PROGRESS IN COLOR, COLORANTS AND COATINGS

available online @ www.pccc.icrc.ac.ir

Progress in Color Colorants Coating 17 (2024), 39-52

Ministry of Science, Research and Technology Institute for Color Science & Technology

Automated Detection of Plant Disease Based on Color Histogram Feature Selection Using Hybrid Random Forest with Adaboost Algorithm

O. S. Greeshma *[1], P. Sasikala[1], S. G. Balakrishnan[2]

[1.] Department of Mathematics, VMKV Engineering College, Salem, P.O. Box: 636308, Tamil Nadu, India.

[2.] Department of CSE, Mahendra Engineering College, Namakkal, P.O. Box: 636308, Tamil Nadu, India.

ARTICLE INFO

Article history:
Received: 01 Feb 2023
Final Revised: 20 July 2023
Accepted: 26 July 2023
Available online: 16 Oct 2023

Keywords:
Random forest
Color histogram
Classification
Feature extraction

ABSTRACT

Multiple microbes can alter a plant's development and agricultural productivity, which has significant implications for the ecosystem and human life. As a result, timely identification, prevention, and prompt treatment are required. Fundamental methods have some drawbacks to plant disease identification like more time-consuming, accuracy, doesn't support multiple plant detection. This paper introduces a hybrid model that uses a random forest classifier combined with the AdaBoost Classifier to classify plant diseases to overcome the above-said drawbacks. So as to individualize normal and abnormal leaves from data sets, the suggested methodology employs the Random Forest with AdaBoost algorithm. The operational processes in our suggested study are preprocessing, segmentation, feature extraction, training the classifier, and classification. The produced datasets of infected and uninfected leaves are combined and processed using the Random Forest classifier to categorize the infected and uninfected photos. Color Histogram is used to gather features from imagery. KNN, Naive Bayes, and SVM are all used to evaluate our suggested technique. Prog. Color Colorants Coat. 17 (2024), 39-52© Institute for Color Science and Technology.

1. Introduction

Plant Disease Detection is essential to prevent crop loss due to pest attacks and various diseases [1]. Timely identification and treatment of diseases can reduce crop losses. Manually inspecting the crop fields for disease identification is time-consuming and requires domain knowledge. Due to these challenges, early-stage leaf disease detection systems are in high demand in agriculture. Automatically detecting crop disease in its earlystage can help control infections and the spread to other plants and reduce plant losses. Agriculture has a significant role in India's economy compared to other countries.

India is the second-largest agricultural producer worldwide [2]. Approximately 75 % of India's inhabitants depend on agriculture, either explicitly or implicitly [3]. In India, due to the rising population, there is a need for advances in agriculture to meet food needs. The agriculture industry requires a massive overhaul to withstand the transforming environment of the Indian economy [4].

Hence, the impact of plant disease and infections from pests on agriculture may affect India's economy by reducing the production quality of food. Early monitoring and proper diagnosis of crop disease using appropriate protection of the crop system may prevent

*Corresponding author: * osgreeshma@gmail.com
Doi: 10.30509/pccc.2023.167095.1202

losses in production quality. Plant diseases can affect several plant components, including the leaf, stem, and seed.

Detecting plant diseases early is a must since it increases yields by more than 60 per cent [5]. Infectious and non-infectious plant diseases are the two types of plant diseases. Infectious plant diseases are mainly provoked by fungi, viruses, and bacteria [6]. Diagnosing diseases swiftly and precisely is vital to deliver effective treatments for diseases. It is primarily hinged on the symptoms of a sick plant. The disease can harm any leaf component, including the blossom, stem, and root.

On the other hand, leaf examination is regarded as the optimal tool for plant diagnosis [7]. The diseased leaf is typically deformed in shape, colour, size, etc.Using a specialist to identify plant leaf disease is antiquated, time-consuming, and inappropriate [8]. As a result, an automated, efficient, and less economical system for detecting diseases from images and suggesting appropriate solutions is required [9]. Innovative agriculture systems rely on automatic leaf disease detection and monitoring to enhance agricultural produce's productivity and maintenance [10]. Researchers have suggested various solutions using image processing and machine Learning (ML) techniques. ML mainly influences recent approaches, and they have shown impressive performance due to the availability of the data, computing resources, and sophistication in the learnability of ML algorithms. Among many approaches, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), multilayer perceptron, Fisher Linear Discriminant (FLD), and random forest classifiers have been prominent in plant leaf disease identification [11-16].

Researchers have suggested various solutions using image processing and machine Learning (ML) techniques. ML mainly influences recent approaches, and they have shown impressive performance due to the availability of the data, computing resources, and sophistication in the learnability of ML algorithms. Machine learning has developed as a robust computing methodology for resolving complicated computer vision challenges. Leaf disease detection in plants at the early stages is complex and challenging. Also, if the principal item in the image merges with other entities termed noises, the execution gets more complex [17, 18]. The development of leaf disease detection techniques has received much attention in earlier studies. It was considered insufficient and challengingeven though the illumination frequently affected the outcomes of feature extraction on the leaf picture, and the features employed may not have been discriminatory.Machine learning enables computers to operate without any need for human involvement. Machine learning allows a system to operate independently and make predictions [19]. There are three machine learning methodologies: supervised, reinforcement, and unsupervised. Pre-processing, segmentation, feature extraction, and classification are the four phases in machine learning applications for leaf disease diagnosis [20].

Ismail et al. [21] presented an approach for detecting and classifying potato plant disease using computer vision techniques. The proposed method is based on colour and texture features. The implemented method is processed in four steps- In the pre-processing and segmentation, LAB colour space and Delta E colour difference method are applied. Later features are extracted based on RGB, HSV and Local Binary Patterns (LBP). Multi Support Vector Machine (SVM) finally classifies the extracted patterns.Moreover, the results of feature subsets of RGB and HSV colour features with the addition of LBP texture features and a classification difference of 3.6% between RGB and HSV colour feature extractors are compared. 90.4 % accuracy is obtained in this method, which is considered a drawback.

Deshpande et al. [22] presented the detection of common fungal diseases, common rust, and northern leaf blight in maize leaves. Two classifiers, namely K-NN and SVM, were used, aiming at the early detection and classification of diseases into common rust, northern leaf blight, multiple infections, or healthy using first-order histogram features and Haar wavelet features based on GLCM features. The highest accuracy of 85% is obtained with K-NN for K=5, and the accuracy obtained with SVM-based classification is 88 %. Mohanapriya et al. [23] introducedthe Naive Bayes classifier to detect unhealthy regions of plant leaves and also to classify them. Initially, the leaf images are collected, colour converted, segmented, feature extracted, and finally,the plant disease is classified.The accuracy of the result obtained is about 97 %. This method takes less measure of time and gets more efficient.Less success rate is obtained. This paper uses machine learning techniques to describe plant leaf disease diagnosis and categorization technology. We proposed an Automated Hybrid detection of plant disease using Random Forest with AdaBoost classifier that applies different pre-processing and feature extraction techniques. Random forest with the AdaBoost classification method is considered to classify disease from plants. We experimented with our proposed model in MATLAB using the Plant Village dataset. The proposed method improves the evaluation parameters when compared with existing methods.

2. OBJECTIVE

Several contributions were presented to develop an automated system for detecting plant leaf diseases. Three kinds of features: colour, texture, and geometric, are suggested. In this research, we present different stages for crop disease prevention using image analysis and machine intelligence to quickly, randomly, automatically and accurately diagnose and identify plant leaf diseases. This article aims to focus on identifying plant diseases based on different approaches, such as the analysis of images. The image analysis technique and the machine learning-based Hybrid Automated detection of plant disease using Random Forest with AdaBoost algorithmare effective and accurate fields for disease detection using plant leaf images. The performance of this system was evaluated using the Plant Village dataset.

Plantdiseasehasincreasedsignificantlyinrecentyears. Many studies in this area have concentrated on establishing solutions for plant disease detection in the early stage. The main objectives are

- To Propose a machine learning-based Hybrid Automated detection of plant disease using Random Forest with AdaBoost algorithm.
- The proposed method is trained in fourphases pre-processing, segregation, feature extraction, and classification.
- We experiment with the efficiency of our postulated methodology based on the Plant Village dataset. The reactivity, veracity, recall, and accuracy are improved.

The structure of the paper can be followed as follows: section 1 describes the introduction of plant diseases and existing methods. Section 2 reports the purpose of the postulated method. Section 3 recounts equivalent activities. Section 4 delineates the proposed mechanism. Section 5 designates the experimental setup. Section 6 explains the outcome and consideration. Section 7 discusses the conclusion. Section 8 expounds on references.

3. RELATED WORK

Many researchers worked on detecting and recognizing plant disease, adopting different concepts. Mohan et al. [24] introduced a technique for disease identification that employs Haar-like features to determine the affected part of the paddy plant. The reliability of disease diagnosis is 83.33 percent. The SIFT (Scale Invariant Feature Transform) characteristic recognizes diseases. K-Nearest Neighbour (K-NN), as well as Support Vector Machine (SVM) classifiers, are employed to categorize the illnesses (Brown Spot, Leaf Blast, and Bacterial Blight). With SVM, the accuracy is 91.1 %, and with K-NN, it's 93.33 %. Focusing on structural changes, Phadikar et al. [25] established an analytical approach to classify Brown Spot and Leaf Blast disease in rice plants. The image's clarity is improved by using the mean filtering approach. The image is segmented using Otsu's classification technique. The image's hue plane is utilized to choose the boundary. The radial dispersion of the colour from the centre to the edge of the spot images is obtained to categorize diseases. They used Bayes and SVM classifiers to diagnose and distinguish paddy plant diseases (Brown Spot and Leaf Blast), with a 79.5 and 68.1 percent reliability for the Bayes and SVM classifiers. Deshmukh and Radhika [26] presented a paddy leaf disease recognition methodology. The infected parts of the leaf are separated using the k-means algorithm. Then features from the infected part of the image are extracted. Contrast, homogeneity, correlation, and energy features calculated from GLCM (Gray-Level Co-occurrence Matrix). Only LL (Low-Low band) is used for discrete wavelet transform among four image subbands. The standard deviation and covariance are also extracted as features. In this system, thirteen features are used for classification. Ultimately, Backpropagation Neural Network (BPNN) algorithm is applied with 40 hidden layers to categorize the kinds (Leaf Blast, Brown Spot, and Healthy Leaf). Al Bashish et al. [27] postulated a software application predicated on image processing for automated leaf disease recognition and categorization. In the initial step, they constructed a colour modification framework for the RGB (Red, Green, and Blue) leaf image. The images are separated into four groups in the next step utilizing the k-means approach. If the leaf is infected with more than one disease, the disease could be in one or more of such four groupings. The contaminated segment's properties are obtained in the third step. For feature extraction, the Color Co-occurrence Matrix (CCM) is employed. Subsequently, five plant illnesses are diagnosed using a neural network with a backpropagation methodology. The accuracy of their method is around 89.5 %. K-means segmentation-based disease detection technique is postulated by Sethy et al. [28]. The infected segment is found through two phases. The input image is initially transformed to L*a*b colour space, and then the k-means method is deployed. The images are adjusted, and the intensity is improved to retrieve the essential details. The pixels are sorted depending on their geometric and colour properties. They have detected healthy and infected areas using k-means (k=3) clustering. Finally, the area of the infected part and the healthy part are measured. Islam et al. [29] postulated a methodology for determining the proportion of paddy leaf pixels damaged by the Leaf Blast disease. They segmented the image using the k-means classification framework and then determined the proportion of disease-affected pixels. At first, the leaf is placed horizontally on a white background to take as input. The original image is segmented into three clusters according to variation of colour using the k-means algorithm. They considered both damaged and undamaged leaf regions to compute the proportion of impacted pixels. The severity of diseases and measures to cure them are suggested for Leaf Blast disease. Parven et al. [30] presented a system that can classify paddy plant diseases using image processing and machine learning technique. They collected normal and diseased paddy plants from various regions. Then the backdrop is confiscated using a mask, and the output image is segmented using k-means clustering. Finally, four diseases (Brown Spot, Sheath Blight, Blast Disease, and Narrow Brown Spot) are classified using SVM. The system showed 94 % accuracy.

GorjiKandi et al. [31] suggested a method for the colour scene transformation between images by applying histogram matching or histogram specification algorithms (Resenfeld and Kak histogram matching algorithm). It is suggested that to transform the colour scene between two images, the histograms of the input image's three R, Gand B channels would be matched to the corresponding histograms of the destination one. The performance of the introduced method was investigated for several images. The obtained results indicated that this method could transform the colour scene between images.

Parvinzadeh et al. [32] suggested eggplant as a potential fruit waste for dyeing rug piles which is a valuable product. However, using some metal salts to obtain the desired colour fastness was a disadvantage. As a result, some synthetic dyes used in dyeing sectors could be replaced by different fruit wastes due to their ability to match the desired colour, added value, economical and environmental aspects and availabilities. In this research, dyeing wool fibres was carried out using the powdered skin of eggplant. For this purpose, the Iranian wool was first treated with some metal salts, including Fe(II), Sn(II), Cu(II), Cr(VI) and Al(III). These salts are commonly used as a mordant to improve the wash and light fastness of naturally dyed textiles. The wool fibre was then dyed with 50 % owfthe powdered skin of eggplant. The colourimetricproperties of the dyed yarns were evaluated using a reflectance spectrophotometer. Results showed that the skin of the eggplant is a potential source for dyeing wool fibres.

4. Proposed mthodology

Most of the traditional classifiers were unsuccessful in reducing the classification error and don't support multi-plant detection. To solve this issue, we propose a random forest classifier combined with the AdaBoost algorithm to detect plant diseases automatically. This section explains the overall workflow of plant disease identification, including our proposed model for classification. In Figure 1, the proposed model's block diagram is depicted.

Figure 1 illustrates the proposed hybrid Random forest with the AdaBoost algorithm for plant disease classification. The proposed method includes four phases (i) Pre-processing, (ii) Segmentation, (iii) Feature extraction using color histogram (iv) Classification using a hybrid Random forest AdaBoost algorithm. Images from the Plant village dataset are provided as input. Initially, the input image undergoes preprocessing stage. Pre-processing is mainly performed to resize the source image to fit into the RGB color space and to reduce noise. Following preprocessing, the resulting image is segmented. Clustering separates an image into distinct sections. The image is then subjected to feature extraction. The basic data gathered from photos to distinguish them is known as features. Feature extraction is done using the color histogram. It also converts RGB images to HSV and HSI. Consequently, the hybrid random forest with AdaBoost algorithm is proposed.

The proposed hybrid Random forest AdaBoost algorithm performs training and testing. Finally, the plant images with diseases are classified.

4.1. Pre-processing

Image preprocessing is required to boost the images for further exploration. Image resolution can have an impact on the computing time. Thus resizing the RGB color space (Figure 2a) is implemented to reduce the computation time. [26]

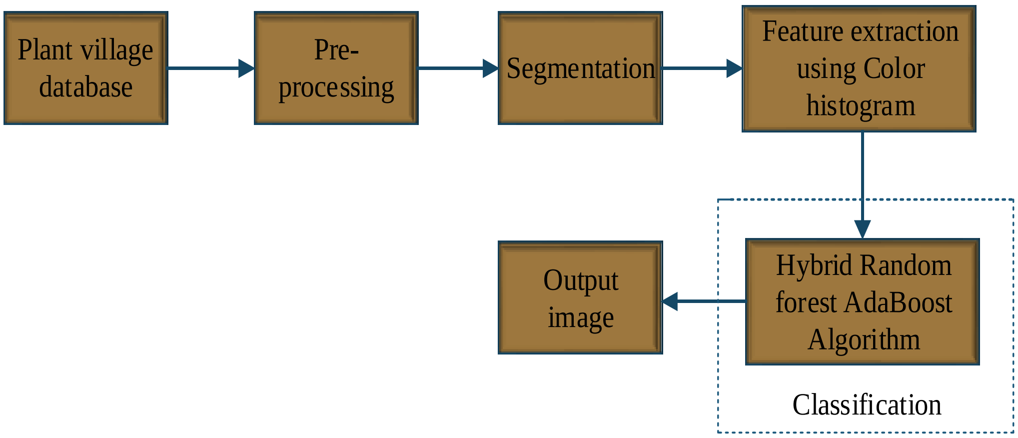

Figure 1: Proposed hybrid Random forest with AdaBoost algorithm.

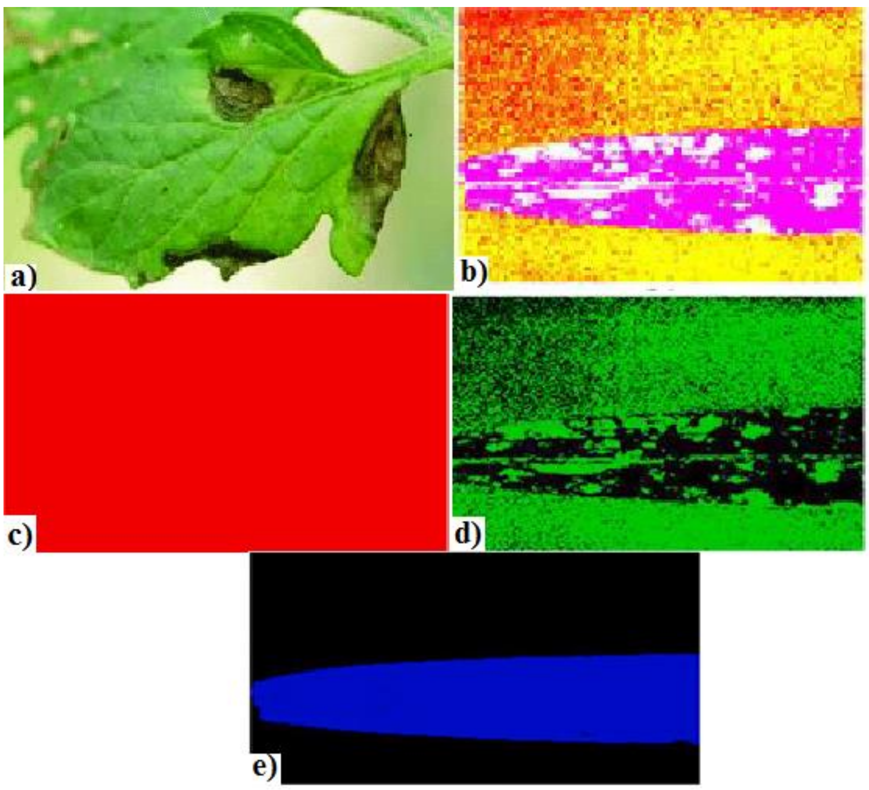

Figure 2: a) Original RGB image, b) L*a*b image, c) L channel, d) a channel and d) b channel.

The image resolution of 5184 × 3456 pixels is resized into 640 × 480 pixels. The RGB to L*a*b transformation is created and formulated on the upcoming equations [27].

- Transfiguring RGB to L*a*b (Eqs. 1-3)

$$L = 0.2126R + 0.7152G + 0.0722 \quad (1)$$

$$a = 1 \cdot 4749\,(0 \cdot 2213R - 0.339G + 0.1177B) + 128 \quad (2)$$

$$b = 0.6245\,(0 \cdot 1949R + 0.6057G - 0.8006B) + 128 \quad (3)$$

Figure 2a presents the original RGB image. Figure 2b presents the RGB to L*a*b color space transition outcome.Figure 2c illustrates the L channel. Figure 2d represents the channel. Figure 2e shows the b channel. In this mechanism, the b channel was preferred as the input for the next step since the leaf area stands out more from the backdrops in the L*a*b colour space than in other channels. Selecting asuitable colour space simplifies the entire procedure.

4.2. Segmentation

Segmentation aims to create a sub-picture known as a region of interest (ROI) image. The K-means grouping technology [28] was applied to individualize the leaf and backdrop. Since the region in the image separates into the leaf, affected area, and backdrop, the number of clusters (K) was first established hinged on the number of classes. In this example, K is three as the area in the image divides into the leaf, infected area, and backdrop. The following are the phases of this methodology [29]:

- Step 3: In the 'L*a*b*' space, colours are grouped using K-means, and the disparity among three colours is determined using Euclidean distance measure.
- Step 4: Every pixel in the image gets a tag with its designated cluster index.
- Step 5: The pixels in the input data are split by colour utilizing pixel labels, resulting in various image segments dependent on the number of clusters.

Figure 3 shows the segmentation results of the sample disease-affected leaf image. Here the image is segmented into 3 clusters because the image has three colors. The first cluster represents the leaf part, the second cluster represents the infected part, and the third cluster represents the background part of the image.

4.3. Feature extraction

The colour histogram is used to gather characteristics. The colour histogram shows how the colours of a picture are represented. The RGB to L*a*b transition was done using an equation 1-3. The histogram is computed after the L*a*b picture is translated to HSI and HSV colour space. The RGB image must be converted to HSV because the HSV model accurately matches how the human eye perceives colours. The histogram plot describes the number of pixels obtainable in the given colour spectrum [30].

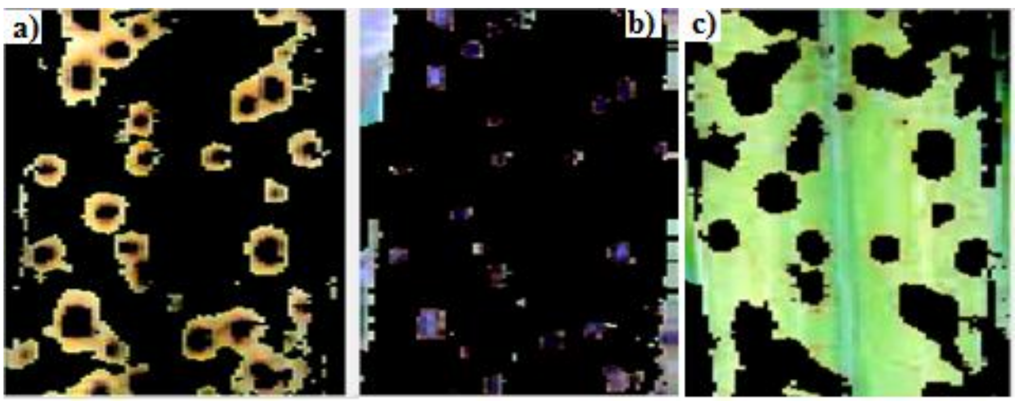

Figure 3: Segmentation results of the sample disease-affected leaf image a) clusters1, b) clusters 2 and clusters 3.

4.3.1. Converting RGB to HSI

HSI color space comprises three channels: Hue (H), intensity (I), and Saturation (S). The channels are generated using the given equations (Eqs. 4-6).

$$S = 1 - \frac{[\min(R,G,B)]]}{I} \tag{4}$$

$$I = \frac{1}{3}(R + G + B) \tag{5}$$

$$H = \begin{cases} \theta, B \leq G \\ 360 - \theta,\ B > G \end{cases} \tag{6}$$

Where, $\theta = \cos^{-1}\left\{\frac{1/2[(R-G)+(R-B)]}{[x(R-G)^2+(R-B)(G-B)]^{1/2}}\right\}$

4.3.2. Converting RGB to HSV

HSV color space comprises 3 channels: Hue (H), Value (V), and Saturation (S). H channel is calculated using equation 6. While the V and S channels are calculated by equations 7 and 8.

$$S = \begin{cases} 0, max(R,G,B) = 0 \\ 1 - \frac{mix(R,G,B)}{max\ (R,G,B)}, otherwise \end{cases} \tag{7}$$

$$V = maxf_0(R,G,B) \tag{8}$$

In this methodology, colour characteristics are generated utilizing colour histogram as this feature is a potent descriptor that has been effectively employed in several prior researches [22]. The histogram is simple to visualize, and the pixel strength dispersion reveals the disparity between infected and uninfected leaves. Twelve histograms were constructed by extracting characteristics from the subdivided leaf area histogram hinged on every channel from four colour spaces—RGB, L*a*b, HSI, and HSV. Firstly, every channel's histogram has 256 intensity values ranging from 0 to 255, resulting in a maximum of 3072 features. Since the number of parameters obtained is large, the histogram was segmented into 4, 8, 16, and 32 bins. For bins 4, 8, 16, and 32, the number of features obtained was 48, 96, 192, and 384, correspondingly.

4.4. Classification using proposed hybrid Random forest with AdaBoost algorithm

The suggested hybrid methodology is adaptable, allowing it to be utilized for categorization and reversion. Compared to previous machine learning algorithms, our suggested hybrid Random forest with AdaBoost produced higher precision with fewer image data sets. The flowchart for the suggested work is shown in Figure 4.

4.4.1. Steps involved in the proposed hybrid random forest with adaboost

Step 1: Start the process

Step 2: Segmented images undergo feature extraction using color histogram

Step 3: Use Hybrid Random forest with AdaBoost to classify plant images.

Step 4: Training and Testing are done using a Hybrid random forest with the AdaBoost algorithm

Step 5: Trained and tested images are used for the classification of plant images.

Step 6: Stop the process.

4.4.2. Training

The image data are dissected into two categories: training and testing. The feature vector for instructional data is created. A random forest algorithm is used to educate the obtained feature vector. In addition, the testing data feature vector is provided to the trained classifier for prediction.

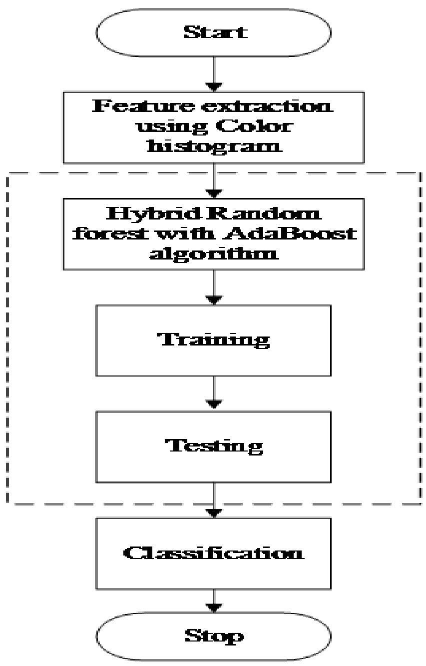

Figure 4: Flowchart of proposed work.

Initially, we grow an initial forest θ_0 with B_0 number of trees and initial feature vector $F_0(.)$. After the feature vector, we compute $u_0 = \sqrt{\#F_0(.)}$. Let the mean and standard deviation of feature weights be μ_0 and σ_0 correspondingly. Then R_0 is the initial subset of features whose subsets are less than $(\mu_0 - 2\sigma_0)$. The feature vector with a deducted set of features is $F_1(.). = F_0(.).$ -R_0. Next, we compute Δ_u and Δ_v using the given equation (Eq. 9).

$$\Delta_u = \#\, T_{n+1} - \#T_n$$

$$\Delta_v = \#\, \check{T}_{n+1} - \#\check{T}_n \tag{9}$$

Where T_{n+1} and T_n are the feature vectors. Based on these values, we trained the classifiers. At any pass n and reduces features set $F_n(.)$ We begin by ranking the characteristics. Eventually, we come upon a novel set of crucial features A_n and set of features to be deducted R_n. With decreased features, we get the feature vector $F_{n+1}(.). = F_n(.).$ -R_n and upgrade the feature bags utilizing the given equation (Eq. 10).

$$\begin{aligned} T_{n+1} &= T_n + A_n \\ \check{T}_{n+1} &= \check{T}_n - R_n - A_n \end{aligned} \tag{10}$$

Finally, we get the training vector θ_{n+1} with feature vector $F_{n+1}(.)$ (Eq. 11).

$$\theta_{n+1} = 1 - A_n = \frac{\binom{\Delta u}{\Delta v}}{\left(\frac{\Delta u + \Delta v}{F_{(n+1)}(.)}\right)} \tag{11}$$

Where θ_{n+1} represents the training vector, $F_{n+1}(.)$ represents the feature vector and Δ_u, Δ_v Represents the set of important and unimportant features.

Algorithm 1 shows the pseudo-code for classification using a Hybrid Random forest with the AdaBoost algorithm. First, the feature vector undergoes training and testing for classification. Calculate feature weights using the mean and standard weights of the feature vector. Update feature bags for training vector. After training, the images undergo testing for classification.

Algorithm 1: Pseudocode for Classification

***Input**: Plant village dataset*

***Output:** Infected plant image*

1. *Procedure Initialize* (θ_0)
2. *Grow initial forest* θ_0 *with feature vector* $F_0(.)$.
3. *While* $V_n \geq F$ *do*

- *Compute mean* μ_0 *and standard deviation* σ_0 *of feature weights in* $\check{T}_n$

4. *Find* $F_1(.). = F_0(.).$ -R_0.

5. *Find* $\Delta_u = \#\, T_{n+1} - \#T_n$ *and*

$$\Delta_v = \#\, \check{T}_{n+1} - \#\check{T}_n$$

6. *Find feature vector* $F_{n+1}(.). = F_n(.).$ -R_n

7. *Update feature bags using equation*

$$T_{n+1} = T_n + A_n$$

$$\check{T}_{n+1} = \check{T}_n - R_n - A_n$$

8. *Training vector*

$$\theta_{n+1} = 1 - A_n = \frac{\binom{\Delta u}{\Delta v}}{\left(\frac{\Delta u + \Delta v}{F_{(n+1)}(.)}\right)}$$

9. *end for*

10. *Compute* $err^{(m)} = \varepsilon^m w.\pi(C_i \neq T^{(m)}(x_i)) \sum_{i=1}^{n} w_i$

11. *Set* $w_i \leftarrow w_i.exp(\alpha^{(m)}.\pi\left(C_i \neq T^{(m)}(x_i)\right)$, *for i= 1,2,....n*

12. *Testing vector* $C^* = argmax \sum_{T=1}^{W} \alpha^*(T) P_T^*(c)$

13. *Update Classification result*

14. *End procedure*

4.4.3. Testing

The testing for classification is accomplished by creating a model from the training vector. Then, by raising the weights, create a second model that aims to fix the defects from the first. A score is allocated to each vector and the final testing vector for classification is well-defined as the linear combination of the vectors from each stage. Error in the training vector is given by equation 12,

$$err^{(m)} = \varepsilon^{m} w.\pi(C_i \neq T^{(m)}(x_i)) \sum_{i=1}^{n} w_i \quad (12)$$

As the training error is calculated, a weighted vector is applied for accessing the class label of the test data. Let the test data consist of B^* number of vectors and α^* represent the weight of vector. If $P_T^*(c)$ represents the feasibility of class label, projected by vector for the input test data, then the actual class label C^* for the input test data is given by (Eq. 13),

$$C^* = argmax \sum_{T=1}^{W} \alpha^*(T) P_T^*(c) \quad (13)$$

Equation 13 gives the output of testing data for classification.

5. Experimental setup:

The Plant Village Dataset is utilized in this paper. It's made up of photos of plant leaves shot in a lab setting. There are 54 306 photos of 14 distinct plant varieties in total, organised into 38 unique categories as species/disease pairs. Apple, Blueberry, Cherry, Corn, Grape, Orange, Peach, Bell Pepper, Potato, Raspberry, Soybean, Squash, Strawberry, and Tomato are among the varieties found in this data. This database contains photos of healthy plants from 12 distinct varieties, as well as photographs of 17 fungal infections, 4 bacterial diseases, 2 viral infections, 2 mould diseases, and one mite disease. Pictures were captured with a regular digital camera outdoor, under various climatic situations, and from many sources, resulting in a more diversified database. This database is useful for employing machine learning methods due to the high number of samples and variety of diseases. Color photos, grey scale images, and fragmented images with the backdrop concealed are all included in the database. Segmented pictures are employed in this paper.

6. Results and Discussion

6.1. Sensitivity

The sensitivity ratio is an essential statistic for extracting plant disease-related characteristics from fragmented plant images. The features gathered aid in determining if they are linked to regular or pathological characteristics. Plant image preprocessing is a key aspect in the unique model of plant disease identification approaches, assuming that numerous plant pictures with same modalities include identical information on image sensibility. Table 1 illustrates the comparison of proposed hybrid Random forest with AdaBoost method with existing methods. The sensitivity of the proposed hybrid Random forest with AdaBoost method is calculated by the formula given below (Eq. 14),

$$\text{Sensitivity} = \frac{TP}{TP+FP} \text{ X } 100\% \quad (14)$$

Table 1: Comparison of proposed hybrid Random forest With AdaBoost method with existing methods.

Total number of datasets	Proposed Method	KNN	Naïve Bayes	SVM
100	70	60	63	58
200	76	65	65	55
300	85	60	70	58
400	87	70	73	63
500	98	80	83	75

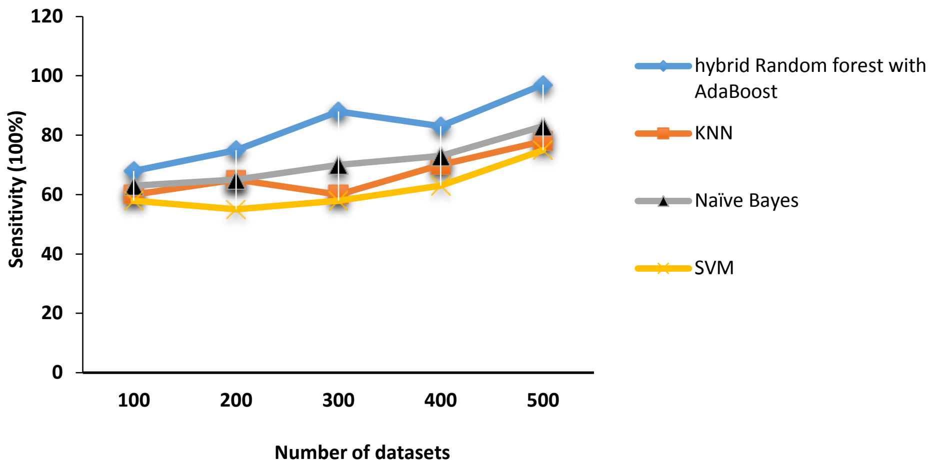

Figure 5: Sensitivity of proposed method with existing method.

Figure 5 shows the sensitivity of hybrid Random forest with AdaBoost. The hybrid Random forest with AdaBoost provides a high excellent sensitivity energy efficiency than the existing methods.

6.2. Specificity

Uniqueness for enhancement as they are the finest predictors of image quality. Considering diverse choosing criteria for high-quality image criterion, many scholars acquiring less values utilizing conventional methodologies criterion may be an essential measure utilized in gathering plant disease correlated characteristics from the fragmented image. The characteristics gathered aid in determining if the characteristics contribute to regular or aberrant characteristics. Table 2 illustrates the comparision of proposed hybrid Random forest with AdaBoost method with existing methods. The specificity of the proposed hybrid Random forest with AdaBoost is calculated by the formula given below (Eq. 15),

$$Specificity = \frac{TP}{TP+FP} X\ 100\% \qquad (15)$$

Figure 6 indicate the specificity of the postulated hybrid Random forest with AdaBoost method. The proposed hybrid Random forest with AdaBoost method has high energy efficiency and specificity than the existing methods.

6.3. Precision

Approaches for extracting features as well as some categorization techniques have been developed. As a result, only a few features were acquired, and plant disease identification had poor reliability. In addition, the KNN, Naive Bayes, and SVM algorithms lack the overlap measure, which is a sequential index and an essential criterion for evaluating the accuracy of any plant disease detection method (Figure 7). The precision ratio of the proposed hybrid Random forest with AdaBoost method is calculated by the formula given below (Eq. 16),

$$Precision\ = \frac{TP}{TP+FP} \qquad (16)$$

Table 2: Comparision of proposed hybrid Random forest with AdaBoost method with existing methods.

Total number of datasets	Proposed Method	KNN	Naive Bayes	SVM
100	70	65	66	65
200	80	73	73	70
300	85	75	80	63
400	83	65	75	60
500	95	80	83	75

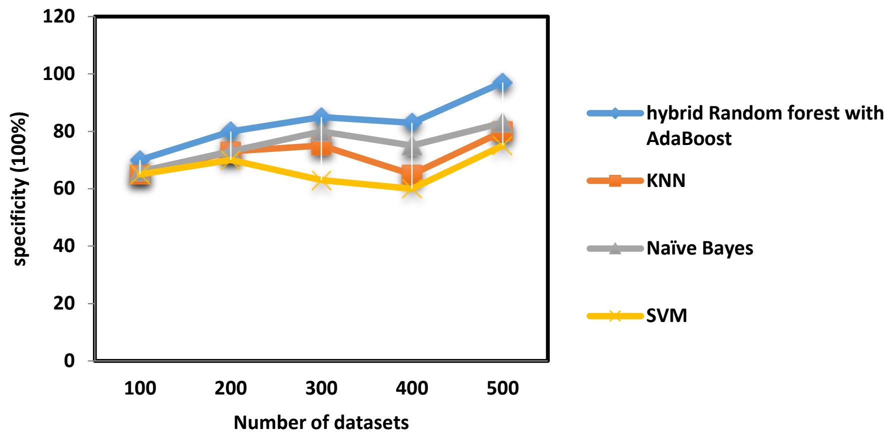

Figure 6: Sensitivity of proposed method with existing method.

Table 3 illustrates the comparision of postulated hybrid Random forest with AdaBoost method with existing methods.

Figure 7 shows the precision ratio of the hybrid Random forest with AdaBoost method. The novel hybrid Random forest with AdaBoost method has high energy efficiency and precision ratio than the existing methods.

6.4. Recall

The recall ratio is expressed as a percentage of the categorized image for all linked projections. The recall ratio is the proportion of noteworthy instances recovered, whereas recall is the proportion of pertinent instances obtained. Table 4 demonstrates that when contrasted to KNN, Naive Bayes, and SVM, the presented hybrid Random forest with AdaBoost approach has an excellent recall ratio. The Recall ratio of the proposed hybrid Random forest with AdaBoost method is calculated by the formula given below (Eq. 17),

$$Recall\ ratio = \frac{TP}{TP+FN} \quad (17)$$

Figure 8 demonstrates the Recall ratio of the hybrid Random forest with AdaBoost method. The hybrid Random forest with AdaBoost method has high energy efficiency and recall ratio than the existing methods.

Table 3: Comparision of proposed method with existing methods.

Total number of datasets	Proposed Method	KNN	Naïve Bayes	SVM
100	70	66.8	66.9	65.8
200	87	71.22	75	70
300	88	66	83	65.4
400	92	80	85	78
500	97	91	92	90

Table 4: Comparision of proposed hybrid Random forest with AdaBoost method with existing methods.

Total number of datasets	Proposed Method	KNN	Naïve Bayes	SVM
100	65	49	50	48
200	76	65	66	51
300	89	74	76	73.4
400	90	82	85	81
500	98.6	85	87	84

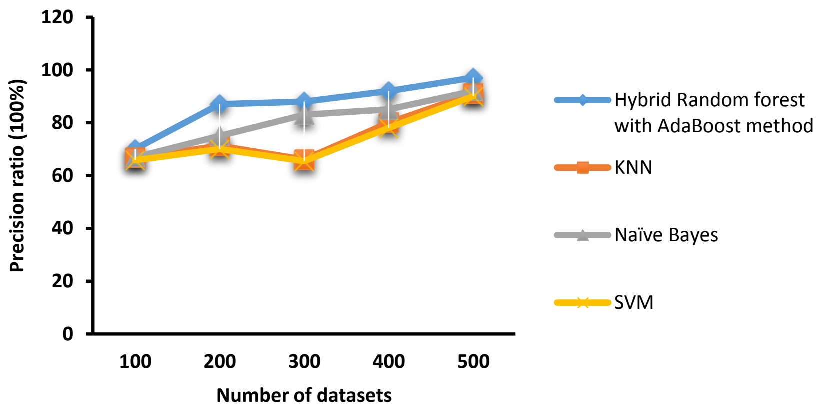

Figure 7: Precision of proposed method with existing method.

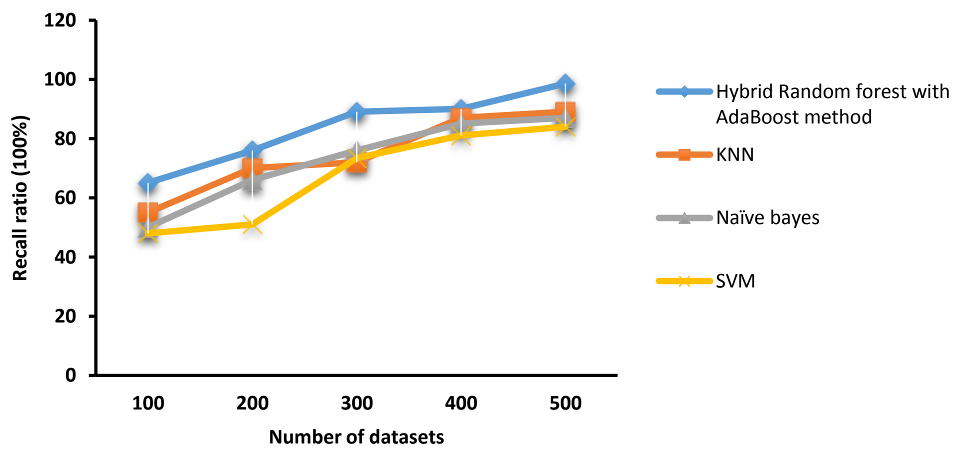

Figure 8: Recall ratio of proposed method with existing method

7. Conclusion

In this paper we propose a hybrid Random forest with AdaBoost methodology for plant disease identification. The postulated methodology is used in classification. The proposed hybrid Random forest with AdaBoost methodology contains four phases Preprocessing, Segmentation, Feature extraction using color histogram, classification using hybrid Random forest with AdaBoost method. The proposed hybrid Random forest with AdaBoost method is trained in three steps. In first phase the images are selected from the Image dataset for training. In the second step testing is done from the trained images. In the third step Classification of plant images is evaluated from tested data. Our proposed method uses Plant Village dataset and is implemented in MATLAB. The results shows the proposed hybrid Random forest with AdaBoost method has achieved a better sensitivity, specificity, recall ratio, precision ratio when compared with existing methods such as KNN, Naive Bayes, SVM.

5. References

1. Mattihalli C, Gedefaye E, Endalamaw F, Necho A. Plant leaf diseases detection and automedicine. Inter Thing. 2018; 1: 67-73.
2. Himani A. An analysis of agriculture sector in Indian economy. IOSR J Human Soc. 2014; 19(1): 47–54. Doi: doi.org/10.9790/0837-191104754.
3. Badage A. Crop disease detection using machine learning: Indian agriculture. Inter Res J Eng Technol. 2018; 5(9): 866-869.
4. Ganatra N, Patel A. A multiclass plant leaf disease detection using image processing and machine learning techniques. Inter J Emer Technol. 2020; 11(2): 1082-1086.
5. Iqbal Z, Khan MA, Sharif M, Shah J H, Ur Rehman MH, Javed K. An automated detection and classification of citrus plant diseases using image processing techniques: A review. Comp Electron Agr. 2018; 153: 12-32.
6. Pawar P, Turkar V, Patil P. Cucumber disease detection using artificial neural network. International Conference on Inventive Computation Technologies (ICICT). 2016; 3:1-5.
7. Saradhambal G, Dhivya R, Latha S, Rajesh R. Plant disease detection and its solution using image classification. Inter J Pure Appl Math. 2018; 119: 879-883.
8. Khairnar K, Rahul D. Disease detection and diagnosis on plant using image processing–a review. Inter J Comp Appl. 2014; 108(13): 36-38.
9. Narayanan BN, Djaneye-Boundjou O, Kebede TM. Performance analysis of machine learning and pattern recognition algorithms for malware classification. Aeros Electron. Conference (NAECON) , 2016; 338-342.
10. Luo L, Tang Y, Zou X, Ye M, Feng W, Li G. Vision based extraction of spatial information in grape clusters for harvesting robots. Biosyst Eng. 2016; 151: 90-104. Doi: org/10.1016/j.biosystemseng.2016.08.026.
11. Raja R, Nguyen TT, Vuong VL, Slaughter DC, Fennimore SA. Rtd-seps: real-time detection of stem emerging points and classification of crop-weed for robotic weed control in producing tomato. Biosyst Eng. 2020; 195: 152-171. Doi: org/10.1016/j.biosystemseng.2020.05.004
12. Zheng J, Fu H, Li W, Wu W, Yu L, Yuan S, Tao WYW, Pang TK, Kanniah KD. Growing status observation for oil palm trees using unmanned aerial vehicle (uav) images. ISPRS J. Photogramm. Remote Sens. 2021; 173: 95-121. Doi: org/10.1016/j.isprsjprs.2021.01.0088.
13. Li B, Zhao X, Zhang Y, Zhang S, Luo B. Prediction and monitoring of leaf water content in soybean plants using terahertz time-domain spectroscopy, Comp Electron Agr. 2020; 170: 105239. Doi: org/10.1016/j.compag.2020.105239
14. Raja R, Nguyen TT, Vuong VL, Slaughter DC, Fennimore SA, Rtd-seps: real-time detection of stem emerging points and classification of crop-weed for robotic weed control in producing tomato, Biosyst Eng. 2020; 195: 152-171. Doi: org/10.1016/j.biosystemseng.2020.05.004.
15. Iqbal MA, Talukder KH. Detection of potato disease using image segmentation and machine learning. International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), 2020; 43-47. doi: 10.1109/WiSPNET48689.2020.9198563.
16. Dhingra G, Kumar V, Joshi HD. A novel computer vision based neutrosophic approach for leaf disease identification and classification. Measurement. 2019; 135: 782-794. Doi.org/10.1016/j.measurement.2018.12.027.
17. Mohan KJ, Balasubramanian M, Palanivel S. Detection and recognition of diseases from paddy plant leaf images. Inter J Comp Appl. 2016; 144(12): 34-41.
18. Santanu P, Sil J, Kumar Das A. Classification of rice leaf diseases based on morphological changes. Intern J Info Electron Eng. 2012; 2(3): 460-463.
19. Deshmukh R, Deshmukh M. Detection of paddy leaf diseases. Inter J Comp Appl. 2015; 2(3): 8-10.
20. Al Bashish D., Malik B., Sulieman B. Detection and classification of leaf diseases using K-means-based segmentation and neural network based classification. Info Technol J. 2011; 10(2): 267-275. DOI: 10.3923/itj.2011.267.275.
21. Kumar SP, Negi B, Bhoi N. Detection of healthy and defected diseased leaf of rice crop using K-means clustering technique. Intern J Comp Appl. 2017; 157(1): 0975-8887.
22. Rashedul I, Rafiqul Islam M. An image processing technique to calculate percentage of disease affected pixels of paddy leaf. Intern J Comp Appl. 2015; 123(12): 28-34.
23. Nargis P. Detection and recognition of paddy plant leaf diseases using machine learning technique. Inter J Innov Technol Expl Eng. (IJITEE). 2020; 9(5): 634-

638.

24. Ramesh S, Vydeki D. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. Inf Proc Agr. 2020; 7(2): 249-260.
25. Marshia Binth-E M, Mehrab Hossain M. Rice doctor: paddy disease detection from leaf image using machine learning, 2020.
26. Saleem G, Akhtar M, Ahmed N, Qureshi W. Automated analysis of visual leaf shape features for plant classification. Comp Electron Agri. 2019; 157: 270-280. Doi: org/10.1016/j.compag.2018.12.038
27. Zhuang J., Luo S., Hou C., Tang Y., He Y., Xue X. Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications. Comp Electron Agr. 2018; 152: 64-73. Doi: org/10.1016/j.compag.2018.07.004.
28. Zhang S, Wu X, You Z, Zhang L. Leaf image based cucumber disease recognition using sparse representation classification. Comp Electron Agr. 2017; 134: 135-141. Doi: org/10.1016/j.compag.2017.01.014.
29. Tian K, Li J, Zeng J, Evans A, Zhang L. Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. Comp Electron Agr. 2019;165: 104962. Doi:org/10.1016/j.compag.2019.104962.
30. Garcia-Lamont F, Cervantes J, Lopez A, Rodriguez L. Segmentation of images by color features: a survey. Neurocomputing. 2018; 292: 1-27. Doi: org/ 10.1016/j.neucom.2018.01.091.

How to cite this article:
Greeshma OS, Sasikala P, Balakrishnan SG. Automated Detection of Plant Disease Based on Color Histogram Feature Selection Using Hybrid Random Forest with Adaboost Algorithm. Prog Color Colorants Coat. 2024;17(1):39-52. Doi: 10.30509/pccc.2023.167095.1202