complexity due to its

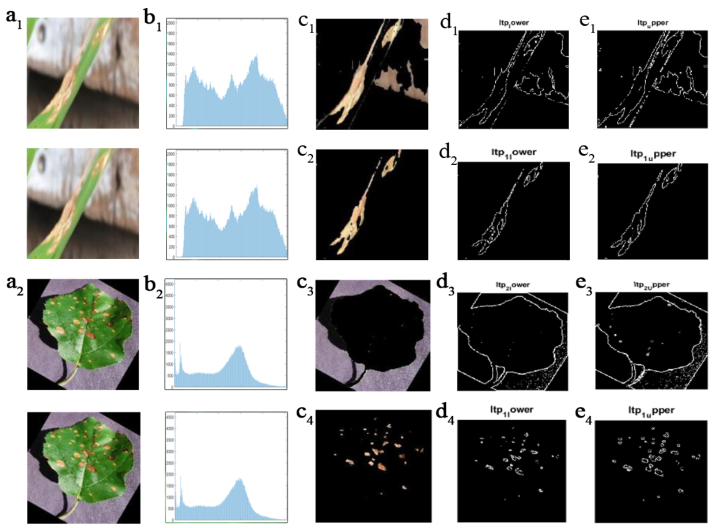

Fig. 3. Effect of image enhancement on recognizing plant leaf disease on critical situations: (a_1) rice blast disease image and (a_2) apple black rot disease image. (b_1), and (b_2) is the leaf image histogram of a_1 and a_2, respectively. (c_1), and (c_3) is the color segmentation results of a_1 and a_2 respectively in traditional K-means clustering having extra noise without image enhancement, and (c_2), and (c_4) is the segmentation results of a_1 and a_2 respectively in our modified color segmentation algorithm with image enhancement. (d_1), (d_3) and (e_1), (e_3) are the lower and upper features of traditional LTP respectively. (d_2), (d_4) and (e_2), (e_4) are the lower and upper features of modified HLTP respectively.

histogram data structure. It also reduces the biasness of similar features in various PLDs because of histogram classification phenomena over features. Variance in histogram comparatively differentiates well. It improves accuracy than the other machine learning algorithms and requires less memory space than CNN. So, it is useful and reliable for recognizing PLDs using the mobile application.

Performance Analysis Two hundred thirty-five(235) plant leaf disease images of twelve classes are used to evaluate our PLD recognition framework's performance. The recognition rate of each class is shown in a confusion matrix in Fig. 5. The summary of performance metrics, including accuracy, precision, recall, and F1-score, are shown in Table. 5.Our PLD recognition framework achieves accuracy, precision, recall, and F1 score of 99.34%, 94.66%, 93.54%, and 94.10%, respectively. For measuring the degree of separability among classes, the ROC curve has been shown in Fig. 6. AUC (The area under the ROC curve) for our proposed framework is 0.97. Minimum AUC is 0.85 for rice sheath-rot, and the maximum of AUC is 1 for five classes such as pepper bacterial-spot, grape black-measles, rice blast, cherry powdery-mildew, and rice tungro.

For further evaluation, we compare the performance of our PLD recognition with the benchmark method proposed by Pantazi et al. in [9] and Singh

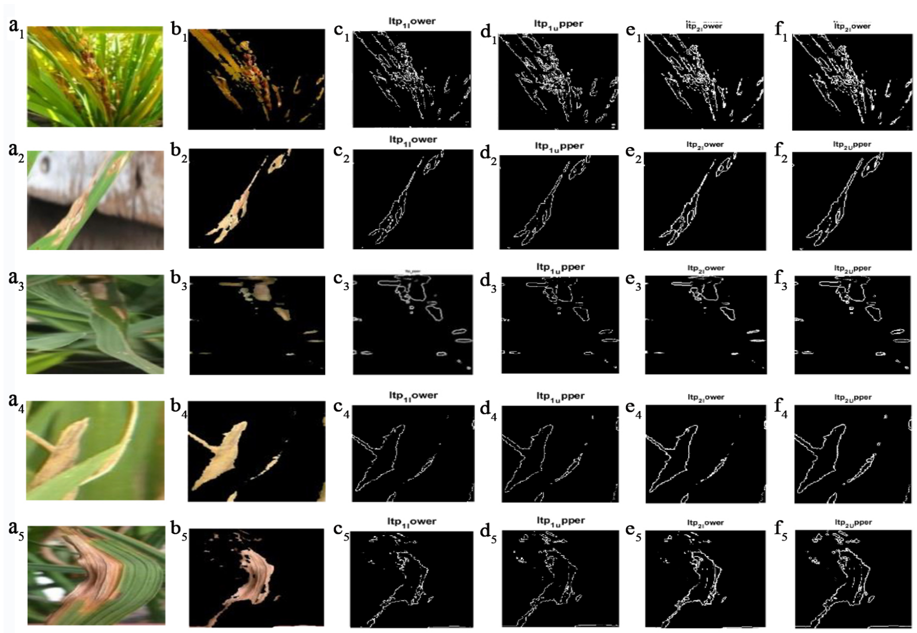

Fig. 4. The processing examples of rice images in our proposed PLD framework under different critical environments: $(a_1 - a_5)$ are the RGB PLD samples. (b) Segmented ROIs after implementation of adaptive centroid-based segmentation. (c) HLTP-1 lower features. (d) HLTP-1 upper features. (e) HLTP-2 lower features and (f) HLTP-2 upper features.

et al. in [11] on our dataset. The proposed method in [9] is significant for its high generalization using histogram features and the ability to overcome the intrinsic challenges(segmentation and different disorders with similar symptoms) under uncontrolled capture conditions. For comparing with the method in [9], the GrabCut algorithm for segmentation, Histogram-based Local Binary Pattern for feature extraction, and one class SVM for classification are executed on our dataset. The proposed method in [11] has significance in the auto initialization of clustering centers and generalization. For comparing with the method in [11], Genetic algorithm for segmentation, Color Co-occurrence method for feature extraction, and SVM for classification is executed on our dataset. From Table 6, it is visual that our proposed PLD recognition framework performs relatively better than other methods proposed in [9] and [11]. Our PLD recognition framework achieves accuracy and F1-score of 99.34% and 94.10%, respectively. These evaluations are superior to the accuracy achieved by the state-of-the-art method.

Moreover, we compare the PLD recognition framework results using histogram-based gradient boosting with the CNN-based PLD recognition model. As we have a small number of PLD images, we augment the PLD images using rotation, shifting, scaling, flipping, change in brightness, and contrast changes. Then, considering the number of network parameters, we execute the state-of-the-art

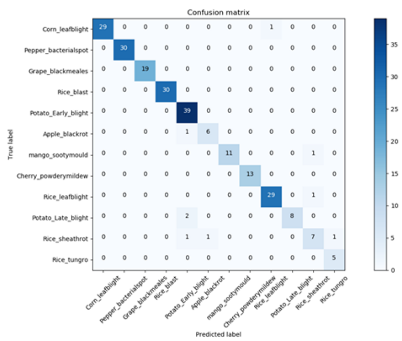

Fig. 5. Confusion matrix for recognizing plant leaf diseases.

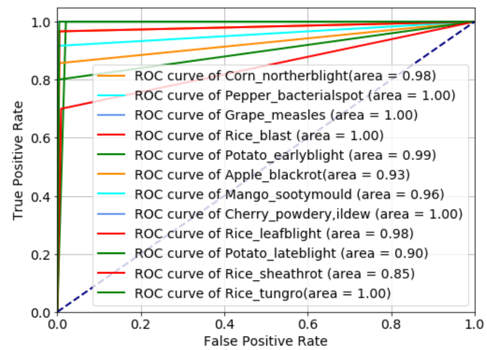

Fig. 6. ROC curve of each plant leaf diseases recognition of our framework.

Table 5. Performance evaluation of each classes using our proposed plant leaf disease recognition framework.

Class	TP	FP	FN	Accuracy	Precision	Recall	F1-score
Corn_northernblight	29	0	1	99.57%	100%	96.67%	98.30%
Pepper_bacterialspot	30	0	0	100%	100%	100%	100%
Grape_blackmeasles	19	0	0	100%	100%	100%	100%
Rice_blast	30	0	0	100%	100%	100%	100%
potato_earlyblight	39	4	0	98.29%	90.69%	100%	95.12%
Apple_blackrot	7	1	1	99.15%	87.5%	87.5%	87.5%
Mango_sootymould	11	0	1	99.57%	100%	91.67 %	95.65%
Cherry_powderymildew	13	0	0	100%	100%	100%	100%
Rice_bacterialleafblight	29	1	1	99.14%	96.67%	96.67%	96.67%
Potato_lateblight	8	0	2	99.15%	100%	80%	88.89%
Rice_sheathrot	7	2	3	97.87%	77.78%	70%	73.69%
Rice_Tugro	5	1	0	99.57%	83.33%	100%	90.91%
Average				**99.34%**	**94.66%**	**93.54%**	**94.10%**

convolutional layers based architecture, AlexNet(input image of 224 x 224) using ImageNet weights and achieves 99.25% accuracy, as shown in Table 7.

Critical Analysis Our framework recognizes well under different illumination, in the natural background, and complex background. However, there are still some misclassifications in detecting disease, as shown in Fig. 7(a-h). By analyzing these misclassifications, it is found that PLD images are misclassified due to multiple disease symptoms and changed symptom's features, such as shape. These challenges are located for future work. Not only information of colors or intensities of ROIs in spatial order, but also geometric features are considered as features.

5 Conclusion and Future Work

In our PLD recognition framework, ROIs are initially detected by modified ACS with automatic initialization of K. Then features have been extracted by HLTP. Finally, classification has been done by a histogram-based gradient boosting classifier. Our proposed PLD framework overcomes existing PLD recognition

Table 6. Comparison of performance evaluation with other state-of-the-art plant leaf disease recognition frameworks.

Class	Our framework		Method in [9]		Method in [11]	
	Accuracy	F1-score	Accuracy	F1-Score	Accuracy	F1-score
Corn_northernblight	**99.57**%	**98.30**%	95.74%	84.85%	97.02%	87.96%
Pepper_bacterialspot	100%	100%	100%	100%	99.15%	96.67%
Grape_blackmeasles	**100**%	**100**%	97.87%	85.71%	97.45%	85%
Rice_blast	**100**%	**100**%	99.15%	96.77%	99.58%	96.77%
potato_earlyblight	98.29%	95.12%	**98.72**%	**96.30**%	97.46%	92.86%
Apple_blackrot	99.15%	**87.5**%	99.15%	83.33%	97.02%	60%
Mango_sootymould	**99.57**%	**95.65**%	97.00%	55.55%	98.30%	76.13%
Cherry_powderymildew	**100**%	**100**%	99.15%	91.72%	98.30%	84.62%
Rice_bacterialleafblight	**99.14**%	**96.67**%	97%	89.66%	96.60%	85.19%
Potato_lateblight	99.15%	**88.89**%	99.15%	84.21%	97.02%	58.83%
Rice_sheathrot	**97.87**%	**73.69**%	96.59%	60%	94.46%	31.58%
Rice_Tugro	**99.57**%	**90.91**%	99.57%	83.33%	98.72%	63.31%
Average	**99.34%**	**94.10%**	**97.59%**	**76.57%**	**98.26%**	**85.02%**

Table 7. Comparison between PLD recognition using histogram-based gradient boosting classifier and state-of-the-art CNN model.

Method/Network	Accuracy	#Network/Learning Parameters	Storage required
Our proposed framework	99.34%	6	0.62MB
AlexNet	99.25%	6.4M	25.6MB

limitations, such as having image backgrounds, similar features in different disorders, and under uneven illumination and orientation for uncontrolled captured images. ACS eliminates the lack of sensitivity of k in K-means clustering [10] and performs effectively irrespective of the image backgrounds and similar features in different disorders. HLTP overcomes other challenges of PLD detection under uncontrolled capturing. Using linear interpolation and dynamic mean threshold, HLTP handles the orientation and variation of neighbors' grey level. In this work, some diseases having fungal and bacterial symptoms such as rice sheath-rot and apple black-rot are recognized in a better rate of, on average, 98.51%, as shown in Table 5. Our PLD recognition framework achieves an average of 99%

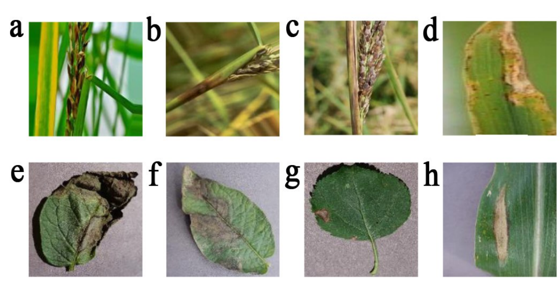

Fig. 7. Some misclassified images: (a), (b), (c) are some false positive rice sheath-rot images. (d) is rice bacterial leaf blight.(e) and (f) are some false positive potato late blight images. (g) is false positive apple black-rot image and (h) is false positive corn northern leaf-blight.

of accuracy for PLD with similar symptoms such as potato early-blight, potato late-blight, and corn northern-blight, as shown in Table 5. However, our proposed framework performs well and having high generalization ability but still has limitations of detecting multiple diseases. It can be solved by concatenating ROIs of multiple diseases.

References

1. A. A. Vasilyev, G.N.S., Vasilyev, A.N.: Processing plants for post-harvest disinfection of grain. Proceedings of the 2nd International Conference on Intelligent Computing and Optimization (ICO 2019) , Advances in Intelligent Systems and Computing **1072**, 501–505 (2019)
2. Borse, K., Agnihotri, P.G.: Prediction of crop yields based on fuzzy rule-based system (frbs) using the takagi sugeno-kang approach. Proceedings of the International Conference on Intelligent Computing and Optimization (ICO 2018), Advances in Intelligent Systems and Computing **866**, 438–447 (2018)
3. Boulent, J., Foucher, S., Théau, J., St-Charles, P.L.: Convolutional neural networks for the automatic identification of plant diseases. Frontiers in plant science **10** (2019)
4. Brahimi, M., Mahmoudi, S., Boukhalfa, K., Moussaoui, A.: Deep interpretable architecture for plant diseases classification. In: Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). pp. 111–116. IEEE (2019)
5. Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture **145**, 311–318 (2018)
6. Guolin Ke, Qi Meng, T.F.T.W.W.C.W.M.Q.Y., Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. pp. 1–3 (2017)
7. Liang, W.j., Zhang, H., Zhang, G.f., Cao, H.x.: Rice blast disease recognition using a deep convolutional neural network. Scientific Reports **9**(1), 1–10 (2019)
8. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Frontiers in plant science **7**, 1419 (2016)
9. Pantazi, X., Moshou, D., A.A. Tamouridou, W.: Automated leaf disease detection in different crop species through image feature analysis and one class classifiers. Computers and Electronics in Agriculture **156**, 96–104 (2019)
10. Sharma, P., Berwal, Y.P.S., Ghai, W.: Performance analysis of deep learning cnn models for disease detection in plants using image segmentation. Information Processing in Agriculture (2019)
11. Singh, V., Misra, A.: Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture **4**, 41–49 (2017)
12. Taha H. Rassem, B.E.K.: Completed local ternary pattern for rotation invariant texture classification. The Scientific World Journal p. 10 (2014)
13. Too, E.C., Yujian, L., Njuki, S., Yingchun, L.: A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture **161**, 272–279 (2019)