SEYBOLD Report
ISSN: 1533 - 9211
https://seyboldreport.net/

AUTOMATED DETECTION OF PLANT DISEASE BASED ON COLOR HISTOGRAM FEATURE SELECTION USING HYBRID RANDOM FOREST WITH ADABOOST ALGORITHM

Greeshma O S[1], Dr. P Sasikala[2], Dr. S G Balakrishnan[3]

[1]Assistant Professor, Department of Computer Science, Sree Sankara Vidyapeetom College, Nagaroor.

[2]Professor & Head, Department of Mathematics,VMKV Engineering College, Salem.

[3]Associate Professor, Department of CSE, Mahendra Engineering College, Namakkal.

Corresponding Author Email: osgreeshma@gmail.com

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.

Key words: Random forest, Color histogram, Classification, Feature extraction.

1. Introduction

Agriculture has a significant role in India's economy. India is the second-largest agricultural producer worldwide [1]. Approximately 75% of India 's inhabitants is dependent on agriculture, either explicitly or implicitly. [2]. In India, due to the rising population, there is a need for advances in agriculture to meet food needs. To withstand the transforming environment of the Indian economy, the agriculture industry requires a massive overhaul [3]. Plant disease impedes a plant's regular development, and it's one of the prime stimuli of pruned productivity that generate financial losses. Plant diseases can affect several plant components, including the leaf, stem, and seed. Detecting plant diseases promptly is a must, since it increases yields by more than 60 percent [4]. A confined disease control approach is one of the primary causes of crop losses. Infectious and non-infectious plant diseases are the two types of plant disease. Infectious plant diseases are mainly provoked by fungus, viruses, and bacteria [5]. To deliver effective treatments for diseases, diagnosing them swiftly and precisely is vital. It is primarily hinged on the symptoms that are apparent in a sick plant. As previously mentioned, illness can harm any component of a leaf, including the blossom, stem, and root. On the other hand, leaf examination is regarded as the optimal tool for plant diagnosis [6]. The diseased leaf is typically deformed in shape, colour, size, etc.

The conventional procedure of using a specialist to identify plant leaf disease is antiquated, time-consuming, and inappropriate [7]. As a result, an automated, efficient, and less economical system for detecting diseases from images and suggesting appropriate solutions is required [8]. In machine learning and computer vision, automated plant recognition is a hot topic [9.] Computer vision is constantly explored in this discipline, where picture or video data are input into the system. Existing work has been conducted for various objectives, including crop and land control, image classification for disease diagnosis, and plant disease detection [10, 11]. Furthermore, if the principal item in the image merges with other entities termed noises, the execution gets more complex [12, 13]. Machine learning has developed as a robust computing methodology for resolving a wide range of complicated computer vision challenges. [14] Machine learning enables computers to operate without any need for human involvement. Machine learning allows a system to operate on its own and make predictions. There are three machine learning methodologies: supervised, reinforcement, and unsupervised [15]. Pre-processing, segmentation, feature extraction, and classification are the four phases in machine learning application for leaf disease diagnosis [16].

This paper speaks for the technology for plant leaf disease diagnosis and categorization applying machine learning techniques. We proposed an Automated Hybrid detection of plant disease using Random Forest with AdaBoost classifier that applies different techniques for preprocessing, and feature extraction. 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

Plant disease has increased significantly in recent years. Many studies in this area have concentrated on establishing solutions for plant disease detection in the early stage. We make the following contributions to this paper:

- Proposed a version of machine learning-based Hybrid Automated detection of plant disease using Random Forest with AdaBoost algorithm.
- The proposed method is trained in four phases preprocessing, 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 postulated method. Section 3 recounts equivalent activities. Section 4 delineates the postulated mechanism. Section 5 designates experimental setup. Section 6 explains outcome and consideration. Section 7 discusses conclusion. Section 8 expounds references.

3. Related work

Many researchers worked on detecting and recognizing plant disease, adopting different concepts. **Mohan et al. [17]** 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 is employed to recognize 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. [18]** 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 color from center 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 reliability of 79.5 and 68.1 percent for Bayes and SVM classifier, accordingly. **Deshmukh and Radhika et al. [19]** presented a methodology for paddy leaf disease recognition. To separate infected parts of the leaf, K-means algorithm is used. Then features from the infected part of the image are extracted. Contrast, homogeneity, correlation, and energy features are calculated from GLCM (Gray-Level Co-occurrence matrix). Only LL (Low-low band) is used for discrete wavelet transform among four subbands of the image. The standard deviation and covariance are also extracted as features. In this system, thirteen features are used for classification. Ultimately, Back-propagation 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. [20]** 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 photos 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, employing a neural network with a back-propagation methodology, five plant illnesses are diagnosed. The accuracy of their method is around 89.5%. K-means segmentation-based disease detection technique is postulated by **Sethy et al. [21].** The infected segment is found through two phases. Initially, the input image is transformed to L*a*b color space, and then the K-means method is deployed. The photos are adjusted, and the intensity is improved to retrieve the essential details. The pixels are sorted depending on their geometric and color 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. [22]** postulated a methodology for determining the proportion of paddy leaf pixels damaged by the Leaf Blast disease. They segmented the actual 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 color using the K-means algorithm. To compute the proportion of impacted pixels, they considered both damaged and undamaged leaf regions. The severity of diseases and measures to cure are suggested for Leaf Blast disease. **Parven et al. [23]** 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. **Ramesh et al. [24]** employed an Improved Deep Neural Network with Jaya Algorithm to recognize and classify paddy leaf diseases. Normal, bacterial blight, brown spot, sheath rot, and blast illnesses are seen in the photos taken in a farming area. The backdrop was eliminated during preprocessing, and RGB photos are modified to HSV. The disease categorization is done with the help of the Jaya Optimization Algorithm and the Improved Deep Neural Network. The methodology is accurate for leaf blast (98.9%), bacterial blight (95.78%), sheath rot (92%), brown spot (94%), and normal leaf image (90.57%). **Mosharof et al. [25]** proposed a working model for paddy disease identification. They looked at Blast, Plant Hopper, and Leaf Folder, three diseases that affect rice. They have applied a convolution neural network for classification. They have also applied the transfer learning model for classification. They developed the model and projected over the data after analyzing it. They discovered that the CNN model had a 99.89 percent reliability rate.

4. Proposed Methodology

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 (Shima Ramesh et al. 2020) combined with the AdaBoost algorithm (Hui Yan et al. 2020) 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.

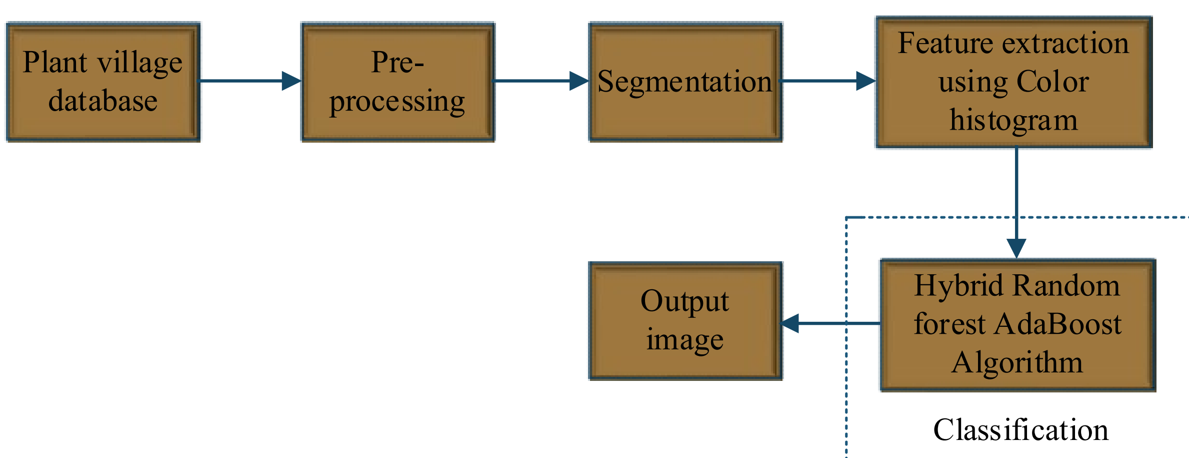


Figure 1: Proposed hybrid Random forest with AdaBoost algorithm

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 (Fig: 2a) is implemented to reduce the computation time. [26]

Figure 2(a): Original RGB image

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

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

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

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

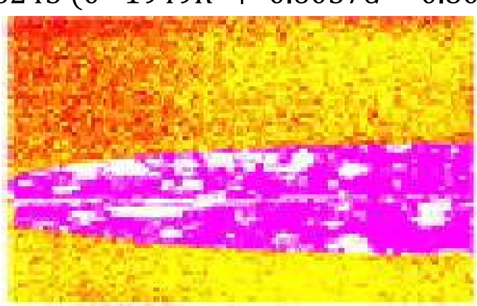


Figure 2(b): L*a*b image

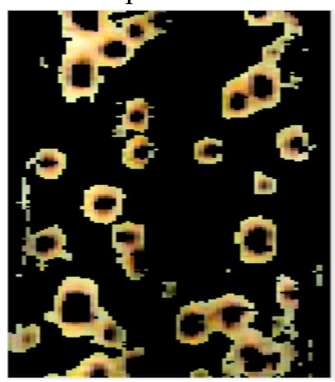
In Fig. 2(b), the outcome of the RGB to L*a*b color space transition is presented. In this mechanism, the b channel was preferred as the input for the next step since the leaf area stands out more from the backdrops in the L*a*b color space than in other channels. Selecting the right color space simplifies the entire procedure.

4.2 Segmentation

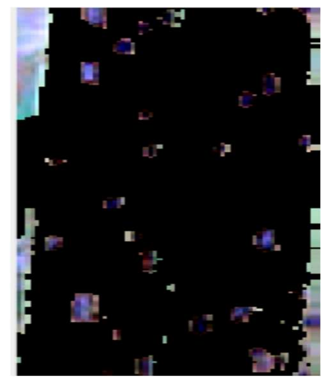
Segmentation aims to create a sub-picture known as a region of interest (ROI) image. The K-means grouping technology [28] was applied to individualize the leaf and backdrop. Since the region in the image separates into the leaf, affected area, and backdrop, the number of clusters (K) was first established 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.

Picture 3(a): Cluster 1



Picture 3(b): Cluster 2



Picture 3 (c): cluster 3

Picture 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].

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.

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

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

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

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.

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

$$V = max f_0(R,G,B) \quad (8)$$

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 Picture 4.

SEYBOLD Report
ISSN: 1533 - 9211

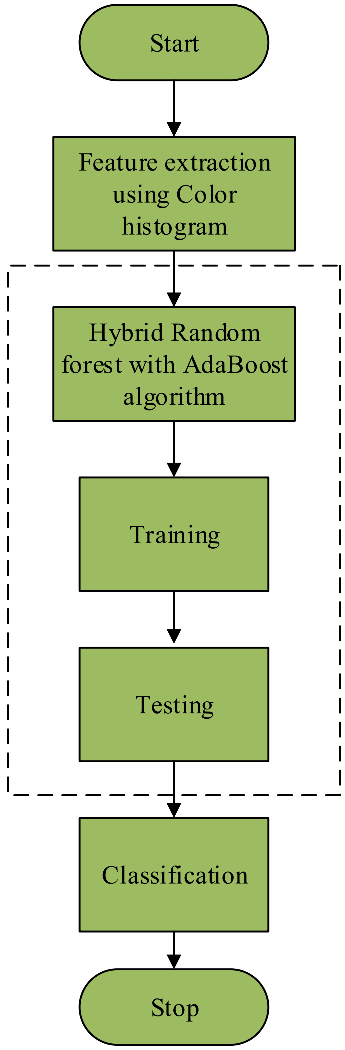


Figure 4: Flowchart of proposed work

4.3.1 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.

4.3.2 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.

Initially, we grow an initial forest $\theta_0$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 $\Delta_u$and Δ_v using the given equation,

$$\Delta_u = \#\, T_{n+1} - \#T_n$$
$$\Delta_v = \#\, \check{T}_{n+1} - \#\check{T}_n \qquad (9)$$

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,

$$T_{n+1} = T_n + A_n$$
$$\check{T}_{n+1} = \check{T}_n - R_n - A_n \qquad (10)$$

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

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

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: 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*

$$\Delta_u = \# T_{n+1} - \# T_n \text{ and}$$
$$\Delta_v = \# \check{T}_{n+1} - \# \check{T}_n$$

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*

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

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*

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.

4.3.3 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,

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

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. Result and discussions

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,

$$Sensitivity = \frac{TP}{TP + FP} X\ 100\%$$

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

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

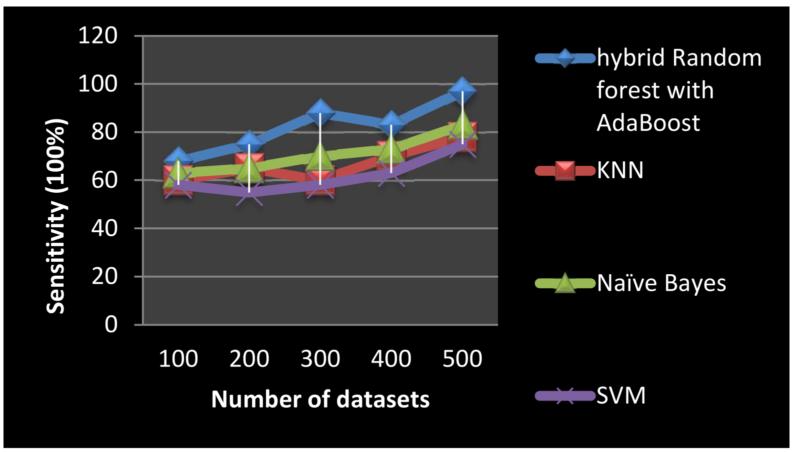


Figure 5: Sensitivity of proposed method with existing method

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,

$$Specificity = \frac{TP}{TP + FP} X\ 100\%$$

ISSN: 1533 - 9211

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

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

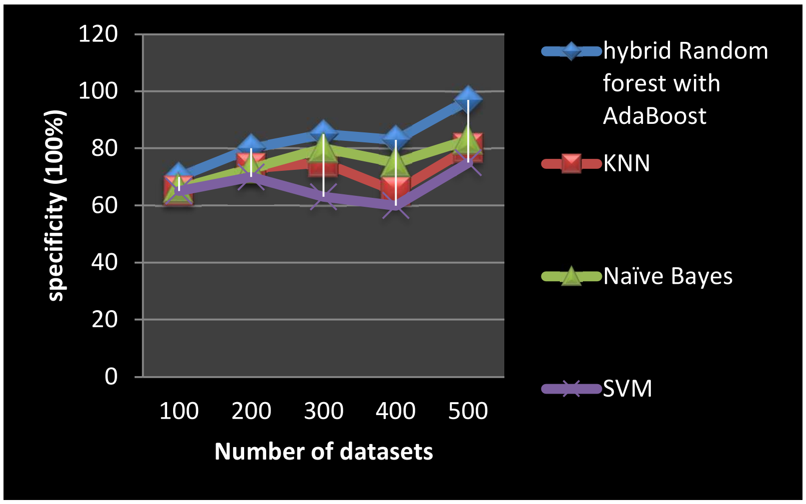


Figure 6: Sensitivity of proposed method with existing method

Picture 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. The precision ratio of the proposed hybrid Random forest with AdaBoost method is calculated by the formula given below,

$$Precision = \frac{TP}{TP + FP}$$

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 3: Comparision of proposed method with existing methods

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

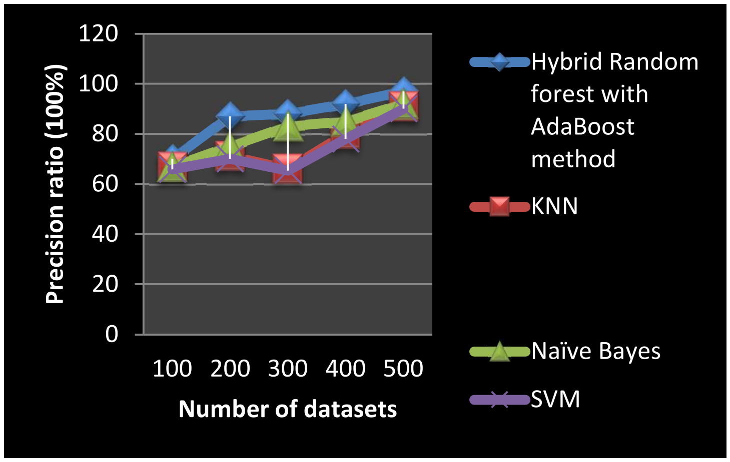


Figure 7: Precision of proposed method with existing method

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,

$$Recall\ ratio = \frac{TP}{TP+FN}$$

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

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

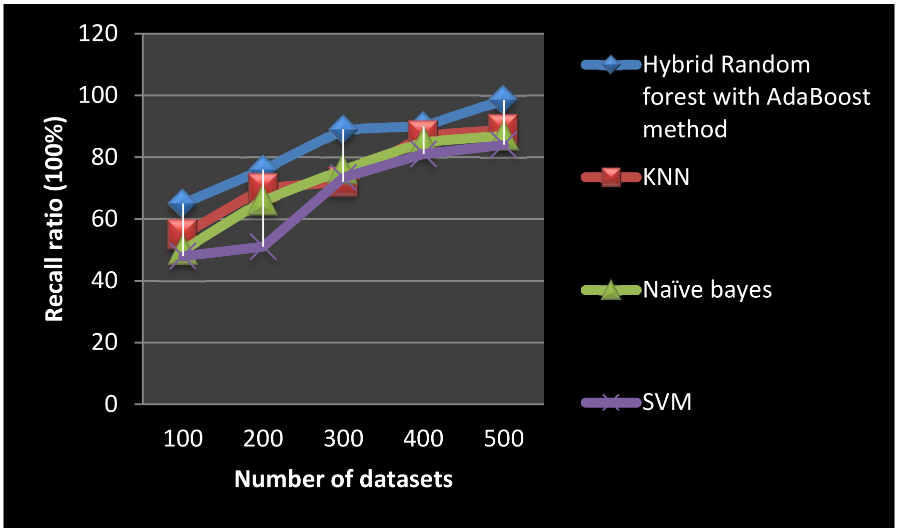


Figure 8: Recall ratio of proposed method with existing method

Picture 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.

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.

8. References

[1]. Mattihalli, C., Gedefaye, E., Endalamaw, F., & Necho, A. (2018). Plant leaf diseases detection and automedicine. Internet of Things, Volume: 1, Page: 67-73.

[2] Himani, 2014. An analysis of agriculture sector in Indian economy. IOSR J. Human. Soc. Volume: 19 (1), Page: 47–54. Doi: doi.org/10.9790/0837-191104754.

[3] Badage A. Crop disease detection using machine learning: Indian agriculture. International. Research. Journal. Engineering. Technology. 2018 September, Volume: 5(9), Page: 866-9.

[4] Ganatra N, Patel A. A multiclass plant leaf disease detection using image processing and machine learning techniques. International Journal. Emerging. Technology 2020 Apr, Volume: 11(2), page: 1082-6

[5] Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., ur Rehman, M. H., & Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing

techniques: A review. Computers and electronics in agriculture, Volume: 153, Page: 12-32
[6] Pawar, P., Turkar, V., & Patil, P. (2016, August). Cucumber disease detection using artificial neural network. In 2016 International Conference on Inventive Computation Technologies (ICICT) (Vol. 3, pp. 1-5). IEEE.
[7] Saradhambal, G., Dhivya, R., Latha, S. & Rajesh, R. (2018). Plant disease detection and its solution using image classification. International Journal of Pure and Applied Mathematics. Volume: 119, Page: 879-883.
[8] Khairnar, Khushal, and Rahul Dagade. " Disease Detection and Diagnosis on Plant using Image Processing–A Review." International Journal of Computer Applications 108, no. 13 (2014): Page: 36-38.
[9] Narayanan, Barath Narayanan, Ouboti Djaneye-Boundjou, and Temesguen M. Kebede. " Performance analysis of machine learning and pattern recognition algorithms for malware classification." In Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS), 2016 IEEE National, page. 338-342. IEEE, 2016
[10] L. Luo, Y. Tang, X. Zou, M. Ye, W. Feng, G. Li, Vision based extraction of spatial information in grape clusters for harvesting robots, Biosyst. Engineering. 151 (2016) 90–104, Doi: org/10.1016/j.biosystemseng.2016.08.026.
[11] R. Raja, T.T. Nguyen, V.L. Vuong, D.C. Slaughter, S.A. Fennimore, Rtd-seps: real-time detection of stem emerging points and classification of crop-weed for robotic weed control in producing tomato, Biosyst. Eng. 195 (2020) 152–171, Doi: org/10.1016/j.biosystemseng.2020.05.004
[12] J. Zheng, H. Fu, W. Li, W. Wu, L. Yu, S. Yuan, W.Y.W. Tao, T.K. Pang, K.D. Kanniah, Growing status observation for oil palm trees using unmanned aerial vehicle (uav) images, ISPRS J. Photogramm. Remote Sens. 173 (2021) 95–121, Doi: org/10.1016/j.isprsjprs.2021.01.0088
[13] B. Li, X. Zhao, Y. Zhang, S. Zhang, B. Luo, Prediction and monitoring of leaf water content in soybean plants using terahertz time-domain spectroscopy, Computer. Electronic. Agriculture. 170 (2020), 105239, Doi: org/10.1016/j.compag.2020.105239
[14] R. Raja, T.T. Nguyen, V.L. Vuong, D.C. Slaughter, S.A. Fennimore, Rtd-seps: real-time detection of stem emerging points and classification of crop-weed for robotic weed control in producing tomato, Biosyst. Eng. 195 (2020) 152–171, Doi: org/10.1016/j.biosystemseng.2020.05.004
[15] M. A. Iqbal and K. H. Talukder, "Detection of Potato Disease Using Image Segmentation and Machine Learning," *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET),* 2020, pp. 43-47, doi: 10.1109/WiSPNET48689.2020.9198563.
[16] G. Dhingra, V. Kumar, H.D. Joshi, A novel computer vision based neutrosophic approach for leaf disease identification and classification, Measurement 135 (2019) 782–794, Doi.org/10.1016/j.measurement.2018.12.027
[17] Mohan K Jagan, Balasubramanian M, Palanivel S. Detection and recognition of diseases from paddy plant leaf images. International Journal Computer Application. 2016; 144(12): 34–

41p.
[18] Phadikar Santanu, Jaya Sil, Asit Kumar Das. Classification of rice leaf diseases based on morphological changes. International Journal Info Electronics Eng. 2012; 2(3): 460–463p.
[19] Deshmukh R, Deshmukh M. Detection of Paddy Leaf Diseases. International Journal Computer Application. 2015; Volume: 2(3): page: 8–10.
[20] Al Bashish Dheeb, Malik Braik, Sulieman Bani-Ahmad. Detection and classification of leaf diseases using K-means-based segmentation and neural network based classification. *Info Technol J*. 2011; 10(2): 267–275p.
[21] Sethy Prabira Kumar, Baishalee Negi, Nilamani Bhoi. Detection of healthy and defected diseased leaf of rice crop using K-means clustering technique. International Journal Computer Application. 2017; 157(1): 0975–8887p.
[22] Islam Rashedul, Md Rafiqul Islam. An image processing technique to calculate percentage of disease affected pixels of paddy leaf. International Journal Computer Application. 2015; 123(12): 28–34p.
[23] Parven Nargis, *et al.* Detection and Recognition of Paddy Plant Leaf Diseases using Machine Learning Technique. International Journal Innovative Technology Explore Engineering (IJITEE). 2020; 9(5): 634–638p.
[24] Ramesh S, Vydeki D. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. Information process Agriculture. 2020; 7(2): 249–260p.
[25] Mosharof Marshia Binth-E, Md Mehrab Hossain. Rice Doctor: Paddy Disease Detection from Leaf Image Using Machine Learning, 2020.
[26] G. Saleem, M. Akhtar, N. Ahmed, W. Qureshi, Automated analysis of visual leaf shape features for plant classification, Computer. Electronics. Agriculture. 157 (2019) 270–280, Doi: org/10.1016/j.compag.2018.12.038
[27] J. Zhuang, S. Luo, C. Hou, Y. Tang, Y. He, X. Xue, Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications, Computer. Electronics. Agriculture. 152 (2018) 64–73, Doi: org/10.1016/j.compag.2018.07.004.
[28] S. Zhang, X. Wu, Z. You, L. Zhang, Leaf image based cucumber disease recognition using sparse representation classification, Computer. Electronics. Agriculture. 134 (2017) 135–141, Doi: org/10.1016/j.compag.2017.01.014.
[29] K. Tian, J. Li, J. Zeng, A. Evans, L. Zhang, Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm, Computer. Electronics. Agriculture. 165 (2019), 104962, Doi:org/10.1016/j.compag.2019.104962.
[30] F. Garcia-Lamont, J. Cervantes, A. Lopez, L. Rodriguez, Segmentation of images by color features: a survey, Neurocomputing 292 (2018) 1–27, Doi: org/ 10.1016/j.neucom.2018.01.091