WILEY | Hindawi

Research Article

Blended Features Classification of Leaf-Based Cucumber Disease Using Image Processing Techniques

Jaweria Kainat,[1] Syed Sajid Ullah [iD],[2] Fahd S. Alharithi [iD],[3] Roobaea Alroobaea [iD],[3] Saddam Hussain [iD],[4] and Shah Nazir [iD][5]

[1]Department of Computer Science, COMSATS University Islamabad, Wah Cantt, Pakistan
[2]Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, USA
[3]Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[4]School of Digital Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Brunei Darussalam
[5]Department of Computer Science, University of Swabi, Swabi, Khyber Pakhtunkhwa, Pakistan

Correspondence should be addressed to Saddam Hussain; saddamicup1993@gmail.com and Shah Nazir; shahnazir@uoswabi.edu.pk

Existing plant leaf disease detection approaches are based on features of extracting algorithms. These algorithms have some limits in feature selection for the diseased portion, but they can be used in conjunction with other image processing methods. Diseases of a plant can be classified from their symptoms. We proposed a cucumber leaf recognition approach, consisting of five steps: preprocessing, normalization, features extraction, features fusion, and classification. Otsu's thresholding is implemented in preprocessing and Tan–Triggs normalization is applied for normalizing the dataset. During the features extraction step, texture and shape features are extracted. In addition, increasing the instances improves some characteristics. Through a principal component analysis approach, serial feature fusion is employed to provide a feature score. Fused features can be classified through a support vector machine. The accuracy of the Fine KNN is 94.30%, which is higher than the previous work in past papers.

1. Introduction

Plant leaf infections are a very inferior risk and require a robust economy for the nation. The development of farming products and accountability for a valuable part of these products are quite imperative.

In image processing, pipelined techniques have created incredible advancement results with high accuracy, yet there are a few concerns as well. Firstly, proficiency exceedingly relies upon features extraction and after that features selection for infected leaf image, in which the highlighted features are extracted, and secondly pipelined procedures are moderately unpredictable. The boisterous images, irregular lighting, and turmoil background in the dataset are unneglectable. This may damage the feature's eminence and reduce the detection rate. For this reason, a viable technique is utilized to dispense with noise and another hazard [1]. The plant leaf disease detection through the naked eye observes the manifestations, incorporating an extensive level of complications, because of this multifaceted nature and countless disease in crops and current pathology issues, even a ranch expert and vegetable pathologists are generally neglected to dissect specific disease and therefore conducted wrong results and deductions.

An automatically computerized strategy to perceive and classify the plant leaf disease would give full support to the plant pathologists for detection of disease through visual perceptions [2]. With the use of graphical processing units' numerous applications related to artificial intelligence (AI), machine learning has a higher rate of growth, which prompts curiosity in models and methodologies. An automated computerized identification is required for the

detection of classification of plant leaf disease [2]. Features extraction [3] and features selection are essential for image illustration. Many features extraction and classification approaches have been introduced for organic products, such as fruits. The hybrid method is used for the recognition and detection method for the citrus fruit disease. Lesion spot detection and geometric and texture feature fusion are used to select the best features. PCA was used for scoring the features. The proposed method utilizes a dataset of citrus diseased fruit named Anthracnose, canker, scab, greening, and melanoses. The proposed algorithm accomplished diverse accuracy which is 97% on citrus disease, 89% results on the consolidated, and 90.4% accuracy on the local dataset [4]. Different steps of image processing in which preprocessing, segmentation, features extraction, and classification are used are depicted in Figure 1.

Classical image processing contains propelled procedure of Computer Vision (CV) for detection of disease and expands the rate of accuracy in results. Approaches for the segmentation incorporate thresholding [5], adaptable thresholding approach [6] segmentation dependent on Neural Network (NN) [7,8], bend segmentation [9], and edge recognition-based segmentation [10]. These methods can be applied for plant leaf disease detection.

The image processing techniques are used to check different diseases in the crop and the diseases in insects of the crop. Deep learning techniques also help in identifying infections. The detection implementation by using the Convolutional Neural Network (CNN) is improved in terms of accuracy, well-defined results, and precision [11].

When it comes to know about the challenges, then different methods are used to compare the existing work with the previous one. The challenges include detection speed problem, occlusion problem, and lighting problem [5].

1.1. Problem Statement. Generally, Computer Vision- (CV-) based techniques for identification and classification consist of mainly five steps including preprocessing, features extraction, features selection, features fusion, and classification. For these types of structures, many challenges are encountered that are to be addressed to increase the efficiency and accuracy rate for classification.

The visually appealing quality of images of cucumber leaves for accurate feature extraction is the first challenge in the preprocessing stage. The maintenance of low- and high-quality contrast of spots on leaves, curved edges may affect the accuracy of classification. The changes required for the dataset include scale spacing for data augmentation. Feature fusion and feature selection use texture and shape features; therefore, an appropriate method for feature extraction and selection is needed for the improvement of accuracy in results and as well as for classification. Moreover, the domain of computer-based approaches indicates that visually interesting features in features selection are the main challenge for the improvement of accuracy performance and recognition rate.

1.2. Author's Contributions. Significant contribution in this approach is the preprocessing, feature extraction, and classification steps. In preprocessing resized, the greyscale transformed image uses the Otsu threshold, which converts the intensity image into a binary image. Tan–Triggs is used for normalization, which is never utilized in existing work. The second influential contribution in this algorithm is an amalgamation of three features. These features are extracted and further fused by using serial-based feature fusion. The dataset is usually too small for cucumber leaf disease, and preprocessing is the primary issue. The preprocessing step is necessary for this cause. The contrast enhancement problem, as well as the light variation problem, can render detecting cucumber leaf disease difficult.

1.3. Paper Organization. The remaining paper is arranged as follows: Section 2 describes literature work. Section 3 presents a proposed step along with diagram and all proceeding steps. Section 4 presents feature extraction and selection; Section 5 presents features fusion; Section 6 provides classification; in Section 7, results and experiments are presented; and the conclusion is described in Section 8.

2. Related Work

In related work, a number of techniques are presented for recognition and classification of plant leave disease. For the detection of crop disease, the CNN model was used for training and testing purposes, based on a plant leaf dataset of different classes. Existing deep learning models are applied, which are specific and are very easily applicable to crops attack. These models are determined with a small amount of data but give the best results for specific objects [6]. These modeling's contain the very high-performing technique for the development of disease recognition. For the sake of better results, the classification method using many classifiers gives satisfactory results [7].

Zhang et al. [12] described an IoT-based approach that handles the problem of discriminant features selection of disease parts. They explained that equal importance of features fusion, clustering, and PHOG features are extracted. A complete process of segmentation is done, and in this way, plant disease can be recognized.

By comparing different conventional techniques with CNN models using the images taken from the dataset, their results for the conventional perform better than a deep model. The losses may occur due to the mismanagement of CLR infection present on the coffee leaves, resulting in the drop of immature leaves. The severity of plant disease can be described with less amount of work with remote sensing bands and different vegetation catalogs. This is because of the multifaceted relationship between remote sensing catalog and leaf disease complexity [9]. Multiscreening with multifeatures is used for disease detection [10].

The diagnosis of leaf diseases is important in the cultivation of crops. The production can be achieved by observation which requires a high degree of learning and
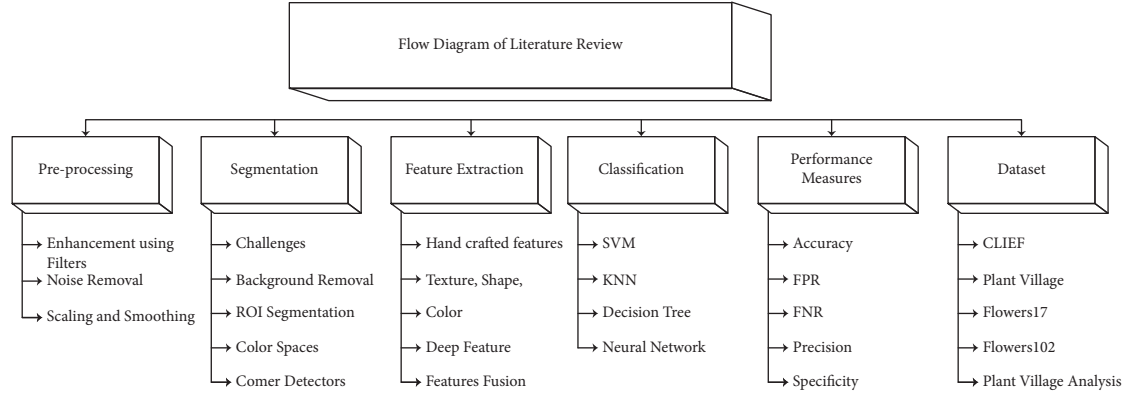
Figure 1: The literature review of different steps in image processing.

practice. The novelty in the domain of Artificial intelligence is the observation of retaining data, which is further enhanced by using different tools merging with the AI techniques helpful in diagnosing the crop diseases. The stimulating task is a visualization of features or interesting parts of a leaf image. This task is difficult in the case of automated detection [13].

The proposed approach for segmentation and detection method for plant leaf disease include the fusion of superpixel and k-mean clustering. Feature extraction method PHOG uses the color components and greyscale image. The accuracy achieved by the proposed method is 92.15% with fivefold cross-validation [12].

3. Proposed Model and Benefit

This section of the proposed model elaborates three major steps like preprocessing, feature extraction, and leave disease classification. Each step is further divided into multiple subparts in a sequence. The first step consists of preprocessing and normalization. Whereas Figure 1 shows the second step of feature extraction based on the texture and shape features of cucumber leaf images, Figure 2 shows the proposed model for cucumber leave disease identification and classification.

3.1. Dataset Expansion. The dataset for cucumber leaf disease is increased by increasing the instances of the dataset. The main reason for enlarging the number of images is to get the improving accuracy and decrease the error rate. A number of instances can be increased by using the MATLAB command and we also increase the dataset by using the flipping function, which also increases the dataset. The main reason for increasing the number of instances is to manage the dataset to get the optimal solution.

3.2. Preprocessing. Preprocessing plays an essential role in image enhancement in order to achieve the milestone to refine edges, eliminate noise, and remove blurriness, and challenge is required for image augmentation. The preprocessing step involves the conversion of a color image into a greyscale image. The intensity of the color image is converted into another channel with the image size of 640×480. Figure 3 shows the cucumber dataset. The purpose of image preprocessing is to encounter the interesting part (i.e., diseased part).

Normalization is a process in which data is rearranged to meet the needs of removing the redundancy from the data and adding all related data (logical data) or interesting parts. All normalization techniques perform in the preprocessing stage. First of all, RGB images from the dataset are converted into greyscale, resized images, and normalization techniques are performed.

Tan–Triggs is a function to normalize the data in a vector and matrix which can be computed through scoring. Our work includes the Tan–Triggs normalization and Otsu threshold to binarize the image into greyscale. The dataset consists of six diseased classes of cucumber containing Downy mildew, Powdery mildew, Anthracnose, Blight, *Corynespora*, and Angular leaf spot. Dataset is collected on the basis of private reference [14] excluding healthy images.

3.3. Tan–Triggs Normalization. Tan–Triggs is used for the enhancement of the local texture features of the image for disease recognition under lighting conditions. The Illumination Normalisation (IN) is used in the prepossessing stage to describe the difference between gamma correction, nonlinear filtering, and Gaussian filtering [15].

Tan–Triggs contains different steps: Gamma correction, the difference of Gaussian, masking, and contrast equalization. Details are as follows [16].

3.4. Illumination Normalization and Reflectance. The amount of illumination is reflected from the image, and components contain the reflectance; these components are

$$m(x, y) = J(x, y) * K(x, y). \tag{1}$$

In the above equation, J is light incident and depends on circumstances, and it varies when compared with reflectance.

We also take the log of J term when the objects are compressed by using the sum.
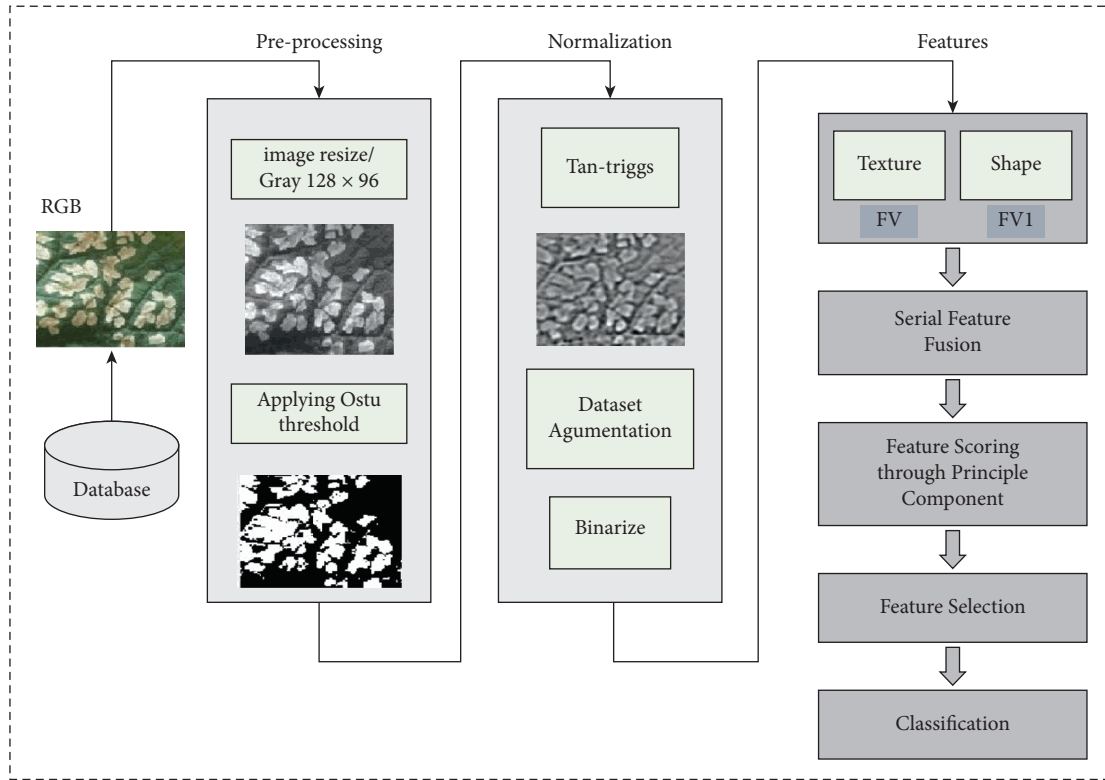
FIGURE 2: A proposed model for cucumber leaves disease identification and classification.
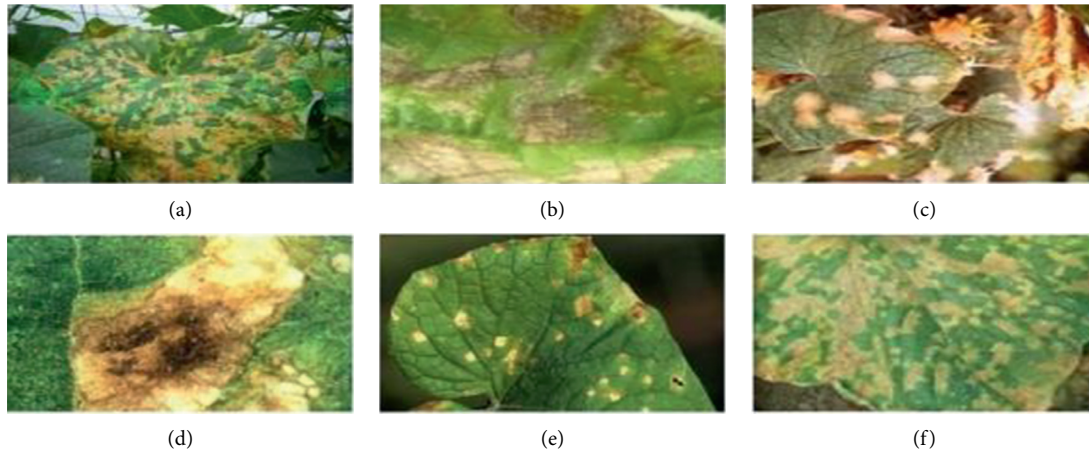

FIGURE 3: Cucumber diseased leaf dataset. (a) Angular leaf spot. (b) Anthracnose. (c) Blight. (d) *Corynespora*. (e) Downy mildew. (f) Powdery mildew.

$$log(J(x, y)) + K(x, y). \tag{2}$$

3.4.1. Gamma Correction.
Gamma correction is a transformation of a grey-level image. Its purpose is to convert every pixel into its intensity as follows:

$$0 \le \gamma \le 1, \tag{3}$$

where $g(J)$ increases the dynamic series of pixels. The dynamic range enhances the brightness while doing compression and enhances the dynamic range in dark areas.

3.4.2. Difference of Gaussian Filtering.
Normalization for enhancing the contrast does not eliminate shading/intensity effects. Shading effects have a low frequency. Shading effects can be removed by using the high range filters known as Dog filters.

Dog filter used for edge detection and Gaussian can be described by standard deviation sigma. Sigma σ) is used to remove noise only. Another Gaussian removes high-frequency details from the pixels. So, we can obtain a high-frequency edge by subtracting low-frequency pixels.

3.4.3. Contrast Equalization. Pixel intensifies and maintains the contrast. We can get pixel intensity by using the following equations:

$$j = \frac{j}{(\text{mean}(j^\alpha))^{(1/\alpha)}},$$

$$j = \frac{j}{(\text{mean}(\min(\tau, j^{\text{alpha}})))^{(1/\alpha)}}, \quad (4)$$

$$j = \tau * \tan h\left(\frac{1}{\tau}\right).$$

Figure 4 shows the grey scale and Tan–Triggs normalized image of a cucumber leaf.

A complete proposed diagram of the proposed model for the cucumber leaf disease recognition is presented in Figure 2.

4. Features Extraction and Selection

Feature extraction in image processing extracts the interesting part of an image (i.e., the diseased part of an image). The selected features reduce the dimensions. In the proposed system, texture and shape features including HOG, LBP, and COLOR features are extracted. As HOG stands for Histogram of Gradient, LBP stands for Local Binary Pattern, and SVM stands for Support Vector Machine. By using many parameters, every feature returns a different feature vector. The detailed diagram is shown in Figure 5. The details of every feature are described as follows.

HOG features are widely used for object detection, which is represented as a single feature vector where each represents a segment of an object. Mostly computed by using the sliding window for each position in a segment of an image with an SVM classifier. Therefore, we use these features for the detection of disease in the cucumber leaves. Preprocessing phases resize the original image into any type of size through which the position of an image can get through several scales. The visualization of the HOG feature for cucumber anthracnose is shown in Figure 6.

The local binary pattern that is used for the texture feature was first to discover by [17] in 2002. The mathematical modeling of LBP can be described by using blocks. The block is overlapped and divided into the same size. The center and neighboring pixels are matched with the greyscale values. The threshold is fixed to compare the block size value with the grey pixel value. If the equivalent grey-level value is greater than the center pixel value, the position is marked as "1." Otherwise, the position is marked as "0" [18].

The 7 layers of CNN are used with the LBP. The LBP stands for Local Binary Pattern that is used for the features extraction of disease [14].

Color features are used to find out the diseases in the plants or crops. Different colors spaces are explored in [11], and then features are extracted from different color channels.

Binary code: 00100011 and LBP.

In the above figure, the first two tables show the difference of pixels, and the third table shows the threshold value and binary code accordingly.

The formal form of the LBP feature can be explained by using the following equation [18]:

$$lbp(L_a, M_a) = \sum_{q=0}^{q-1} 2^q f(i_q - i_a). \quad (5)$$

In the above equation, the terms $\mathbf{L_a}$, $\mathbf{M_a}$ show the location of the pixel that is located at the center; i_q and i_a show the brightness of adjacent pixels where the term $\mathbf{f}$ shows symbolic function:

$$f(a) = \begin{cases} 1, & \text{if } x \ge 0, \\ 0, & \text{else.} \end{cases} \quad (6)$$

The local binary pattern has the ability to adapt to the circumstances of different texture features in detail and is used in many fields, including texture recognition [19] and hyperspectral image classification [20]. The diseased anthracnose leaf has LBP feature extraction with pixel-wise LBP image shown in Figures 5 and 7, showing the calculation of features extraction and features scoring of HOG, LBP, and color, respectively.

5. Features Fusion

Feature fusion plays an important role in the field of machine learning and computer vision. Many features are joined together to make a new feature vector by applying the serial-based fusion. To get the optimal solution, the HOG, LBP, and color feature vectors are fused.

Features reduction is substantial in most disease recognition processes as it eliminates the unwanted features and removes the redundant features from the images [14]. As a result, we will get an accurate classification. The fusion method is very helpful for getting better results. This paper implements a probability process for the removal of inappropriate features [21]. Features are named as feature factors F_1, F_2, and F_3; these factors belong to R, where R represents the real values. Hog, LBP, and color features are to be considered, which represent the positive value features.

The feature vector can be described by using the following equation:

$$Fv_{i \times H} = \{v_{i \times 1}, v_{i \times 2}, v_{i \times 3}, \ldots, v_{i \times H}\}. \quad (7)$$

Equation (7) shows the fused feature vector of HOG, where $i = 1$, F is the feature vector, and H shows the HOG feature.

$$Fv_{j \times L} = \{v_{j \times 1}, v_{j \times 2}, v_{j \times 3}, \ldots, v_{j \times L}\}. \quad (8)$$
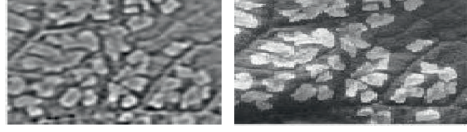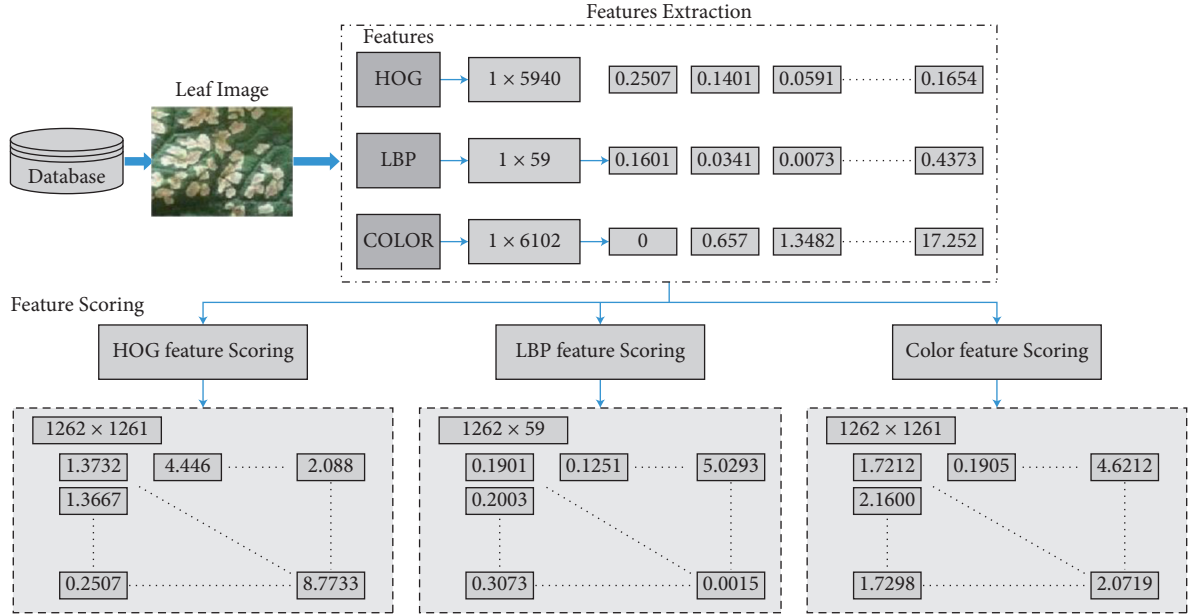
FIGURE 4: Tan–Triggs normalization.

FIGURE 5: Features extraction and features scoring of HOG, LBP, and color, respectively.
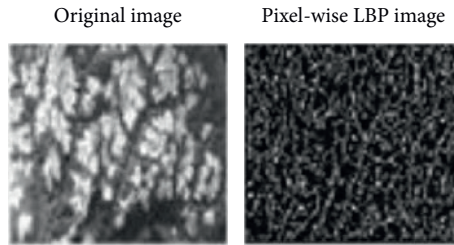

FIGURE 6: HOG visualization.

FIGURE 7: Grey image with LBP pixel-wise extraction.

Equation (8) shows the fused feature vector of LBP, where $j = 1$, F is the feature vector, and L shows the LBP feature.

$$Fv_{k \times C} = \{v_{k \times 1}, v_{k \times 2}, v_{k \times 3}, \ldots, v_{k \times C}\}. \tag{9}$$

Equation (9) shows the fused feature vector of color, where $k = 1$, F is the feature vector, and C shows the color feature.

Features are fused by using the following equation, where D shows the fused vector:

$$\text{FUSED } (FV)_{1 \times D} = \sum_{i=1}^{3} \{Fv_{i \times H}, Fv_{j \times L}, Fv_{k \times C}\}. \tag{10}$$

HOG, LBP, and color features are extracted and fused together. Figure 8 shows the complete feature extraction of texture, color, and shape features.
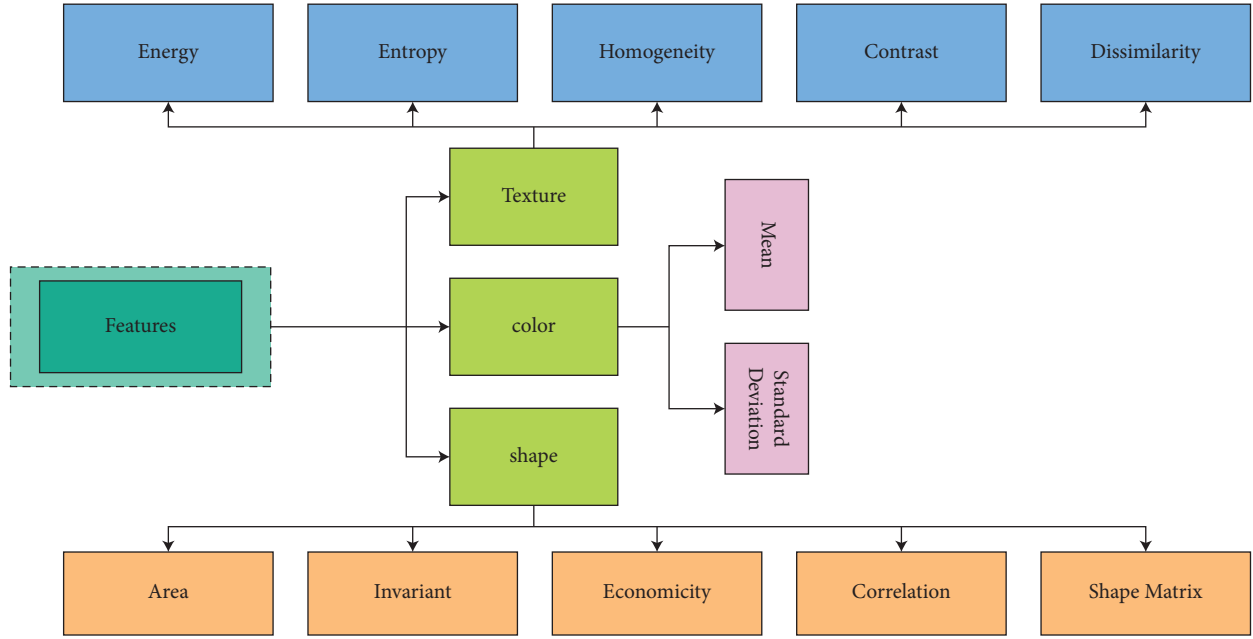
FIGURE 8: Features extraction of texture, color, and shape features.

6. Classification

Classification plays an essential role in the field of image processing. To attain satisfactory results from the multi-SVM, many experiments have been performed by using the selected feature. In the category of classification, many classifiers are used to classify the disease. KNN classifiers include medium, weighted, and fine KNN.

The geometric family of classifiers includes the diversity of SVM. These can be substantiated by using the functions quadratic, Gaussian kernel, linear, and cubic. The tree category contains the ensemble boosted and bagged trees. Probability classifications include naïve Bayes, multikind of naïve Bayes, Bayesian logistic regression, and Bayesian net. These classifiers are used to evaluate the performance of the cucumber leaf dataset, which is based on different performance measures later used in the result section.

7. Experimental Results and Analysis

A set of models are presented for the final analysis of the suggested framework. In the first experiment, HOG and BRISK features are fused and later fed to the classifiers. In the second test, all types of features (shape, texture, and colors) are fused and given to the classifiers, and in the final test, reduced features using the proposed method are supplied to the classifiers. The key reason behind these experiments is to analyze the performance of each step, involved in this proposed methodology.

7.1. Dataset and Performance Measures. For the evaluation of experimental results, cucumber leaf dataset is used, which is collected through private reference. The six classes for cucumber leaf dataset are angular leaf spot, anthracnose, blight, coryneform, downy mildew, and powdery

mildew. The total number of images in the dataset is 1262. Three types of experiments are performed with the help of features extraction. All experiments contain different feature subsets with the variation of HOG, LBP, and COLOR features. Each experimental result varies from the other based on selecting features. Classification results can be measured by using classification methods including Fine KNN, subspace KNN, Bagged trees, Weighted KNN, SVM cubic, Quadratic SVM, boosted trees, Fine Tree, Cosine KNN, and Medium tree. The performance is calculated by using many classification methods. Performance measures include False Positive Rate (FPR), Sensitivity, Classification Rate (CR), Specificity, False Negative Rate (FNR), Precision, Accuracy, and Time.

In true positive rate means diseased samples, and true negative means both actual and predicted values are negative. False-negative means that actual values are positive and predicted values are negative. The mathematical results can be taken by using the formulas for CR, FPR, Sensitivity, Specificity, Precision, and FNR as follows [22]:

$$CR = \frac{(CC\ samples)}{(all\ samples)} \times 100\%,$$

$$FNR = \frac{(false\ negative)}{(all\ negatives\ output)},$$

$$FPR = \frac{(false\ positive)}{(all\ positives\ output)}, \qquad (11)$$

$$senstivity = \frac{true\ positive}{true\ positive + false\ negative},$$

$$presicion = \frac{true\ positive}{TP + FP}.$$

There are many cross-validation (CV) methods available in MATLAB, including *k*-fold, HoldOut, LeaveMout. The method performed by machine learning mainly involves tuning, assembly of the trained model, and performance evaluation of the proposed model. CV methods besides the performance measure also perform estimation and configuration. This is enthusiastically inclined and effective. An effective bootstrap method for cross-validation (BBC-CV) is used for selecting the best results and performance [23]. Features were combined with Principal Component Analysis (PCA) with key point's I-e principal component and index. We perform 10-fold cross-validation on each trial to get better results and generate classifying data for the cucumber dataset. All tests were performed on MATLAB 2018a using the personnel laptop with Windows 8.1, 64-bit operating system, and 6 GB RAM.

7.1.1. Angular Leaf Spot.

The spots may consist of different colors of light yellow, and the condition of leaves becomes very severe in case of less nitrogen. Angular leaf spot means that spots appear on veins of leaf caused by the virus pathogens, *P. syringae* PV. Lachrymans. This disease is mostly caused by infection and the pseudomonas virus. For this purpose, medicated protection element Plant Growth-Promoting Rhizobacteria (PGPR) mediated ISR is recycled [24].

7.1.2. Powdery Mildew.

Powdery mildew is a severe disease that causes mildew toxicities in leaves of cucumber. Infected fluid from leaves causes reductions in growth, premature vegetation, as a result of loss in economic. A fungal virus that is commonly known as phytopathogenic is present in powdery mildew. There is fast and divergent increase of the fungus-virus in infected leaves preserved with Milana. Milana may be suitable in the case of plant disease defense in the incorporated organization of powdery mildew [25].

7.1.3. Downy Mildew.

Fungus bacterial diseased spreads from the leaves, caused by the infection known as *Pseudoperonospora*. The growing reason for downy mildew is the increased rate of leaf temperature. Unhealthy regions with complex backgrounds show infected parts [26].

7.1.4. Anthracnose.

Anthracnose is caused by yeast effects and the fungal virus is known as *Colletotrichum*. These viruses cause resistance possessions in the growth rate of cucumber leaves [3].

7.1.5. Corynespora.

The infection in leaves starts from slight spots of brown color with increasing the yellow glory. Leaves become uneven in shape and result in leafless plants. Symptoms rapidly increase in the fungal virus known as *Corynespora* [8].

7.2. Results.

For measurable results, three separate tests are performed using a distinct number of features in each experiment. The description for each experiment which contains the number of features and number of diseased classes is shown in Table 1.

The description analysis for each experiment is performed on 1262 images with multifeatures selection.

7.2.1. Experiment 1.

In experiment 1, the total number of images are 1262 with six diseased classes of a cucumber leaf. For these first results, in the first experiment, HOG, LBP, and color features are extracted. 100 features for HOG and LBP are taken at 10-fold cross-validation. Fine KNN and subspace KNN give the highest accuracy of 94.30% than other classification methods. The accuracy gained for both fine KNN and subspace KNN is similar. The instances for the feature vector (FV) are double. Performance measures include specificity, sensitivity, precision, FNR, FPR, and time. The graphical illustration of certain classification methods is shown in Figure 9. The other classifiers including bagged KNN, weighted KNN, and SVM cubic are also described as below. The first experiment gives the best results as compared with other experiments from the above discussion and analysis our proposed method gives the performance for the feature vector size for HOG is 1×5940, LBP is 1×59, and color is 1×6120. In the confusion matrix ensemble, fine KNN confirms the classification results shown in Table 2. Table 3 shows the experimental results of experiment 1. The confusion matrix of cucumber leaf dataset of Test 1 on ensemble fine KNN is shown in Table 2.

7.2.2. Experiment 2.

In the second experiment, HOG, LBP, and color features are extracted. 300 features for HOG and LBP are taken at 10 cross-validations. In the second experiment, fine KNN and subspace KNN give 94.60% and 94.50% accuracy with 84.0% specificity and sensitivity, 84.6% precision, 5.4% FNR rate, and 0.011 FPR rate, respectively. These classifiers give the highest accuracy as compared to others. The instances for the feature vector (FV2) are double. Performance measures include specificity, sensitivity, precision, FNR, FPR, and time. The confusion matrix confirms the classification results in Table 4. Table 5 shows the performance evaluation of experiment 2. Graphical representation of classification methods in terms of accuracy and time for experiment 2 is shown in Figure 10.

7.2.3. Experiment 3.

In the third experiment, HOG and LBP features are extracted. 500 features for HOG and LBP are taken at 10 cross-validations. In the first experiment, fine KNN and subspace KNN give 94.2% and 82.5% specificity and 83.66% precision, 5.8% FNR, and 0.011 FPR rate, respectively. These classifiers give the highest accuracy as compared to others. The instances for the feature vector (FV2) are double for the reason of improving accuracy. Performance measures include specificity, sensitivity, precision, FNR, FPR, and time. The confusion matrix approves the result of the classification method in Table 6. The remaining performance of classifiers includes Fine SVM, cubic SVM, and fine Gaussian SVM, fine tree, weighted

TABLE 1: Selected features in experiments.

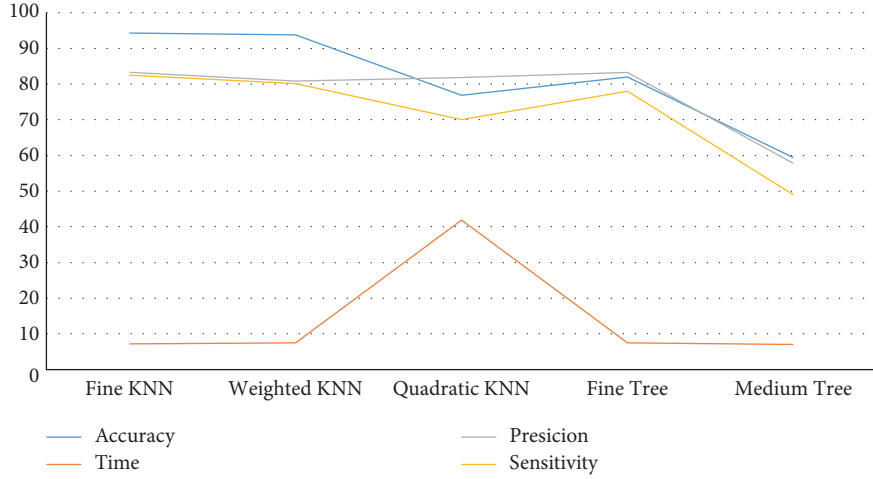|		Selected features		
Experiment	Classes	HOG	LBP	Color
Experiment 1	6	100	100	50
Experiment 2	6	150	100	0
Experiment 3	6	100	150	60

FIGURE 9: Graphical representation of classification methods in terms of accuracy and time for experiment 1.

TABLE 2: Confusion matrix of cucumber leaf dataset of test 1 on ensemble fine KNN.

		Classification classes					
Classification classes	Total images	Angular leafspot	Anthracnose	Blight	*Corynespora*	Downy mildew	Powdery mildew
Angular leafspot	204	**96%**	2%	1%	1%		
Anthracnose	54		**99%**	1%			
Blight	130			**98%**			5%
Corynespora	22		25%	56%	**16%**	2%	2
Downy mildew	257	2%	2%	3%	5%	**87%**	1%
Powdery mildew	348			1%			**99%**

Bold shows highest values in terms of accuracy.

TABLE 3: Result for experiment 1.

Classifier	Performance evaluation					
	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	FNR (%)	FPR
Fine KNN	**94.30**	**82.5**	**82.5**	**83.30**	**5.7**	**0.011**
Subspace KNN	**94.30**	**82.5**	**82.5**	**82.50**	7.2	0.011
Bagged trees	94.00	96.0	96.0	79.80	6.0	0.013
Weighted KNN	93.80	80.1	80.1	80.83	6.2	0.013
SVM cubic	91.90	65.5	91.6	94.40	8.1	0.031
Quadratic SVM	76.70	68.9	72.0	81.80	23.3	0.055
Fine tree	82.00	78.0	78.0	83.20	18.0	0.040
Boosted trees	66.20	56.0	56.0	72.40	33.8	0.071
Cosine KNN	68.70	60.0	60.0	64.80	31.3	0.071
Medium tree	59.40	49.4	49.4	57.80	59.4	0.093

Bold shows the highest values in terms of accuracy.

KNN and their specificity, sensitivity, precision, FNR, and FPR, respectively, are also shown in Table 7. The accuracy improves by the subspace classifier by using 10-fold cross-validation. Graphical representation of classification methods in terms of accuracy, time, precision, and specificity for experiment 2 is shown in Figure 11.

TABLE 4: Confusion matrix of cucumber leaf dataset of test 2 on ensemble fine KNN.

Classification classes	Total images	Classification classes					
		Angular leafspot	Anthracnose	Blight	*Corynespora*	Downy mildew	Powdery mildew
Angular leafspot	204	**97%**		1%	1%		
Anthracnose	54		**99%**	1%			
Blight	130			**98%**	1%		1%
Corynespora	22		25%	56%	**23%**	2%	2
Downy mildew	257	1%	2%	3%	5%	**87%**	1%
Powdery mildew	348						**100%**

Bold shows the highest values in terms of accuracy. Fine KNN and subspace KNN have highest accuracy.

TABLE 5: Performance evaluation of experiment 2.

Classifier	Performance evaluation					
	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	FNR (%)	FPR
Fine KNN	**94.6**	**84.0**	**84.0**	**84.6**	**5.4**	**0.011**
Subspace KNN	94.5	83.8	83.8	84.3	5.6	0.011
Bagged trees	94.2	80.5	80.5	81.0	5.8	0.013
Weighted KNN	93.6	81.0	81.0	82.5	6.4	0.013
SVM cubic	91.6	90.8	90.8	93.6	8.4	0.018
Fine tree	81.8	76.6	76.6	81.6	18.2	0.045

Bold shows the highest values in terms of accuracy. The Fine KNN classifier has the highest accuracy of 94.6%, specificity of 84.0%, sensitivity of 84.0%, precision of 84.6%, FNR rate of 5.4, and FPR is 0.011.
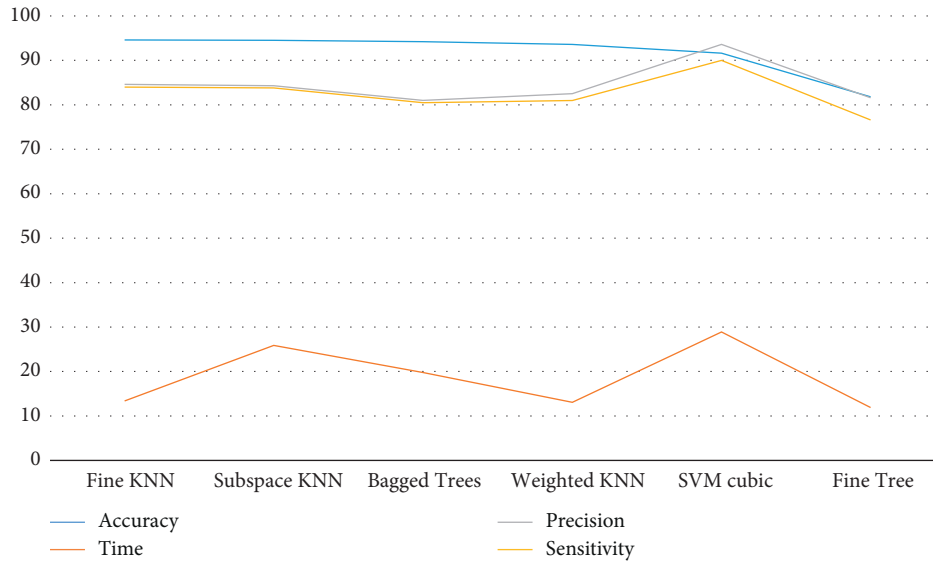

FIGURE 10: Graphical representation of classification methods in terms of accuracy and time for experiment 2.

TABLE 6: Confusion matrix of cucumber leaf dataset of test 1 on subspace-KNN.

Classification classes	Total images	Classification classes					
		Angular leafspot	Anthracnose	Blight	*Corynespora*	Downy mildew	Powdery mildew
Angular leafspot	204	**96%**	2%	1%	1%		
Anthracnose	54		**99%**	1%			
Blight	130			**96%**	1%	2%	1%
Corynespora	22		12%	25%	**18%**	45%	22
Downy mildew	257	1%	2%	3%	5%	**87%**	1%
Powdery mildew	348						**99%**

The highest values are shown in bold.

TABLE 7: Performance evaluation of experiment 3.

| Classifier | Performance evaluation | | | | | |
	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	FNR (%)	FPR
Subspace KNN	**94.2**	**82.5**	**82.5**	**83.66**	**5.8**	**0.011**
Bagged trees	93.0	95.2	94.8	79.16	7.0	0.015
Fine KNN	93.9	83.1	82.1	83.66	6.1	0.013
Cubic SVM	91.8	93.6	91.0	94.00	8.2	0.018
Fine Gaussian SVM	71.2	62.2	62.2	81.8	28.2	0.068
Fine tree	82.1	77.6	77.6	83.2	17.9	0.08
Weighted KNN	93.4	95.6	90.7	96.9	6.6	0.016

Bold shows the highest values in terms of accuracy. The subspace classifier has the highest accuracy of 94.2%, specificity of 82.5%, sensitivity of 82.5%, precision of 83.66%, FNR rate of 5.8, and FPR is 0.011.
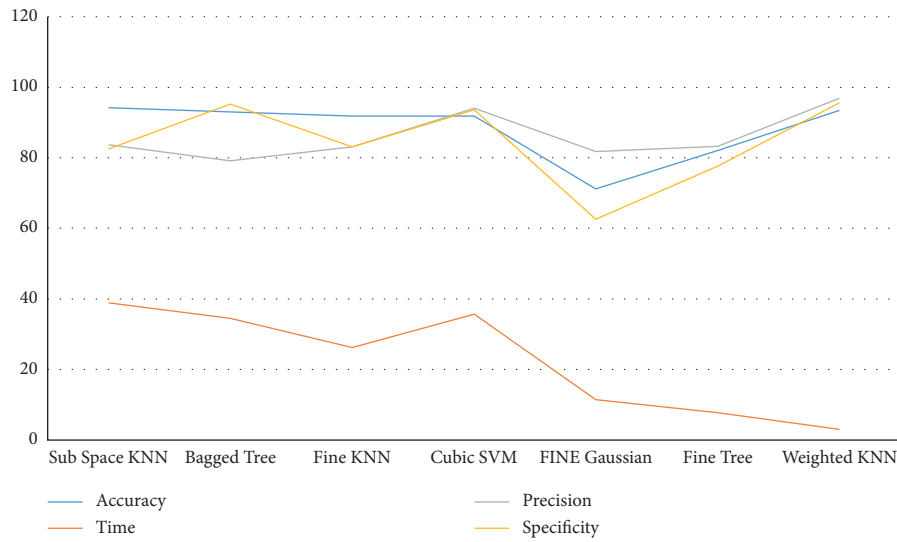

FIGURE 11: Graphical representation of classification methods in terms of accuracy, time, precision, and specificity for experiment 2.

8. Conclusion

The proposed approach is mainly used for cucumber leaf detection, based on preprocessing, normalization, feature extraction, feature selection, fusion, and classification. From above all discussion and influences, it is concluded that cucumber leaf disease is addressed by using the feature extraction of HOG, LBP, and color features. These texture and shape features help in the recognition of the disease in leaves and color features help in the recognition of diseased part of a leaf. Furthermore, feature selection and feature fusion are important to improve the accuracy of different performance measures, including accuracy, specificity, sensitivity, and precision. The proposed algorithm method shows 94.6% accuracy, which is better than that of existing work.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All the authors contributed equally to this work and were involved in its development at every phase. The submitted version of the work has been read and approved by all authors.

Acknowledgments

References

[1] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Computers and Electronics in Agriculture*, vol. 154, pp. 18–24, 2018.

[2] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 283–292, 2018.

[3] H. D. Gottstein and J. Kuc, "Induction of systemic resistance to anthracnose in cucumber by phosphates," *Phytopathology*, vol. 79, no. 2, pp. 176–179, 1989.

[4] M. Sharif, M. A. Khana, Z. Iqbal, M. F. Azam, M. I. UllahLali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Computers and Electronics in Agriculture*, vol. 150, pp. 220–234, 2018.

[5] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, pp. 1–18, 2021.

[6] X. E. Pantazi, D. Moshou, A. A. Tamouridou, and S. Kasderidis, "Leaf disease recognition in vine plants based on local binary patterns and one class support vector machines," in *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations,* Thessaloniki, Greece, Springer, September 2016.

[7] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deepplant: plant identification with convolutional neural networks," in *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP),* Quebec City, Canada, IEEE, September 2015.

[8] M. Kwon, B. R. Kang, B. H. Cho, and Y. C. Kim, "Occurrence of target leaf spot disease caused by Corynespora cassicola on cucumber in Korea," *Plant Pathology*, vol. 52, no. 3, p. 424, 2003.

[9] A. Chemura, O. Mutanga, M. Sibanda, and P. Chidoko, "Machine learning prediction of coffee rust severity on leaves using spectroradiometer data," *Tropical Plant Pathology*, vol. 43, no. 2, pp. 117–127, 2018.

[10] P. A. Marin Zapata, S. Roth, D. Schmutzler, T. Wolf, E. Manesso, and D. A. Clevert, "Self-supervised feature extraction from image time series in plant phenotyping using triplet networks," *Bioinformatics*, vol. 37, no. 6, pp. 861–867, 2021.

[11] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Recent advances in image processing techniques for automated leaf pest and disease recognition–A review," *Information processing in agriculture*, vol. 8, no. 1, pp. 27–51, 2021.

[12] S. Zhang, H. Wang, W. Huang, and Z. You, "Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG," *Optik-International Journal for Light and Electron Optics*, vol. 157, pp. 866–872, 2018.

[13] X. Pantazi, D. Moshou, and A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," *Computers and Electronics in Agriculture*, vol. 156, pp. 96–104, 2019.

[14] S. Zhang and Z. Wang, "Cucumber disease recognition based on Global-Local Singular value decomposition," *Neurocomputing*, vol. 205, pp. 341–348, 2016.

[15] X. Tan and W. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[16] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proceedings of the International Workshop on Analysis and Modeling of Faces and Gestures*, Springer, Rio de Janeiro, Brazil, October 2007.

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[18] Z. Xia, X. Ma, Z. Shen, X. Sun, N. N. Xiong, and B. Jeon, "Secure image lbp feature extraction in cloud-based smart campus," *IEEE Access*, vol. 6, 2018.

[19] R. Mehta and K. Egiazarian, "Dominant rotated local binary patterns (DRLBP) for texture classification," *Pattern Recognition Letters*, vol. 71, pp. 16–22, 2016.

[20] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3681–3693, 2015.

[21] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1391–1402, 2005.

[22] C. Xie, C. Yang, and Y. He, "Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities," *Computers and Electronics in Agriculture*, vol. 135, pp. 154–162, 2017.

[23] I. Tsamardinos, E. Greasidou, and G. Borboudakis, "Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation," *Machine Learning*, vol. 107, no. 12, pp. 1895–1922, 2018.

[24] L. Liu, J. Kloepper, and S. Tuzun, "Induction of systemic resistance in cucumber against bacterial angular leaf spot by plant growth-promoting rhizobacteria," *Phytopathology*, vol. 85, no. 8, pp. 843–847, 1995.

[25] F. Daayf, A. Schmitt, and R. Belanger, "The effects of plant extracts of Reynoutria sachalinensis on powdery mildew development and leaf physiology of long English cucumber," *Plant Disease*, vol. 79, no. 6, pp. 577–580, 1995.

[26] M. Lindenthal, U. Steiner, H. W. Dehne, and E. C. Oerke, "Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography," *Phytopathology*, vol. 95, no. 3, pp. 233–240, 2005.