Hindawi
Advances in Human-Computer Interaction
Volume 2021, Article ID 9367778, 15 pages
https://doi.org/10.1155/2021/9367778

Research Article

A Perspective View of Cotton Leaf Image Classification Using Machine Learning Algorithms Using WEKA

Bhagya M. Patil[1] **and Vishwanath Burkpalli**[2]

[1]*PDA College of Engineering, Kalaburgi, Karnataka, India*
[2]*Department of Information Science & Engineering, PDA College of Engineering, Kalaburgi, Karnataka, India*

Correspondence should be addressed to Bhagya M. Patil; patil.bhagya@gmail.com

Received 1 May 2021; Revised 2 June 2021; Accepted 1 July 2021; Published 15 July 2021

Academic Editor: Francesco Bellotti

Cotton is one of the major crops in India, where 23% of cotton gets exported to other countries. The cotton yield depends on crop growth, and it gets affected by diseases. In this paper, cotton disease classification is performed using different machine learning algorithms. For this research, the cotton leaf image database was used to segment the images from the natural background using modified factorization-based active contour method. First, the color and texture features are extracted from segmented images. Later, it has to be fed to the machine learning algorithms such as multilayer perceptron, support vector machine, Naïve Bayes, Random Forest, AdaBoost, and K-nearest neighbor. Four color features and eight texture features were extracted, and experimentation was done using three cases: (1) only color features, (2) only texture features, and (3) both color and texture features. The performance of classifiers was better when color features are extracted compared to texture feature extraction. The color features are enough to classify the healthy and unhealthy cotton leaf images. The performance of the classifiers was evaluated using performance parameters such as precision, recall, *F*-measure, and Matthews correlation coefficient. The accuracies of classifiers such as support vector machine, Naïve Bayes, Random Forest, AdaBoost, and K-nearest neighbor are 93.38%, 90.91%, 95.86%, 92.56%, and 94.21%, respectively, whereas that of the multilayer perceptron classifier is 96.69%.

1. Introduction

In India, agriculture is the main occupation, and two-thirds of the population is dependent on agriculture directly or indirectly. The yield of the crop depends on the growth, and diseases might affect the outcome of the crop. However, for farmers, it will be challenging to identify the disease with naked eyes. Therefore, identifying plant diseases at the early stage will benefit in diagnosing and preventing unnecessary crop loss. Among different parts of the plant, the leaf is the part that affects the crop yield if it gets affected. Visible symptoms can help in the detection of disease, and plant pathologists can suggest a suitable pesticide. In earlier days, disease identification was performed by taking the leaf sample and checking the disease type or the condition using a microscope, or else experts would identify the disease. If sufficient facilities are not there, then the farmers need to contact the experts for further action. However, this approach will be time-consuming. Therefore, the automatic detection of the disease will help the farmers to overcome the yield loss. There are numerous image processing methods to achieve this and are used for processing the images and identifying the disease. Here, different leaf disease classification is performed, taking, for example, leaves of pomegranate [1], grapes [2], tomato [3], and custard apple [4] for disease analysis. Figure 1 shows the basic image processing steps for classification.

Usually, the acquired leaf images with the natural background are filtered using the Gaussian mask. The filtered images are used for segmentation using the modified factorization-based active contour method, followed by feature extraction. However, if we use all the features in the classification, training time will be more; therefore, the selection of features is performed [5].

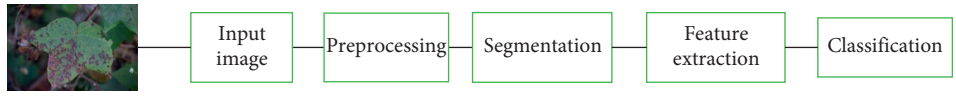

Figure 1: Leaf disease classification.

Finally, the classification of cotton leaf images is performed using different classifiers as shown in Figure 2.

Among the different steps involved in this process, after feature extraction, feature selection is used to improve the performance of the classifier. There are various color and texture features extracted to get the accurate classification accuracy. In the literature, many authors have submitted surveys on the performances of different classifiers. In this paper, based on the features selected [6], the performance of the classifiers is compared. The classifiers such as neural network, support vector machine, AdaBoost, Naïve Bayes, and Random Forest are considered in this study.

Different classifiers have advantages and disadvantages concerning different parameters such as training data size. Using these classifiers, authors have researched different leaf disease classification, especially in cotton. These classifiers have application in breast cancer diagnosis problem, heart disease prediction system [7], prediction of economic events [8], text categorization [9], skin disease diagnosis [10], medical science [11], face recognition [12], health science [13], brain tumor diagnosis [14], terrain classification [15], real-time facial expression recognition [16], discrimination of breast tumors in ultrasonic images [17], cancer genomics [18], detection of skin cancer [19], skin lesion segmentation [20], malignant melanoma detection [21], and disease detection in pomegranate leaf and fruit [22].

Apart from the survey of leaf classification, even an automatic detection system can be helpful in disease identification, and many researchers have introduced different methods for disease classification.

The organization of this paper is as follows. Section 2 provides the material and method, Section 3 provides related works, Section 4 describes methodology, and Section 5 gives the results and discussion followed by the conclusion in Section 6.

2. Material and Method

2.1. Database. For this study, cotton leaf images were considered. The images were captured with a natural background from the cotton field under controlled conditions. The database consists of nearly 300 images. There are two categories, healthy and diseased, used for training and testing. The sample images are shown in Figure 3.

The database consists of 150 healthy images, whereas the number of diseased images is nearer to 150. The diseases captured are Alternaria, Grey Mildew, and Cercospora leaf spot. The images were captured from various regions of Karnataka, India. The images acquired were larger, so before processing, it was resized to 256×256 size.

3. Related Works

Many researchers have published papers on leaf disease classification. We will summarize some of the papers as follows.

Gupta et al. [23] proposed an improved artificial plant optimization algorithm that helps to classify healthy and diseased parts. The database used for experimentation is a private dataset consisting of 236 images. A histogram of oriented gradients (HOG) was used for feature extraction. The performance of the proposed algorithm is 97.45%, and it was compared with k-nearest neighbors, support vector machine, Random Forest, and convolutional neural network. Kumari et al. [24] presented a classification of cotton diseased leaf spots using image processing techniques. K-means clustering was used for segmenting the diseased part of the leaf. Later, extracted features such as contrast, correlation, energy, and homogeneity were fed to the neural network classifier; the accuracy of the neural network classifier is 92.7%. Krithika and Grace Selvarani [25] proposed grape leaf disease classification using color histograms, and GLCM features were extracted. The KNN classifier was used for classification. Sarangdhar et al. [26] proposed a system for detecting and controlling the cotton disease classification. Five different types of diseases were identified, and the implementation was performed using an app. The overall accuracy of the classifier is 83.26%. The paper also deals with soil quality monitoring systems.

Panigrahi et al. [27] focused on maize leaf disease classification using different classifiers. The performance analysis of classifiers is analyzed performed, and it was found that the Random Forest classifier suits well for their dataset. The accuracy of the classifier is 79.23%. Vaishnnave et al. [28] performed groundnut leaf disease classification using a KNN classifier. Here, the authors categorized four different diseases. The fast feature method is used for feature extraction, and the same is feed to the KNN classifier. Usually, the database images used for the experiment have a plain background. Mokhtar [29] introduced a tomato leaf disease classification using image processing techniques. The GLCM features were extracted, and it helps to identify the healthy and diseased parts. In this paper, 800 images were used for experimentation. The performance of the SVM classifier is higher when compared to that of the other classifiers. Shrivastava and Pradhan [30] used 172 color features for each 14 different color spaces. The features were extracted from 619 images. The performance of seven different classifiers is used for comparison, and among them, the SVM classifier has achieved 94.65% accuracy. The author classified the rice plant diseases into four classes. Hossain et al. [31] proposed that the KNN classifier achieves better classification using the Arkansas plant disease database and Reddit leaf disease database. The authors converted the input RGB

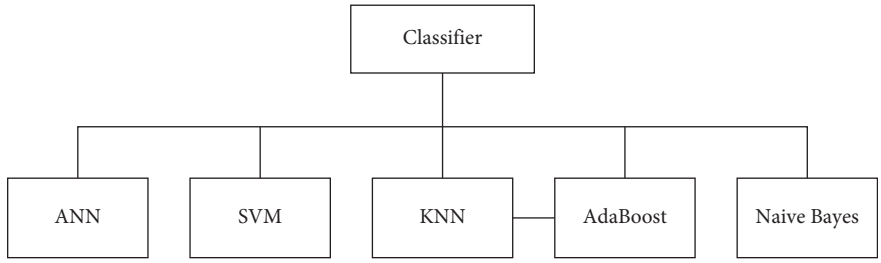

Figure 2: Different supervised classifiers.

(a) (b)

Figure 3: (a) Healthy leaf image. (b) Diseased leaf image.

image to l* a* b* model, and later color segmentation was performed. Later, color and texture features were extracted and fed to the KNN classifier, and the accuracy of the KNN classifier was 96.76%. Allen [32] worked on the Ethiopia maize disease leaf dataset, and the author claims that the research carried out was not proposed by anyone before. In this study, preprocessed RGB is converted to gray, and image enhancement is performed to improve the image quality. Furthermore, texture, color, and morphological features are extracted. They were fed to the classifier, and the accuracy was 95.63%. Basavaiah and Arlene Anthony [33] introduced a model for tomato leaf disease classification using a Random Forest classifier. The dataset consisted of 500 images, and they were resized to 500×500. The features such as color histograms, local binary patterns, and Hu moments are extracted. Furthermore, a dataset of 300 images is used for training, and testing is done for 200 images. The classification was performed using a decision tree classifier and a Random Forest classifier. The experiment resulted in 90% and 94% accuracy for the decision tree and Random Forest classifiers, respectively.

Though extensive research was done on the leaf disease classification, the database used in the maximum papers is leaf images with a plain background. Apart from that, extracting individual leaf from the image is a challenging task.

This paper contributes to the following:

(i) The cotton leaf images were segmented from the natural background

(ii) It can extract the leaf based on the user contour

(iii) Color features are enough to identify healthy and unhealthy leaf images

Table 1 gives a brief overview of the author's contribution to leaf disease classification.

After discussing the author's contributions in the leaf disease classification, we briefly describe the well-known classifiers. The classifiers such as *K*-nearest neighbor (KNN), adaptive boosting (AdaBoost), support vector machine (SVM), Random Forest, Naïve Bayes classifier, and artificial neural network (ANN) are considered for the survey. These classifiers contributed a lot to the image processing field, so we use these classifiers to show the classifier's performance for our database. Furthermore, we give a brief description of the classifiers which we used for the cotton database.

The neural network classifier was introduced by Alexander Bain and William James in 1890. It was inspired by resembling the brain neurons.

Since it resembles the human brain, the algorithms are patterned accordingly. The neural network has the following advantages: it can handle imperfect data and detect all possible interactions between predictor variables. Hence, it is used in regression analysis, classification, and data processing. It has numerous applications in agriculture, and there has been extensive research in this field. Here, we focus on leaf disease classification, wherein we are using the cotton leaf image database. Different types of disease classification are bacterial blight, powdery mildew, etc. The leaf disease classification is carried out using neural network by Kumar et al. [34]. Disadvantages of the method are that it requires greater computational resources and is prone to overfitting problems [35]. However, the performance of the network is good when compared to other classification models. The classification accuracy depends on the features extracted to train the model and also on the dataset. It even relies on the

Table 1: Literature survey review.

Sl	Author and title	Dataset	Preprocessing	Segmentation	Feature extraction	Classification technique	Accuracy (%)
1	Gupta et al., "Artificial plant optimization algorithm to detect infected leaves using machine learning" (2020)	236 dataset images	NA	NA	Histogram of oriented gradients (HOG)	Improved artificial plant optimization	97.45
2	Kumari et al., "Leaf disease detection: feature extraction with *k*-means clustering and classification with ANN."(2019)	Tomato and cotton leaf images	NA	*K*-means clustering	Contrast, correlation, energy, homogeneity, mean, standard deviation, variance	Artificial neural network	92.7
3	Krithika and Grace Selvarani, "An individual grape leaf disease identification using leaf skeletons and KNN classification" (2017)	Grape leaf images	Convert RGB to HSV and l* a* b* color space	Segmentation performed by extracting H and color channels	GLCM features	KNN	80
4	Sarangdhar et al., "Machine learning regression technique for cotton leaf disease detection and controlling using IoT" (2017)	Cotton leaf images	Gabor filter and median filter	Color transformation	Color moment, texture features	SVM	83.26
5	Panigrahi et al. "Maize leaf disease detection and classification using machine learning algorithms" (2020)	Maize leaf images	RGB to grayscale	Labelled edge detection	Shape, color, and texture	Naive Bayes (NB), decision tree (DT), *K*-nearest neighbor (KNN), support vector machine (SVM), and Random Forest (RF)	79.23
6	Vaishnnave et al, "Detection and classification of groundnut leaf diseases using KNN classifier" (2019)	Groundnut leaf images	RGB to HSV	HSV conversion from binary image	Color, texture, morphology	KNN	75
7	Mokhtar et al., "SVM-based detection of tomato leaves Diseases" (2015)	Tomato leaves	Image enhancement-erosion and dilation	Background removal—background subtraction, single leaf extraction manually cropped	GLCM	SVM	99.83
8	Shrivastava and Pradhan, "Rice plant disease classification using color features: a machine learning paradigm" (2021)	Rice plant	Convert RGB to other forms	NA	Color features extracted from each color space	SVM	94.65

Table 1: Continued.

Sl	Author and title	Dataset	Preprocessing	Segmentation	Feature extraction	Classification technique	Accuracy (%)
9	Hossain et al., "A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier" (2019)	Arkansas plant disease database and Reddit-plant leaf disease datasets	RGB to l* a* b* model	Color segmentation	Color and texture (GLCM)	KNN	96.76
10	Alehegn, "Ethiopian maize diseases recognition and classification using support vector machine" (2019)	Ethiopian maize diseases dataset.	RGB to gray scale conversion	*K*-means clustering	Color, texture, and morphological	SVM	95.63
11	Basavaiah and Arlene Anthony, " A tomato leaf disease classification using multiple feature extraction techniques" (2020)	Tomato leaf images	NA	NA	Color histograms, Hu moments, Haralick, local binary pattern features	Random Forest	94

network weights and the number of times the model is trained.

Vapnik introduced support vector machine (SVM) at AT&T Bell Laboratories with colleagues. It is used to categorize unlabelled data. The advantage of using this model is that there is less risk of overfitting. It helps to efficiently classify unlabelled data also. Here, the classifiers use hyperplane, which helps in separating the data points; therefore, this classifier is used in many applications like leaf disease classification. The disadvantage of using this classifier is the time taken to train the model.

In 1951, the K-nearest neighbor algorithm (KNN) classifier was introduced by Evelyn Fix and Joseph Hodges. This classifier is used in regression and classification. In this, k is defined by the user, and it can be any integer. Choosing the value of k differs based on the dataset, and the k value decides the classifier accuracy. This classifier is used in many applications such as text classifier, visual recognition, Wisconsin–Madison breast cancer diagnosis problem, classification of heart disease, and prediction of economic events (text categorization (Guo et al. [36])). Later, hybrid classifiers combined the KNN classifier with other classifiers [37], so that the classification accuracy was improved. It is used in classification and regression. Its application is in leaf disease classification like grapes [38]. Likewise, we have used this classifier for our database also.

The Random Forest classifier algorithm was introduced by Tin Kam ho in 1995 using the random subspace method. This often gives higher accuracy than the single decision tree. This method is used for classification and regression. Random Forest has many advantages such as simple implementation, fast operation, and its application in various fields. It has an effective method for estimating missing data and maintaining accuracy when much data are missing. Hence, the classifier is used in different sectors such as banks and healthcare. Its application in the agricultural sector is leaf classification [39].

One of the disadvantages of this classifier is that it needs more resources and computational power to build many trees to combine the different tree outputs. Since many trees are needed to be united, the time taken to train the classifier will be more.

The AdaBoost classifier presented by Yoav Freund and Robert Schapire in 2003 is a short form of adaptive boosting. It is the first boosting algorithm introduced by Freud and Schapire in 1996. It is a combination of weak classifiers, and during its training, it selects the features that will improve the classifier's predictive power. Since it has many advantages, it was used for classification in multiclass extensions, single-class problems, multilabel problems, etc. In this paper, Subasi et al.[39] proposed an ensemble AdaBoost classifier, which is used to find the human activity using a sensor. Here, the activity recognition is achieved using wearable sensors. The different physical activities were checked by the model proposed by the authors and proved that their model is better when compared to others. It will be used for leaf disease classification [40–42].

Naïve Bayes classifier is based on the Bayes theorem and is widely used in a classification task. The name Naïve is used since it assumes that fed features are considered independent of each other. It means even if you change any one feature, it will not affect the other features. Because of this feature, it is used in many applications.

4. Methodology

In the literature, researchers used machine learning algorithms for classification. This paper presents the introduction to well-known classifiers and presents the classification model as shown in Figure 4.

The first process is to get the leaf image input from the database, and it is preprocessed by resizing the image to 256×256. After reducing the size of the leaf image, extracting the interested region from the natural background is considered segmentation. This step is the crucial step for the next image processing. There are many segmentation algorithms, but we are using the modified factorization-based active contour method (MFACM). The results of the proposed method are good, and it can be observed from the figure. Followed by this, feature extraction needs to be performed. Following is a summary of the feature extraction.

Features play a vital role in extracting meaningful information from an image. From the study, we can say that various feature extraction methods are available. Features such as shape, color, and texture can be extracted so that it helps in further understanding of the image. For example, we can retrieve color features by converting the RGB image to an HSV image, and then using each channel, the color mean is calculated. Color mean, hue, mean saturation, mean value, and mean standard deviation can also be calculated. Apart from color, texture features can also be extracted, which helps in extraction of the leaf pattern. Gray level co-occurrence matrix (GLCM) method is the most commonly used method for texture features. Like this, local binary pattern and wavelet transform features can also be used for feature extraction in the same way. Finally, a summary of the classification algorithms is provided as follows.

4.1. Random Forest Classifier. Random Forest is the simplest and diverse method to solve classification problems. Here, the forest term means ensemble of decision trees and is usually trained using the bagging method as shown in Figure 5. The bagging method combines different learning models to get good accurate results. Based on the maximum voting of each tree class label, the classifier output is decided.

The advantage of this classifier is as follows:

(1) It is easy to measure the relative importance of each feature for prediction

The disadvantage of this classifier is as follows:

(1) Too many decision trees will lead to a slow algorithm

Chaudhary et al. [43] introduced a modified Random Forest classifier for multiclass groundnut leaf disease classification problem. In this paper, a modified Random Forest classifier uses a Random Forest classifier, an attribute evaluator method, and an instance filter method. To show the performance of the proposed method, the author compared existing machine learning algorithms such as SVM, neural network, and logistic regression with the proposed model to check which classifier will be suitable for their dataset. An accuracy of 97.80% is achieved on five UCI machine learning repository benchmark datasets using the projected model.

4.2. Naïve Bayes Classifier. A Naive Bayes classifier [44] is based on the Bayes theorem and is a probabilistic machine learning model used for classification tasks, as shown in Figure 6.

The fundamental Naive Bayes assumption is that each feature makes an independent and equal contribution to the outcome.

The advantages of this classifier are as follows:

(1) It is faster, and it can predict class easily

(2) It solves multiclass prediction problems

The disadvantage of this classifier is as follows:

(1) It is hard to find independent features

Khan et al. [45] familiarized plant accurate recognition and classification using the Naïve Bayes classifier. The features used for classification are texture and shape features. The training of the classifier was performed on 30 different species datasets. The ROC curve is 0.981, which specifies that the accuracy of the classifier is good.

4.3. Feedforward Neural Networks. Feedforward is an artificial neural network [46] and is a biologically inspired algorithm. Here, the information passes in only one direction forward and never comes backward. One of the simplest forms of feedforward network is single-layer perceptron, and another form is multilayer perceptron. The single-layer perceptron has a single layer of output nodes, as shown in Figure 7. Based on the weight, series are fed as input to get the output.

Multilayer perceptron (MLP) [47] consists of multiple layers of computational units or perceptron interconnected to the output layers, as shown in Figure 8. It used the concept of backpropagation learning for training data.

MLP has advantages concerning solving any complex problem with greater efficiency. It has a lot of applications in the field of speech recognition, image recognition, and classification.

The advantages of this classifier are as follows:

(1) It helps in solving the complex problem

(2) Adaptive learning makes the network extract patterns from inaccurate data

The disadvantage of this classifier is as follows:

(1) Sometimes, it might take a longer time for training a large dataset

Since the multilayer perceptron has a lot of advantages, this classifier is used in the field of leaf disease classification. Suhaili Beeran [48] used MLP for watermelon leaf disease classification. The color features are extracted and fed to the classifier. The accuracy of 75.9% is achieved for 200 leaf samples.

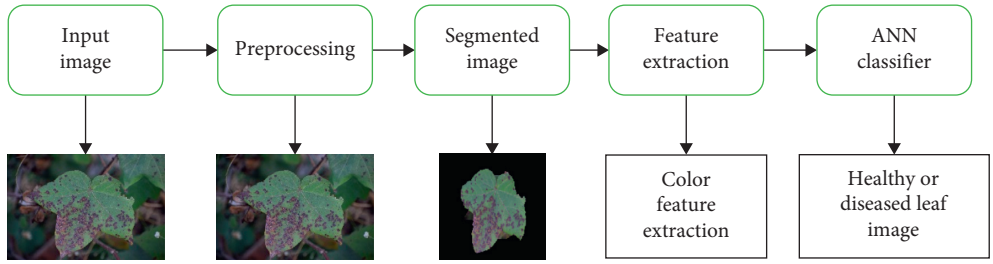

Figure 4: Proposed model.

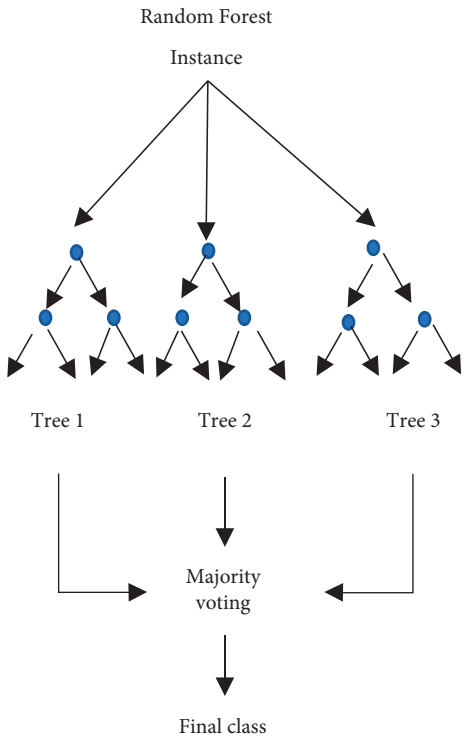

Figure 5: Random Forest.

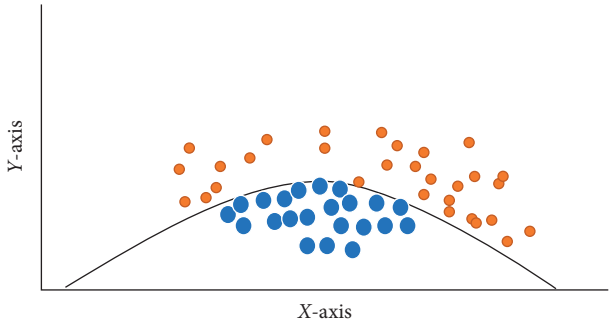

Figure 6: Naïve Bayes.

Shak et al. [49] used MLP for healthy and unhealthy leaf classification. With 90 training samples, the accuracy of the classifier is 97.15%. The accuracy reduces as the number of training samples reduces since the test dataset is more when compared to the training dataset. Next, MLP has marked its place in watermelon leaf disease classification [50].

Though MLP is extensively used in disease classification, the dataset used for classification was simple. The leaf dataset images were with a white or black background, which helps the classifier to outstand as the feature extraction will be easy. In this paper, the cotton dataset is with complex background, and the performance of the classifiers was compared.

4.4. Adaptive Boosting (AdaBoost) Classifier. AdaBoost should meet two conditions:

(1) The classifier should interactively train different weighted training examples

(2) In each iteration, minimizing training errors aims to provide an excellent match for these instances

This method typically selects the subset of training data randomly. Choosing the training set based on the accurate forecast of the last training iteratively trains the AdaBoost machine learning model [51, 52]. It allocates the higher weight to incorrectly categorize observations to have a high likelihood of classification in the next iteration. It also assigns weight to the qualified classifier according to the accuracy of the classifier in each iteration. Elevated weight will be given to the more accurate classifier.

This process iterates until the complete training data suit without any error or until the maximum estimator number specified is reached. To identify, a "vote" across all of the learning algorithms created is performed as shown in Figure 9.

The advantage of this classifier is as follows:

(1) It is less vulnerable to the overfitting problem

The disadvantage of this classifier is as follows:

(1) It is sensitive to noisy data and outliers

4.5. Support Vector Machine (SVM) Classifier. SVM [53, 54] is a supervised machine learning algorithm that was used for classification and regression. It is formally defined by separating the hyperplane, as shown in Figure 10. A hyperplane is a line that helps in separating the data points. The SVM constructs a hyperplane in high-dimensional space or infinite-dimensional space. These hyperplanes help in classifying the data, and there can be more than one hyperplane. The hyperplane, which is at the maximum distance from data points, was considered for classification. The

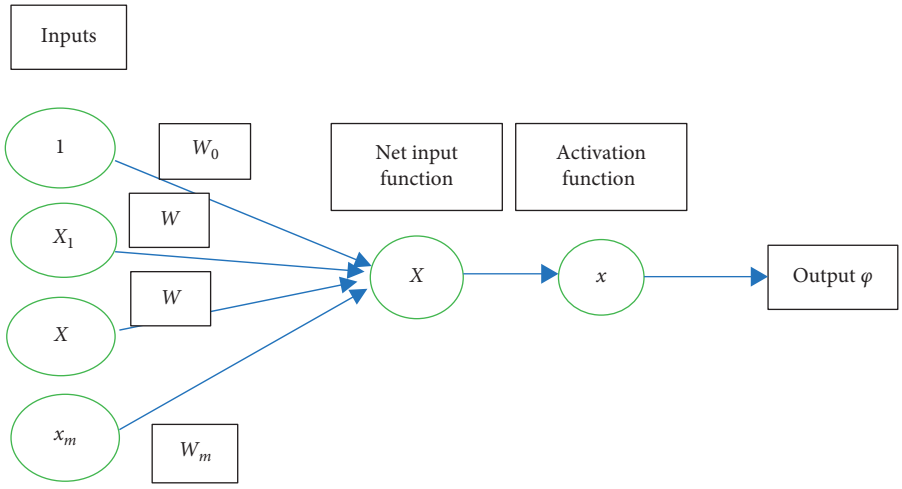

Figure 7: Feedforward network.

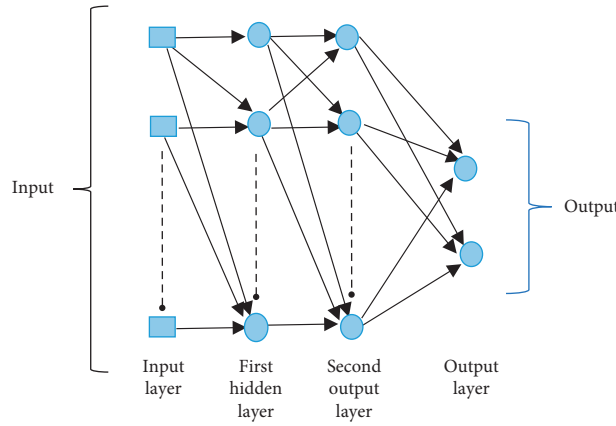

Figure 8: Multilayer perceptron.

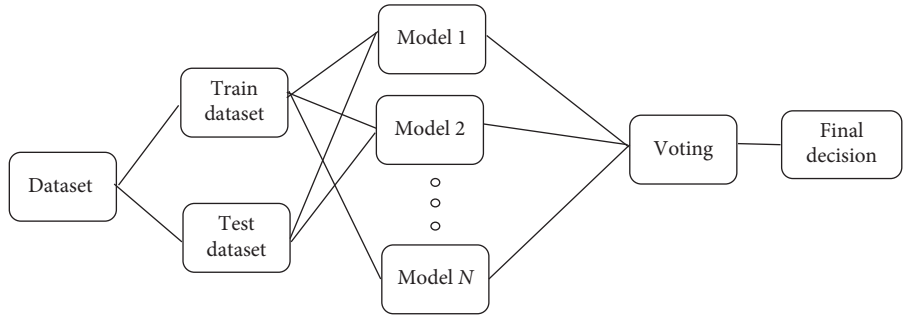

Figure 9: AdaBoost classifier.

classifier is used for high-dimensional spaces. A support vector machine [53, 55] constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space that can be used for classification, regression, or other tasks such as detecting outliers. Automatically, the hyperplane that has the largest distance to the nearest training data point in any class (so-called functional margin) achieves a good separation since, in general, the greater the margin, the lower the classifier's generalization error. SVM has its application in text classification, bioinformatics, hand-written recognition, and image classification.

The advantages of this classifier are as follows:

(1) Classification accuracy is high

(2) It works well for a smaller dataset

The disadvantages of this classifier are as follows:

(1) Training a large dataset will take a longer time

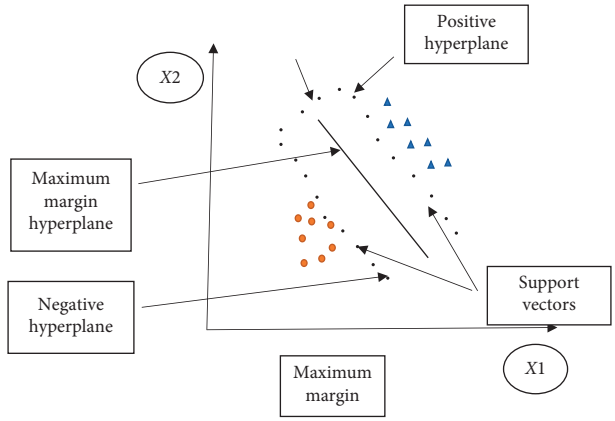

Figure 10: SVM classifier.

(2) Noise sensitivity

Priya and Antony Selvadoss [50] proposed a leaf recognition algorithm using support vector machine (SVM). Here, 12 features were extracted, and the classifier uses the features extracted for classification. This process was carried out on the Flavia dataset and a real dataset. The author compared the SVM classifier with the KNN classifier to show that the SVM has more accuracy and less training time.

4.6. KNN Classifier. It is one of the simplest supervised classification algorithms. The KNN algorithm stores all available data and classifies, based on similarity, a new data point. This implies that it can be conveniently categorized into a well-suite group using the KNN algorithm [56] as new data emerge. It can be used for classification and regression. It is often referred to as a lazy learner algorithm because it does not automatically learn from the training set but instead stores the dataset and performs an operation on the dataset at the time of classification.

At the training point, the KNN algorithm only stores the dataset and then classifies the new data point to a very close group that is nearer to it, as shown in Figure 11.

The KNN working is based on the selection of the K value so that the Euclidean distance can be calculated for k number of neighbors. The categories are done based on the distance between the data points. The query point will belong to the category where there are a maximum number of neighbors.

The advantages of this classifier are as follows:

(1) It is very simple to be implemented

(2) The performance will be good if the training data are large

(3) Does not take training time

Disadvantage of this classifier is as follows:

(1) The computation cost is high

Challenges faced by classifiers are as follows:

(1) Overfitting of the training data because of less datasets

(2) Underfitting of the training data because of removing noise from the data

(3) Time required to train models

The images are segmented from the complex background for cotton leaf disease classification, and removing the background is challenging. The background removal is considered a segmentation technique, and to achieve that, we used a modified factorization-based active contour method. This method helps in recognizing the required leaf image from the image. Later, texture and color features are extracted and fed to the classifier for classification. In the literature, there exist supervised learning classifier algorithms such as artificial neural networks, support vector machine, KNN classifier, AdaBoost, Naïve Bayes classifier, and Random Forest classifier. In this, we are comparing the performance of the classifiers based on the features selected. Features such as color and texture are chosen. The analysis is done on whether texture features or the color features or whether both texture and color features are enough to get the classification accuracy.

In this paper, we focused on the classification of leaf images as healthy and unhealthy. For this binary classification, only color features are enough, and if we further extend it to disease classification, then color features will not be sufficient.

4.7. Weka Tool. Waikato Environment for Knowledge Analysis, developed at the University of Waikato, New Zealand, is free software licensed under the GNU General Public License. It helps analyze machine learning algorithms [57], and the software is written in Java, and it can run on any platform.

5. Results and Discussion

The experiment was carried out on a cotton leaf image dataset. The images were captured in various fields using a digital camera with a resolution of 4048×4048. These images of larger size are difficult to be processed, so they are resized to 256×256. The healthy and diseased classification accuracy using different classifiers is compared based on the number

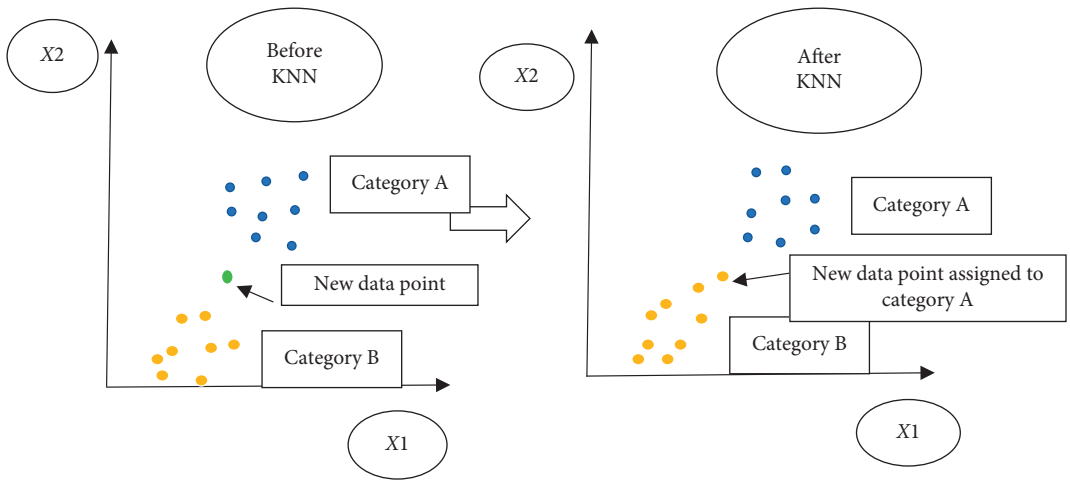

Figure 11: KNN classifier.

Table 2: Evaluation parameters.

Parameter	Equation
Accuracy	TP + TN/(TP + TN + FP + FN)
Precision	TP/TP + FP
Recall	TP/TP + FN
F-measure	(2* precision* recall)/(precision + recall)
TP rate	TP/TP + FN
FP rate	FP/FP + TN
MCC	$\frac{(TP \times TN) - (FP \times FN)}{\sqrt{(FP + TP)(TP + FN)(TN + FP)(TN \times FN)}}$

Table 3: Classification accuracy based on texture and color feature extraction.

Parameters	Random Forest	Bayes	Multilayer perceptron	AdaBoost	SVM	KNN
Accuracy	0.9256	0.8429	0.9669	0.9008	0.9752	0.9173
Precision	0.926	0.840	0.967	0.900	0.976	0.917
Recall	0.926	0.843	0.967	0.901	0.975	0.917
F-measure	0.924	0.840	0.967	0.900	0.975	0.917
MCC	0.825	0.625	0.923	0.767	0.943	0.808

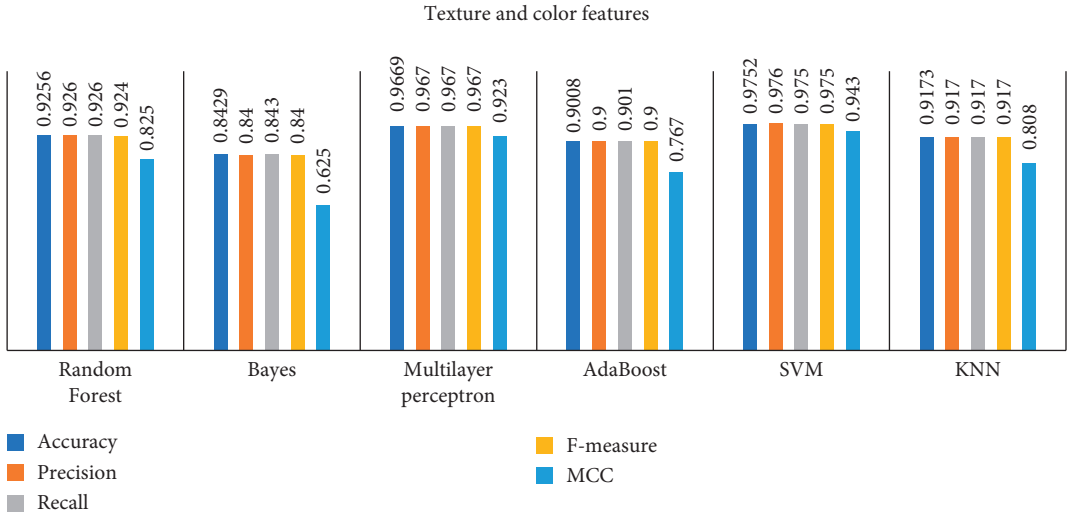

Figure 12: Classifier evaluation measures for texture and color features.

Table 4: Classification accuracy based on texture feature extraction.

Parameters	Random Forest	Bayes	Multilayer perceptron	AdaBoost	SVM	KNN
Accuracy	0.7039	0.6694	0.8925	0.7273	0.6943	0.876
Precision	0.675	0.586	0.781	0.621	0.779	0.724
Recall	0.703	0.644	0.788	0.661	0.778	0.726
F-measure	0.668	0.598	0.781	0.617	0.778	0.726
MCC	0.221	0.035	0.486	0.087	0.554	0.353

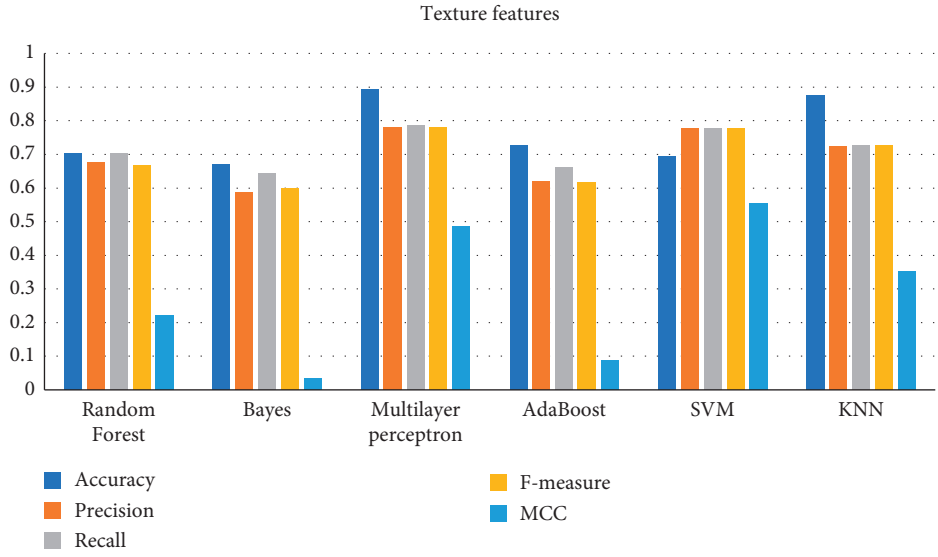

Figure 13: Classifier evaluation based on 8 features.

Table 5: Classification accuracy based on color feature extraction.

Parameters	Random Forest	Bayes	Multilayer perceptron	AdaBoost	SVM	KNN
Accuracy	0.9586	0.9091	0.9669	0.9256	0.9338	0.9421
Precision	0.959	0.911	0.967	0.926	0.934	0.943
Recall	0.959	0.909	0.967	0.926	0.934	0.942
F-measure	0.959	0.906	0.967	0.926	0.834	0.942
MCC	0.904	0.786	0.923	0.829	0.847	0.867

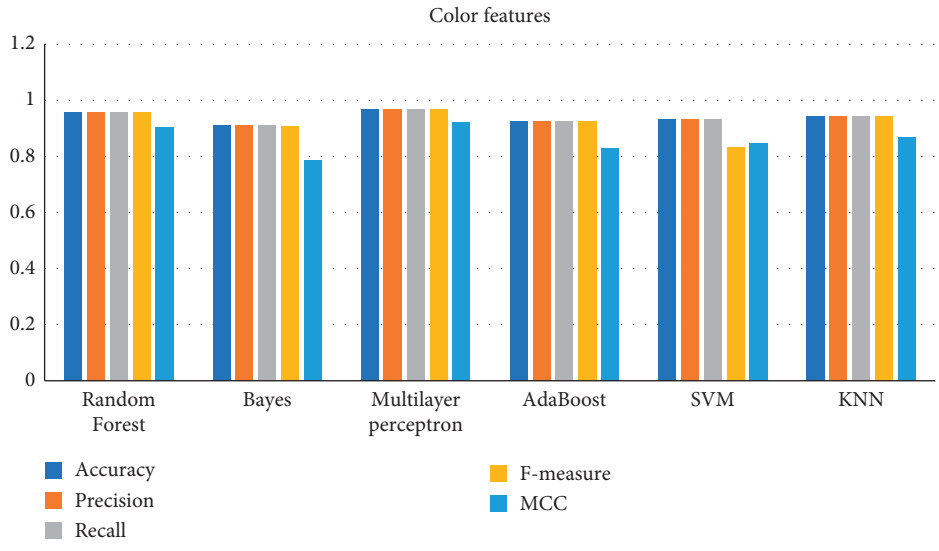

Figure 14: Classifier evaluation based on 4 color features.

Table 6: Classifier performance based on features.

Classifier	12 features	4 features	8 features
Random Forest	92.562	95.86	70.39
Bayes	84.29	90.91	66.94
Multilayer perceptron	96.69	96.69	89.25
AdaBoost	90.086	92.56	72.73
SVM	97.52	93.38	69.42
KNN	91.73	94.21	87.6

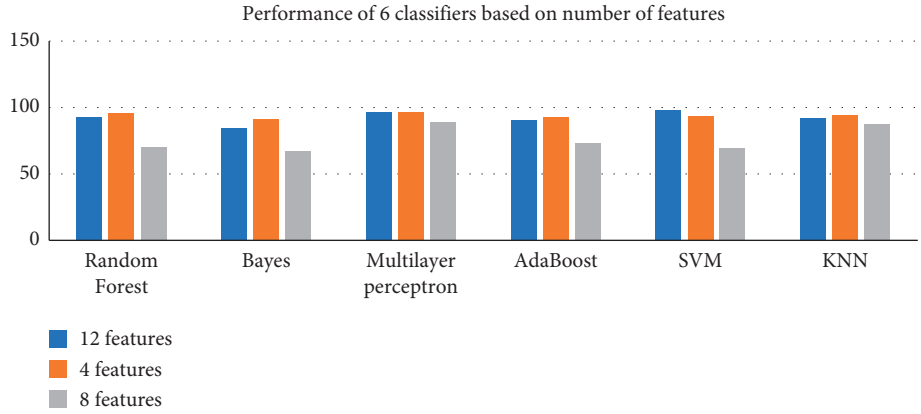

Figure 15: Performance of 6 classifiers.

of features. The feature attributes are color hue, color saturation, color moment, entropy, correlation, energy, contrast, mean, homogeneity, RMS, and standard deviation.

The study of different classifiers is compared based on the number of features extracted. The tool used for the experiment is WEKA (Waikato Environment for Knowledge Analysis, developed at the University of Waikato, New Zealand). The results are compared based on the tool output.

5.1. Evaluation Measures. The classification evaluation measures used for the comparison are accuracy, true-positive (TP) rate, false-positive (FP) rate, precision, recall, *F*-measure, Matthews correlation coefficient (MCC), and class. The parameters are calculated using the following equations in TP, FP, false negative (FP), and true negative (TN) as shown in Table 2).

Table 3 gives the details of different classifier results based on 12 features extracted from nearly 120 images. Here, 12 features were extracted from segmented images: 8 texture features and 4 color features for identifying diseased and nondiseased classification.

Figure 12 shows the classifier accuracy of Random Forest, Bayes, multilayer perceptron, AdaBoost, SVM, and KNN. From the figure, we can observe 12 features affecting the classifier performance. The SVM classifier accuracy performance is more when compared to other classifiers. 5 parameters, namely, accuracy, precision, TP rate, FP rate, and recall values are compared between different classifiers.

In the first case, both texture and color features are extracted and fed to different classifiers. Since all the features may not give the classifiers better performance, we chose 8 texture features to analyze the classifier behaviour. From Table 4, it can be seen that the classifier does not perform so well, and we can also conclude that the multilayer perceptron performed well.

Figure 13 shows the evaluation measures for all the classifiers, which replicate accuracy, precision, recall, *F*-measure, and MCC.

Table 5 and Figure 14 show classification evaluation based on 4 color feature extraction. The multilayer perceptron performs well relative to other classifiers.

The 5 classifiers are compared based on the features extracted. The features, which were used for classification, are texture and color features. The classifier performance is analyzed based on which features are considered. Figure 14 and Table 6 compare different classifiers based on which features are fed as an input to reduce the classifier's training computation time and improve the classification accuracy. The multilayer perceptron performs well when compared to other classifiers for all different types of features.

From Figure 15, we can observe that extracting color features would be enough for classifying a leaf as healthy or diseased. However, the limitation of the proposed method is that if we want to classify diseased types, then color feature alone will not be enough, so we need to extract texture features. Both color and texture features help in leaf disease classification.

6. Conclusion

Leaf disease classification is an essential task in the field of agriculture. The disease identification helps the farmer to find out what precautions can be taken further. The classification can be performed using different machine learning algorithms, and it is used for the cotton leaf database. The

segmentation is performed as the cotton images are taken from the field, and the background is complex. The segmented output images later undergo color and texture feature extraction. This paper shows that the color features are enough to find the classification between healthy and unhealthy images. The same features are used to feed into the WEKA tool, which helps analyze different classifiers. Comparison is performed for 4 color features, 8 texture features, and 12 (texture and color) features. It can be observed that color features are enough for improving the classification accuracy. The survey shows that the accuracy of artificial neural network is better than that of the other classifiers such as Naïve Bayes, Random Forest, SVM, KNN, and AdaBoost. It can be concluded that there is no need to extract different texture descriptors since color features can help identify healthy and unhealthy leaves. In the future, the work can be prolonged to disease classification using texture features or using a deep learning model.

Data Availability

The dataset is available in the public repository: Mendeley Data-Cotton Leaf Dataset.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Bhagya M. Patil and Vishwanath Burkpalli contributed equally to this study.

References

[1] M. G. Sánchez, V. Miramontes-Varo, J. A. Chocoteco, and V. Vidal, "Identification and classification of botrytis disease in pomegranate with machine learning," *Advances in Intelligent Systems and Computing*, vol. 1229, pp. 582–598, 2020.

[2] M. Ji, L. Zhang, and Q. Wu, "Automatic grape leaf diseases identification via united model based on multiple convolutional neural networks," *Information Processing in Agriculture*, vol. 7, no. 3, pp. 418–426, 2020.

[3] H. Sabrol and S. Kumar, "Plant leaf disease detection using adaptive neuro-fuzzy classification," in *Advances in Computer Vision. CVC 2019 Advances in Intelligent Systems and Computing*, K. Arai and S. Kapoor, Eds., vol. 943Cham, Switzerland, Springer, 2020.

[4] A. Gargade and S. Khandekar, "Custard apple leaf parameter analysis, leaf diseases, and nutritional deficiencies detection using machine learning," in *Advances in Signal and Data Processing*Springer, Singapore, 2021.

[5] S. Kumar and K. Vanaja, "Analysis of feature selection algorithms on classification: a survey," *International Journal of Computer Applications*, vol. 96, 2014.

[6] M. Sarkar and T. Y. Leong, "Application of K-nearest neighbors algorithm on breast cancer diagnosis problem," *Proceedings. AMIA Symposium*, vol. 1, pp. 759–763, 2000.

[7] N. Khateeb and M. Usman, "Efficient heart disease prediction system using K-nearest neighbor classification technique"" in *Proceedings of the International Conference on Big Data and Intenet of Thing*, London, UK, December 2017.

[8] S. B. Imandoust and M. Bolandraftar, "Application of k-nearest neighbor (knn) approach for predicing economic events: theoretical background," *International Journal of Engineering Research and Applications*, vol. 3, no. 5, pp. 605–610, 2013.

[9] Z. Chen, L. J. Zhou, X. D. Li, J. N. Zhang, and W. J. Huo, "The Lao text classification method based on KNN," *Procedia Computer Science*, vol. 166, pp. 523–528, 2020.

[10] D.-M. Filimon and A. Albu, "Skin diseases diagnosis using artificial neural networks," in *Proceedings of the 2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 189–194, Timisoara, Romania, May 2014.

[11] J. Patel and R. Goyal, "Applications of artificial neural networks in medical science," *Current Clinical Pharmacology*, vol. 2, no. 3, pp. 217–226, 2007.

[12] H. Thai, "Applying artificial neural networks for Face recognition," *Advances in Artificial NeuralSystems*, vol. 2011, Article ID 673016, 16 pages, 2011.

[13] J. Hatwell, M. M. Gaber, and R. M. Atif Azad, "Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 250, 2020.

[14] A. Minz and C. Mahobiya, "MR image classification using adaboost for brain tumor type," in *Proceedings of the 2017 IEEE 7th International Advance Computing Conference (IACC)*, pp. 701–705, Hyderabad, India, January 2017.

[15] N. H. Nguyen and D. M. Woo, "Terrain classification using adaboost algorithm based on co-occurrence and haar-like features," in *Advanced Multimedia and Ubiquitous Engineering*, J. Park, HC. Chao, H. Arabnia, and N. Yen, Eds., vol. 352, Berlin, Germany, Springer, 2015.

[16] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real time facial expression recognition with adaboost,"vol. 3, pp. 926–929, in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, pp. 926–929, IEEE, Cambridge, UK, August 2004.

[17] A. Takemura, A. Shimizu, and K. Hamamoto, "Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the AdaBoost algorithm with feature selection," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 598–609, 2009.

[18] S. Huang, P. P. Pacheco, S. Narrandes et al., "Applications of support vector machine (SVM) learning in cancer Genomics," *Cancer Genomics & Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[19] A. Murugan, S. A. Nair, and K. P. S. Kumar, "Detection of Skin cancer using SVM, random forest and kNN classifiers," *Journal of Medical Systems*, vol. 43, pp. 1–9, 2019.

[20] R. D. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)," *Asian Pacific Journal of Cancer Prevention*, vol. 20, pp. 1555–1561, 2019.

[21] S. Bakheet, "An SVM framework for malignant melanoma detection based on optimized HOG features," *Computation*, vol. 5, 2017.

[22] T. Deshpande, S. Sengupta, and K. S. Raghuvanshi, "Grading & identification of disease in pomegranate leaf and fruit," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4638–4645, 2014.

[23] D. Gupta, P. Sharma, K. Choudhary et al., "Artificial plant optimization algorithm to detect infected leaves using machine learning," *Expert Systems*, Article ID e12501, 2020.

[24] C. U. Kumari, S. Jeevan Prasad, and G. Mounika, "Leaf disease detection: feature extraction with K-means clustering and

classification with ANN," in *Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1095–1098, Erode, India, April 2019.

[25] N. Krithika and A. Grace Selvarani, "An individual grape leaf disease identification using leaf skeletons and KNN classification," in *Proceedings of the 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, March 2017.

[26] Sarangdhar, A. Adhao, and V. R. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT," in *Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 2, Coimbatore, India, April 2017.

[27] K. P. Panigrahi, H. Das, A. K. Sahoo, and S. C. Moharana, "Maize leaf disease detection and classification using machine learning algorithms," *Advances in Intelligent Systems and Computing*, Springer, Singapore, pp. 659–669, 2020.

[28] M. P. Vaishnnave, K. S. Devi, P. Srinivasan, and G. A. P. Jothi, "Detection and classification of groundnut leaf diseases using KNN classifier," in *Proceedings of the IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1–5, Pondicherry, India, March 2019.

[29] U. Mokhtar, "SVM-based detection of tomato leaves diseases," in *Intelligent Systems 2014 Advances in Intelligent Systems and Computing*, et al. vol. 323, Cham, Switzerland, Springer, 2015.

[30] V. K. Shrivastava and M. K. Pradhan, "Rice plant disease classification using color features: a machine learning paradigm," *Journal of Plant Pathology*, vol. 103, no. 1, pp. 17–26, 2021.

[31] E. Hossain, H. Farhad, and R. Mohammad Anisur, "A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier," in *Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, Cox'sBazar, Bangladesh, February 2019.

[32] E. Alehegn, "Ethiopian maize diseases recognition and classification using support vector machine," *International Journal of Computational Vision and Robotics*, vol. 9, no. 1, pp. 90–109, 2019.

[33] J. Basavaiah and A. Arlene Anthony, "Tomato leaf disease classification using multiple feature extraction techniques," *Wireless Personal Communications*, vol. 115, no. 1, pp. 633–651, 2020.

[34] P. L. Kumar, K. V. K. Goud, G. V. Kumar, and P. S. Kumar, "Enhanced weighted sum back propagation neural network for leaf disease classification," *Materials Today: Proceedings*, 2020.

[35] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996.

[36] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification"," in *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, R. Meersman, Z. Tari, and D. C. Schmidt, Eds., vol. vol 2888Berlin, Germany, Springer, 2003.

[37] R. G. Devi and P. Sumanjani, "Improved classification techniques by combining KNN and random forest with naive bayesian classifier," in *Proceedings of the 2015 IEEE International Conference on Engineering and Technology (ICETECH)*, pp. 1–4, Coimbatore, India, March 2015.

[38] B. Sandika, "Random forest based classification of diseases in grapes from images captured in uncontrolled environments," in *Proceedings of the 2016 IEEE 13th International Conference on Signal Processing (ICSP)*, March 2016.

[39] A. Subasi, D. H. Dammas, R. D. Alghamdi et al., "Sensor based human activity recognition using adaboost ensemble classifier," *Procedia Computer Science*, vol. 140, pp. 104–111, 2018.

[40] M. K. Choudhary and S. Hiranwal, "Feature selection algorithms for plant leaf classification: a survey," in *Proceedings of International Conference on Communication and Computational Technologies*, June 2021.

[41] N. K. Korada, N. Sagar Pavan Kumar, and Y. V. N. H. Deekshitulu, "Implementation of naïve Bayesian classifier and ada-boost algorithm using maize expert system," *International Journal of Information Sciences and Techniques (IJIST)*, vol. 2, 2012.

[42] R. Krishna and K. V. Prema, "Soybean crop disease classification using machine learning techniques," in *Proceedings of the 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, IEEE, Udupi, India, October 2020.

[43] A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Information Processing in Agriculture*, vol. 3, no. 4, pp. 215–222, 2016.

[44] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[45] B. Khan, P. K. Shukla, M. K. Ahirwar, and M. Mishra, "Strategic analysis in prediction of liver disease using different classification algorithms," *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*, vol. 7, pp. 437–449, 2021.

[46] S. B. Maind and P. Wankar, "Research paper on basic of artificial neural network," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, pp. 96–100, 2014.

[47] T. L. Fine, *Feedforward Neural Network Methodology*, Springer Science & Business Media, Berlin, Germany, 2006.

[48] K. Suhaili Beeran, "Classification of watermelon leaf diseases using neural network analysis," in *Proceedings of the 2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC)*, IEEE, Langkawi, Malaysia, April 2013.

[49] S. Shak, "Leaf disease classification using artificial neural network," *Jurnal Teknologi*, vol. 77, 2015.

[50] C. A. Priya and T. Antony Selvadoss, "An efficient leaf recognition algorithm for plant classification using support vector machine," in *Proceeding of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, March 2012.

[51] X. Li, L. Wang, and E. Sung, "AdaBoost with SVM-based component classifiers," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 5, pp. 785–795, 2008.

[52] C. Sandesh Kumar, V. K. Sharma, A. K. Yadav, and A. Singh, "Perception of plant diseases in color images through adaboost," *Advances in Intelligent Systems and Computing*, Springer, Singapore, 2021.

[53] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[54] Y. Zhang, "Support vector machine classification algorithm and its application," in *Information Computing and Applications. ICICA 2012 Communications in Computer and*

Information Science, C. Liu, L. Wang, and A. Yang, Eds., vol. 308Berlin, Germany, Springer, 2012.

[55] D. A. Pisner and D. M. Schnyer, "Chapter 6-Support Vector Machine," in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, Cambridge, MA, USA, 2020.

[56] K. Balakrishna and M. Rao, "Tomato plant leaves disease classification using KNN and PNN," *International Journal of Computer Vision and Image Processing*, vol. 9, no. 1, pp. 51–63, 2019.

[57] P. K. Jain and R. Pamula, "Sentiment analysis in airline data: customer rating based recommendation prediction using WEKA," *Machine Learning Algorithms for Industrial Applications*, Springer, Cham, Switzerland, pp. 53–65, 2021.