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

Bhagya Patil (patil.bhagya@gmail.com)
REVA University  https://orcid.org/0000-0003-2278-9535

Vishwanath Barkpalli
PDA College of Engineering: Poojya Doddappa Appa College of Engineering

Survey paper

Keywords: cotton disease classification, machine learning algorithms, cotton database, color features

DOI: https://doi.org/10.21203/rs.3.rs-502091/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Cotton is one of the major crops in India where 23% of cotton gets exported to other countries. Hence, the cotton yield depends on the crop growth, and it gets affected because of diseases. In this paper, cotton disease classification is performed using different machine learning algorithms. For this research, the cotton database was created by capturing images in the field under controlled conditions. The same database is used for segmenting the images using modified factorization-based active contour. The color and texture features are extracted from segmented images and later its fed to the machine learning algorithms like Multilayer perceptron, Support vector machine, Naïve Bayes, Random forest, Ada Boost, K nearest neighbor. The performance of the classifiers is better when color features are extracted than texture features extraction. The color features are enough to classify the healthy and unhealthy cotton leaf images. Among the different classifiers, Multilayer perceptron gives nearly 96.69% which is greater than other classifiers.

1. Introduction

In India, agriculture is the main occupation and two-third of the population is dependent on agriculture directly or indirectly. Hence yield of the crop depends on the growth and diseases might affect the yield of the crop. So, identifying the plant disease at the early stage will benefit in diagnosing and preventing unnecessary crop loss. Among different parts of the plant, the leaf is the part if gets affected then it affects the crop yield. The detection of disease can be recognized by visible symptoms and plant pathologists can suggest a suitable pesticide. For this, there are numerous image processing methods, and is one of the techniques used for processing the images and identify the disease. There are different leaf disease classification is performed example pomegranate [23], grape [29] (Krithika et al), cotton[5], maize (Panigrahi et al) [28], etc. Figure 1 shows the basic image processing steps for classification.

Normally, the acquired leaf images with the natural background are filtered using the gaussian mask. The filtered image is used for the segment using the Modified factorization based active contour method and followed by extracting features. But if we use all the features in the classification then training time will be more so, selection of features is performed [40]. Finally, the classification of cotton leaf images is performed using different classifiers shown in Fig. 2.

Among the different steps involved in this process, after feature extraction selection of features to improve the performance of the classifier. There are various color and texture features extracted to get accurate classification accuracy. In literature, many authors have submitted surveys on different classifier’s performance. In this paper, based on the features selected [41] the performance of the classifiers is compared. The classifiers like a neural network, support vector machine, Adaboost, naive bayes, random forest are taken into study.

The different classifiers have advantages and disadvantages concerning different parameters like training data size etc. On these classifiers’ authors have researched in different leaf disease classification especially in cotton. These classifiers have application in breast cancer diagnosis problem[2], Heart Disease Prediction System[3], predicting economic events[4], Text Categorization[5], Skin diseases diagnosis[7], medical science[8], Face recognition[9], health science[10], Brain tumor[11], Terrain Classification[12], Real-time facial expression recognition[13], Discrimination of breast tumors in ultrasonic images[15], Cancer Genomics[16], Detection of Skin Cancer[19], Skin Lesion Segmentation[20], Malignant Melanoma Detection[23], disease detection in pomegranate leaf and fruit[25].

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

The organization of this paper is as follows, Sect. 2 is for material and the method, Sect. 3 is related works, Sect. 4 methodology and Sect. 5 results and discussion followed by the conclusion.

2. Material And The Method

Database:
For the study, the cotton leaf images are considered. The images are captured from the cotton field under controlled conditions. Images are captured with a natural background in the field. The database consists of nearly 3000 images. There are 2 categories Healthy and diseased, which were used for training and testing. The sample images are shown below in Fig. 3.

3. Related Works

Many researchers published a paper on leaf disease classification. The classifiers like K-Nearest Neighbour (KNN), Adaptive Boosting (AdaBoost), Support Vector Machine (SVM), Random forest, Bayes classifier, Artificial Neural Network (ANN). These classifiers contributed a lot to the field of image processing, so we use these classifiers to show the classifiers performance for our database. Further we give the brief description about the classifiers which we used for cotton database.

A Neural network 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 advantages like it has ability to handle imperfect data, ability to detect all possible interactions between predictor variables. Hence, it is used in regression analysis, classification, and data processing. It has numerous applications in the field of agriculture and there has been extensive research in this field. Here we focus on leaf disease classification, wherein we are using the cotton leaf images database. Different types of disease classification like bacterial blight, powdery mildew, etc. The leaf disease classification is carried out using a neural network by authors [7]. Disadvantages require greater computational resources, are prone to overfitting [Tu JV. Et al 34]. But still, 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 network weights and no of times the model is trained.

Support Vector Machine (SVM) was introduced by Vapnik 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 in efficiently classify unlabelled data also. Here the classifiers use hyperplane which helps in separating the data points. Since this classifier is used in many applications like leaf disease classification [Patil et al [8], Adhao et al[5], medical field[9], etc. The disadvantage of using this classifier is that training time takes a longer time.

In 1951, K nearest neighbors algorithm(K-NN) classifier is introduced by Evelyn Fix and Joseph Hodges. This classifier 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 k value decides the classifier accuracy. This classifier is used in many applications like text classifier, visual recognition, Wisconsin-Madison breast cancer diagnosis problem, classifying heart disease, predicting economic events (Imandoust et al.[4]), text categorization (Guo G et al.[1]), etc. Later, hybrid classifiers came into existence were combining of KNN classifier with other classifiers (R. G. Devi et al. [6]) so that classification accuracy was improved. It is used in classification and regression. Its application is in leaf disease classification like grapes [38], maize [35], groundnut [32], etc. Likewise, we have used this classifier for our database also

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 lot of applications like implementation is not complex, fast in operation, and it has its role in various fields. It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing. Hence, the classifier is used in different sectors like Bank, Healthcare, etc. Its application is in the agriculture sector that is leaf disease classification [39]. One of the disadvantages of this classifier is it needs more resources and computational power to build many trees so that it can combine the different trees output. Since, many trees need to combined and for this the time taken to train the classifier will be more.

Adaboost classifier presented by Yoav Freund and Robert Schapire in 2003 and is a short form of Adaptive boosting. It is the first boosting algorithm introduced by Freund and Schapire in 1996. It’s a combination of weak classifiers and during its training, it selects the features which will improve the predictive power of the classifier. Having so many advantages it was used for
classification in multi-class extensions, single class problems, multi label problems, etc. It will be used for leaf disease classification [korada et al [36], Krishna et al [37]].

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

Table 1 gives a brief overview of the authors contribution to leaf disease classification.
| Sl | Author and title | Dataset | Preprocessing | Segmentation | Feature extraction | Classification technique | Accuracy |
|----|-----------------|---------|---------------|--------------|-------------------|----------------------------|---------|
| 1  | Krishna, Rajashree, and K. V. Prema. "Soybean crop disease classification using machine learning techniques." IEEE, 2020 | Soybean | NA | NA | NA | Bagging classifier | 88.4% |
| 2  | Kumari, Ch Usha, S. Jeevan Prasad, and G. Mounika. "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, N., and A. Grace Selvarani. "An individual grape leaf disease identification using leaf skeletons and KNN classification." 2017 | Grape leaf images | Convert RGB to HSV and la*b* color space. | Segmentation performed by extracting H and a color channels | GLCM feaures | KNN | 80% |
| 4  | Sarangdhar, Adhao Asmita, and V. R. Pawar. "Machine learning regression technique for cotton leaf disease detection and controlling using IoT." 2017 | Maize leaf images | Gabor filter and median filter | Color tranformation | Color moment, Texture features | SVM | 83.26 |
| 5  | Panigrahi, Kshyanapra Panda, 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  | M. P. 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, morpholoogy | KNN | 75% |
| SI | Author and title | Dataset | Preprocessing | Segmentation | Feature extraction | Classification technique | Accuracy |
|----|------------------|---------|---------------|--------------|-------------------|--------------------------|----------|
| 7  | Mokhtar U. et al., SVM-Based Detection of Tomato Leaves Diseases. 2015 (eds) Springer | Tomato leaves | Image enhancement – erosion and dilation | Background removal-background subtraction, single leaf extraction manually cropped | GLCM | SVM | 99.83% |
| 8  | Shrivastava et al. 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% |
| 9  | Hossain, Eftekhar 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 | Aravind K et al. Grape Crop Disease Classification Using Transfer Learning Approach.2019 | Grape crop | NA | NA | NA | Convolution Neural Network(CNN) | 99.23% |
| 11 | C. A. Priya et al, "An efficient leaf recognition algorithm for plant classification using support vector machine," 2012 | Flavia dataset, Real dataset | RGB to gray scale conversion | NA | 12 leaf features, PCA | SVM | 94.5% - Flavia Dataset 96.8% |
| 12 | Alehegn, Enquhone. "Ethiopian maize diseases recognition and classification using support vector machine." (2019) | Ethiopian maize diseases dataset. | RGB to gray scale conversion | K means clustering | Colour, texture and morphological | SVM | 95.63% |
| 13 | Basavaiah et al, 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% |

4. Methodology

4.1 Random Forest classifier:

Random Forest is the simplest and diverse method to solve classification problems. Here the forest term is meant ensemble of decision trees and usually trained using the bagging method as shown in Fig. 4. The Bagging method is combining different
learning models to get good accuracy results. Based on the each tree class labels maximum voting the classifier output is decided.

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

Disadvantages: 1) Too many decision trees will lead to the slow algorithm

Basavaiah et al [53] introduced a model for tomato leaf disease classification by means of random forest classifier. The dataset consists of 500 images and resizing of size 500x500. The features colour histograms, local binary patterns, Hu moments are extracted. Further, dataset of size 300 images is used for training, and testing is done for 200 images. The classification is performed using a decision tree classifier and random forest classifier. The experiment resulted in 90% and 94% accuracy for decision tree and random forest classifiers respectively.

Chaudhary et al [54], introduced a modified random forest classifier for multi-class groundnut leaf disease classification problems. 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 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 Naïve Bayes classifier [58] is based on Bayes theorem and is a probabilistic machine learning model that’s used for classification tasks as shown in Fig. 5.

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

Advantages:

1. It is faster and it can predict class easily
2. It solves multi-class prediction problems

Disadvantages:

1. It’s hard to find independent features

Padao et al [55], familiarized plant accurate recognition and classification using Naïve Bayes classifier. The features used for classification are texture and shape features are extracted. Training of classifier is performed on 30 different species datasets. The ROC curve is 0.981 which specifies the accuracy of the classifier is good.

4.3 Feedforward neural networks

Feedforward is a form of an artificial neural network [57] and it is inspired by a biologically inspired algorithm. Here the information passes only one direction forward and never comes backward. One of the simplest form feedforward networks is single layer perceptron and another form is multilayer perceptron. The single layer perceptron has a single layer of output nodes as shown in Fig. 6. Based on the weight series are fed as input to the output.

Multilayer perceptron (MLP) [47] consists of multiple layers of computational units or perceptron which are interconnected to the output layers as shown in the Fig. 7. It used the concept of backpropagation learning for training data.

MLP has advantages concerning for solving any complex problem with greater efficiency. It has a lot of applications in the field of speech recognition, image recognition and classification [48].

Advantages:
1. It helps in solving the complex problem
2. Adaptive learning makes the network extract the patterns from imprecise data.

Disadvantages:

1. Sometimes it might take a longer time for training a large dataset

Since multilayer perceptron has a lot of advantages which led to the usage of this classifier in the field of leaf disease classification. Shak et al [48] used MLP for healthy and unhealthy leaf classification. With a training sample of 90, the accuracy of the classifier is 97.15%. The accuracy reduces as the training sample reduces since the test data set is more when compared to the train dataset. Next, MLP has marked its place in watermelon leaf disease classification [49]. Author Kutty et al [49], used MLP for watermelon leaf disease classification. The color features are extracted and feed to the classifier. The accuracy of 75.9% is achieved for 200 leaf samples.

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

4.4 Adaptive Boosting (Ada Boost) classifier-

Ada Boost[14, 59] was proposed by Yoav Freund and Robert Schapire in 1996 and it is an iterative collective method as shown in Fig. 8. It helps in a combination of multiple poor performing classifiers so that classifier accuracy will be more. The basic idea behind Adaboost is to set the weights of classifiers and training the data sample in each iteration so that it ensures the correct predictions of unusual observations.

Two conditions should be met by Adaboost:

1. Different weighted training examples should be interactively trained by the classifier.
2. In each iteration, by minimizing training errors, it aims to provide an excellent match for these instances.

This method normally selects randomly the subset of training data. By choosing the training set based on the accurate forecast of the last training, it iteratively trains the AdaBoost machine learning model. It allocates the higher weight to incorrectly categorized observations so that these observations will 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 suits without any error or until the maximum estimator number specified is reached. To identify, perform a “vote” across all of the learning algorithms you created.

Advantages:

1. It is less vulnerable to the overfitting problem

Disadvantages:

1. It is sensitive to noisy data and outliers

In the paper, author Subasi et al [35] proposed ensemble Adaboost classifier 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 authors proposed model and proved that their model is better when compared to others.

4.5 Support vector machine (SVM) classifier-
SVM [60] is a supervised machine learning algorithm that can be used for classification as well as for regression. It is formally defined by separating the hyperplane as shown in Fig. 9. A hyperplane is the line that helps in separating the data points. The SVM constructs 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 maximum distance from data points will be considered for classification. The classifier is used for high dimensional spaces. A support vector machine [17,18] constructs in a high- or infinite-dimensional space a hyperplane or set of hyperplanes 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, image classification.

Advantages:

1. Classification accuracy is high
2. Works well for a smaller dataset

Disadvantages:

1. Training a large dataset will take a longer time
2. Noise sensitivity

Priya et al [51], 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 flavia dataset and a real dataset. The author compared SVM classifier with the KNN classifier to show that the SVM has more accuracy and takes less training time.

Alehegn et al [52], worked on the Ethiopia maize disease leaf dataset and the author claims that the research carried out is not proposed by anyone. In this, pre-processing RGB to gray conversion, image enhancement is performed to improve the image quality. Further, texture, color, and morphological features are extracted. They are fed to the classifier and the accuracy is 95.63%.

### 4.6 K-NN classifier:

It is one of the simplest supervised classification algorithms. The K-NN [61] 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 K-NN algorithm [1] as new data emerges. 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 data into a group that is very close to the new data when it receives new data as shown in Fig. 10.

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

Advantages:

1. It is very simple to implement.
2. The performance will be good if the training data is large.
3. No Training time
Disadvantages: The computation cost is high.

Hossain et al [45], proposed the leaf disease classification using the KNN classifier. In this paper, the Arkansas plant disease database and Reddit-plant leaf disease datasets are used for their research. The input image RGB to l*a*b* model so that color segmentation is performed. A segmented image is used further to get the color features to be extracted. The features are fed into the KNN classifier and an accuracy is 76.63%.

Krithika, N et [29], presented individual grape leaf disease identification using a KNN classifier. Author proposed tangential direction image segmentation. The color and GLCM features are extracted and further fed to the KNN classifier to get greater accuracy.

For cotton leaf disease classification, the images are segmented from the complex background, and removing the background is a challenging task. The background removal is considered as 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 literature, there are supervised learning classifier algorithms like Artificial neural network, Support vector machine, K-NN classifier, AdaBoost, Naïve bayes classifier, Random forest classifier, etc. In this, we are comparing the performance of the classifiers based on the features selected. Features like color features or texture features are selected. 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, work is focused on 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 won't 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 in analyzing machine learning algorithms [56, 57] and 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 images dataset. The images were captured in various fields using a digital camera resolution of 4048x4048. For processing, these images of larger size are difficult to process so resizing the images to 256x256. The healthy and diseased classification accuracy, using different classifiers is compared based on the number of features. The feature attributes are color Hue, color saturation, color Hue, Color moment, Entropy, Correlation, energy, contrast, mean, homogeneity, RMS, 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.

Evaluation measures:

The classification evaluation measures used for the comparison are Accuracy, TP rate, FP rate, Precision, Recall, F-measure, class.

Table 2 gives the details of different classifier results based on 12 features extracted from nearly 120 images. In this, 12 features are extracted from segmented images. 8 texture features and 4 color features for identifying diseased and non-diseased classification.
Table 2
Classification accuracy based on Texture and color feature extraction

| Parameters  | Random Forest | Bayes | Multilayer perceptron | Ada Boost | SVM   | K-NN  |
|-------------|---------------|-------|-----------------------|-----------|-------|-------|
| Accuracy    | 92.56         | 84.29 | 96.69                 | 90.08     | 97.52 | 91.73 |
| 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 11 shows the classifier accuracy of Random forest, Bayes, Multilayer perceptron, Ada Boost, SVM, and K-NN. From the figure, we can observe 12 features affecting the classifier performance. The performance of SVM classifier accuracy is more when compared to other classifiers. 5 parameters, namely Accuracy, Precision, TP Rate, FP Rate, Recall values comparison between different classifiers.

Table 3
Classification accuracy based on Texture feature extraction

| Parameters  | Random Forest | Bayes | Multilayer perceptron | Ada Boost | SVM   | K-NN  |
|-------------|---------------|-------|-----------------------|-----------|-------|-------|
| Accuracy    | 70.39         | 66.94 | 89.25                 | 72.73     | 69.43 | 87.6  |
| 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 |

In the first case, both texture and color features are extracted and fed to different classifiers. Since all the features may not give better performance of the classifiers, so we chose 8 texture features to analyze the classifier behaviour. From Table 3, it can be seen that behaviour of the classifier doesn’t perform so well and we can also conclude that Multilayer perceptron performed well.

Figure 12 shows the evaluation measures for all the classifiers, which replicates the values of Accuracy, Precision, Recall, F-measure, MCC.

The Table 4 and Fig. 13 gives the 4 color, features consideration for classifier performance. The multilayer perceptron performs well relative to other classifiers.

Table 4
Classification accuracy based on color feature extraction

| Parameters  | Random Forest | Bayes | Multilayer perceptron | Ada Boost | SVM   | K-NN  |
|-------------|---------------|-------|-----------------------|-----------|-------|-------|
| Accuracy    | 95.86         | 90.91 | 96.69                 | 92.56     | 93.38 | 94.21 |
| 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 |

The 5 classifiers are compared based on the features extracted. The features which are used for classification are texture and color features. The classifier performance is analyzed based on which features are considered. Figure 14 and Table 5, gives the comparison of different classifiers based on which features are fed as an input to reduce the classifier's training computation.
time and to improve the classification accuracy. The multilayer perceptron performs well when compared to other classifiers for all different types of 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      |
| Ada Boost           | 90.086      | 92.56      | 72.73      |
| SVM                 | 97.52       | 93.38      | 69.42      |
| K-NN                | 91.73       | 94.21      | 87.6       |

Conclusions

Leaf disease classification is an important 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 are later features like colors and texture features are extracted. In this paper, we are showing that the color features are enough to extract to find the classification between healthy and unhealthy images. The same features are used to feed into the WEKA tool which helps in the analysis of 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 is enough for improving the classification accuracy. From the survey, it can be seen that artificial neural network accuracy is better than the other classifiers like Naive Bayes, random forest, SVM, K-NN, AdaBoost. In the future, the work can be prolonged to disease classification.

Declarations

Ethics approval and consent to participate- Informed consent was obtained from all individual participants included in the study.

Consent for publication- NA

Availability of data and materials- It’s not available publicly.

Competing interests- There was no conflict of interest.

Funding- No funds, grants, or other support was received.

Authors’ contributions- All authors contributed in this research

Acknowledgements- NA

References
1. Guo G, Wang H, Bell D, Bi Y, Greer K. “KNN Model-Based Approach in Classification”. In: Meersman R, Tari Z, Schmidt DC, editors. On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Vol. 2888. Berlin: Springer; 2003. Lecture Notes in Computer Science.

2. Sarkar M, Leong TY. "Application of K-nearest neighbors algorithm on breast cancer diagnosis problem." Proceedings. AMIA Symposium (2000): 759 – 63.

3. Nida Khateeb M, Usman, "Efficient Heart Disease Prediction System using K-Nearest Neighbor Classification Technique", Proceedings of the international conference on Big Data and Internet of Thing, 2017.

4. Imandoust SB, Bolandraftar M. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. Int J Eng Res Appl. 2013;3(5):605–10.

5. Sarangdhar A, Asmita, Pawar VR. "Machine learning regression technique for cotton leaf disease detection and controlling using IoT." 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA). Vol. 2. IEEE, 2017.

6. Devi RG, Sumanjani P, "Improved classification techniques by combining KNN and Random Forest with Naive Bayesian classifier," 2015 IEEE International Conference on Engineering and ATechnology (ICETECH), Coimbatore, 2015, pp. 1–4, doi: 10.1109/ICETECH.2015.7274997.

7. Filimon D, Albu A, "Skin diseases diagnosis using artificial neural networks," 2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, 2014, pp. 189–194, doi: 10.1109/SACI.2014.6840059.

8. Patel JL, Goyal RK. Applications of artificial neural networks in medical science. Curr Clin Pharmacol. 2007 Sep;2(3):217 – 26. doi: 10.2174/157488407781668811. PMID: 18690868.

9. Thai H Le, "Applying Artificial Neural Networks for Face Recognition", Advances in Artificial NeuralSystems, vol. 2011, Article ID 673016, 16 pages, 2011.

10. Hatwell J, Gaber MM, Atif Azad RM. Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences. BMC Med Inform Decis Mak. 2020;20:250.

11. Minz A, Mahobiya C. "MR Image Classification Using Adaboost for Brain Tumor Type," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, 2017, pp. 701–705, doi: 10.1109/IACC.2017.0146.

12. Nguyen NH, Woo DM. Terrain Classification Using Adaboost Algorithm Based on Co-occurrence and Haar-like Features. In: Park J, Chao HC, Arabnia H, Yen N, editors. Advanced Multimedia and Ubiquitous Engineering. Vol. 352. Berlin: Springer; 2015. https://doi.org/10.1007/978-3-662-47487-7_45. Lecture Notes in Electrical Engineering.

13. Wang Y, Ai H, Wu B, Huang C. (2004, August). Real time facial expression recognition with adaboost. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 3, pp. 926–929). IEEE.

14. Li X, Wang L, Sung E. AdaBoost with SVM-based component classifiers. Eng Appl Artif Intell. 2008;21(5):785–95.

15. Takemura A, Shimizu A, Hamamoto K. Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the AdaBoost algorithm with feature selection. IEEE Trans Med Imaging. 2009;29(3):598–609.

16. Huang S, et al. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer genomics proteomics vol. 2018;15(1):41–51. doi:10.21873/cgp.20063.

17. Zhang Y. Support Vector Machine Classification Algorithm and Its Application. In: Liu C, Wang L, Yang A, editors. Information Computing and Applications. ICICA 2012. Communications in Computer and Information Science. Vol. 308. Berlin: Springer; 2012.

18. Cervantes J, Garcia-Lamont F, Rodriguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020;408:189–215.

19. Murugan A, Nair SA, Kumar KPS. Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers. J Med Syst. 2019;43:1–9.

20. Seeja RD, Suresh A. Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM). Asian Pac J Cancer Prev. 2019;20:1555–61.
21. Pisner DA, Schnyer DM. Chapter 6-Support vector machine. In: Machine Learning; Mechelli, Vieira A S., editors.; Academic Press: Cambridge, 2020; pp. 101–21.
22. Bakheet S. An SVM Framework for Malignant Melanoma Detection Based on Optimized HOG Features. Computation 2017.
23. Shi Z, He L, Suzuki K, Nakamura T, Itoh H. Survey on Neural Networks Used for Medical Image Processing. Int J Comput Sci. 2009;3:86–100.
24. Narayanan U, Unnikrishnan A, Paul V, Joseph S, "A survey on various supervised classification algorithms," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 2118–2124, doi: 10.1109/ICECDS.2017.8389824.
25. Deshpande T, Sengupta S, Raghuvarshi KS. "Grading & identification of disease in pomegranate leaf and fruit.". International Journal of Computer Science Information Technologies. 2014;5(3):4638–45.
26. Sethy PK, Barpanda NK, Rath AK, et al. Deep feature based rice leaf disease identification using support vector machine. Comput Electron Agric. 2020;175:105527.
27. Kumari CU, Jeevan Prasad S, Mounika G, "Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1095–1098, doi: 10.1109/ICCMC.2019.8819750.
28. Panigrahi K, Panda. "Maize leaf disease detection and classification using machine learning algorithms. In: " Progress in Computing, Analytics and Networking. Singapore: Springer; 2020. pp. 659–69.
29. Krithika N, Grace Selvarani A. "An individual grape leaf disease identification using leaf skeletons and KNN classification." 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2017.
30. Kumari Ch, Usha S, Jeevan Prasad, Mounika G. "Leaf disease detection: Feature extraction with k-means clustering and classification with ann." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.
31. Prajapati BS, Vipul K, Dabhi, Harshadkumar B Prajapati. "A survey on detection and classification of cotton leaf diseases." 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, 2016.
32. Vaishnavve MP, Devi KS, Srinivasan P, Jothi GAP, "Detection and Classification of Groundnut Leaf Diseases using KNN classifier," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1–5.
33. Chethan KS, et al. "Mobile Application for Classification of Plant Leaf Diseases Using Image Processing and Neural Networks. In: " Data Intelligence and Cognitive Informatics. Singapore: Springer; 2021. pp. 287–306.
34. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol. 1996 Nov;49(11):1225-31. doi: 10.1016/s0895-4356(96)00002-9. PMID: 8892489.
35. Subasi A, et al "Sensor based human activity recognition using adaboost ensemble classifier." Procedia Computer Science 140 (2018): 104–111. Panigrahi, Kshyanaprava Panda, et al. "Maize leaf disease detection and classification using machine learning algorithms." Progress in Computing, Analytics and Networking. Springer, Singapore, 2020. 659–669.
36. Korada N, Kumar N, Sagar Pavan, Kumar, Deekshitulu YVN.H. "Implementation of naive Bayesian classifier and ada-boost algorithm using maize expert system." International Journal of Information Sciences and Techniques (IJIST) 2 (2012).
37. Krishna R, Prema KV. "Soybean crop disease classification using machine learning techniques." 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, 2020.
38. Sandika B, et al. "Random forest based classification of diseases in grapes from images captured in uncontrolled environments." 2016 IEEE 13th International Conference on Signal Processing (ICSP). IEEE, 2016.
39. Choudhary M, Kumar, and Saroj Hiranwal. "Feature Selection Algorithms for Plant Leaf Classification: A Survey." Proceedings of International Conference on Communication and Computational Technologies. Springer, Singapore, 2021.
40. Kumar S, Vanaja K, Ramesh. "Analysis of feature selection algorithms on classification: a survey." (2014).
41. Mokhtar U, et al. SVM-Based Detection of Tomato Leaves Diseases. In: Filev D et al, editors. Intelligent Systems'2014. Advances in Intelligent Systems and Computing. Vol. 323. Cham: Springer; 2015. https://doi.org/10.1007/978-3-319-11310-4_55.

42. Shrivastava VK, Pradhan MK. Rice plant disease classification using color features: a machine learning paradigm. J Plant Pathol. 2021;103:17–26.

43. Agarwal M, Sinha A, Gupta SK, Mishra D, Mishra R. Potato Crop Disease Classification Using Convolutional Neural Network. In: Somani A, Shekhawat R, Mundra A, Srivastava S, Verma V, editors. Smart Systems and IoT: Innovations in Computing. Vol. 141. Springer, Singapore: Smart Innovation, Systems and Technologies; 2020.

44. Hossain E, Hossain MF, Mohammad Anisur Rahaman. "A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier." 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2019.

45. Aravind K, Anirudh RP, Mukesh R, Ashiwin K, Vikas R G. (2019) Grape Crop Disease Classification Using Transfer Learning Approach. In: Pandian D., Fernando X., Baig Z., Shi F, editors Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB). ISMAC 2018.

46. Fine TL. Feedforward neural network methodology. Springer Science & Business Media; 2006.

47. Shak S, "Leaf disease classification using artificial neural network." Jurnal Teknologi (2015).

48. Kutty S, Beeran, et al. "Classification of watermelon leaf diseases using neural network analysis." 2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC). IEEE; 2013.

49. Priya C, Arun T, Balasaravanan. and Antony Selvadoss Thanamani. "An efficient leaf recognition algorithm for plant classification using support vector machine." International conference on pattern recognition, informatics and medical engineering (PRIME-2012). IEEE, 2012.

50. Priya CA, Balasaravanan T, Thanamani AS, "An efficient leaf recognition algorithm for plant classification using support vector machine," International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), Salem, India, 2012, pp. 428–432.

51. Alehegn E. Ethiopian maize diseases recognition and classification using support vector machine. International Journal of Computational Vision Robotics. 2019;9(1):90–109.

52. Basavaiah J, Arlene Anthony A. Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques. Wireless Pers Commun. 2020;115:633–51.

53. Chaudhary A, Kolhe S. Raj Kamal. "An improved random forest classifier for multi-class classification." Information Processing in Agriculture 3.4 (2016): 215–22.

54. Padao FReyF, Elmer A, Maravillas. "Using Naive Bayesian method for plant leaf classification based on shape and texture features." 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2015.

55. Khan B, et al. "Strategic analysis in prediction of liver disease using different classification algorithms." Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning. IGI Global, 2021. 437–449.

56. Jain P, Kumar, Pumula R. "Sentiment Analysis in Airline Data: Customer Rating Based Recommendation Prediction Using WEKA." Machine Learning Algorithms for Industrial Applications. Cham: Springer; 2021. pp. 53–65.

57. Maind SB. Priyanka Wankar. "Research paper on basic of artificial neural network." International Journal on Recent Innovation Trends in Computing Communication 2.1 (2014): 96–100.

58. Pal M. "Random forest classifier for remote sensing classification. International journal of remote sensing. 2005;26(1):217–22.

59. Kumar C, Sandesh, et al. "Perception of Plant Diseases in Color Images Through Adaboost." Innovations in Computational Intelligence and Computer Vision. Singapore: Springer; 2021. pp. 506–11.

60. Cervantes J, et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020;408:189–215.
Figures

Figure 1
Leaf disease classification

Figure 2
Different supervised classifiers

Figure 3
a) Healthy leaf image b) Diseased leaf image
Figure 4

Random forest

Figure 5

Naïve bayes
Schematic of Rosenblatt’s perceptron.

Figure 6

Feedforward network

Figure 7

Multilayer Perceptron
Figure 8
Adaboost classifier

Figure 9
SVM classifier

Figure 10
KNN classifier
Figure 11

Classifier evaluation measures for texture and color features

TEXTURE FEATURES

![Graph showing evaluation measures for texture features]
Figure 12

classifiers evaluation based on 8 features

![Graph showing classifiers evaluation based on 8 features](image)

Figure 13

classifiers evaluation based on 4 features

![Graph showing classifiers evaluation based on 4 features](image)

Figure 14

Six classifiers performance

![Graph showing performance of 6 classifiers based on number of features](image)