Plant Disease Detection Using Machine Learning

Shima Ramesh
Assistant Professor: department of electronics and communication,
MVJ college of Engineering.
Bangalore, India

Mr. Ramachandra Hebbar,
Senior Scientist, ISRO, RRSC-S,
Marathalli, Bangalore, India,
hebbar4@gmail.com

Niveditha M, Pooja R, Prasad Bhat N, Shashank N
Research Scholar: department of electronics and communication,
MVJ college of Engineering,
Bangalore, India

Mr. P V Vinod
Scientist, ISRO, RRSC-S,
Marathalli, Bangalore, India
ramasubramoniams@gmail.com

***Abstract*— Crop diseases are a noteworthy risk to sustenance security, however their quick distinguishing proof stays troublesome in numerous parts of the world because of the non attendance of the important foundation. Emergence of accurate techniques in the field of leaf-based image classification has shown impressive results. This paper makes use of Random Forest in identifying between healthy and diseased leaf from the data sets created. Our proposed paper includes various phases of implementation namely dataset creation, feature extraction, training the classifier and classification. The created datasets of diseased and healthy leaves are collectively trained under Random Forest to classify the diseased and healthy images. For extracting features of an image we use Histogram of an Oriented Gradient (HOG). Overall, using machine learning to train the large data sets available publicly gives us a clear way to detect the disease present in plants in a colossal scale.**

Keywords— ***Diseased and Healthy leaf, Random forest, Feature extraction, Training, Classification.***

I. INTRODUCTION

The agriculturist in provincial regions may think that it's hard to differentiate the malady which may be available in their harvests. It's not moderate for them to go to agribusiness office and discover what the infection may be. Our principle objective is to distinguish the illness introduce in a plant by watching its morphology by picture handling and machine learning.

Pests and Diseases results in the destruction of crops or part of the plant resulting in decreased food production leading to food insecurity. Also, knowledge about the pest management or control and diseases are less in various less developed countries. Toxic pathogens, poor disease control, drastic climate changes are one of the key factors which arises in dwindled food production.

Various modern technologies have emerged to minimize postharvest processing, to fortify agricultural sustainability and to maximize the productivity. Various Laboratory based approaches such as polymerase chain reaction, gas chromatography, mass spectrometry, thermography and hyper spectral techniques have been employed for disease identification. However, these techniques are not cost effective and are high time consuming.

In recent times, server based and mobile based approach for disease identification has been employed for disease identification. Several factors of these technologies being high resolution camera, high performance processing and extensive built in accessories are the added advantages resulting in automatic disease recognition.

Modern approaches such as machine learning and deep learning algorithm has been employed to increase the recognition rate and the accuracy of the results. Various researches have taken place under the field of machine learning for plant disease detection and diagnosis, such traditional machine learning approach being random forest, artificial neural network, support vector machine(SVM), fuzzy logic, K-means method, Convolutional neural networks etc.…

Random forests are as a whole, learning method for classification, regression and other tasks that operate by constructing a forest of the decision trees during the training time. Unlike decision trees, Random forets overcome the disadvantage of over fitting of their training data set and it handles both numeric and categorical data.

The histogram of oriented gradients (HOG) is an element descriptor utilized as a part of PC vision and image processing for the sake of object detection. Here we are making utilization of three component descriptors:

1. Hu moments
2. Haralick texture
3. Color Histogram

Hu moments is basically used to extract the shape of the leaves. Haralick texture is used to get the texture of the leaves and color Histogram is used to represent the distribution of the colors in an image.

978-1-5386-7523-6/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDI3C.2018.00017

II. LITERATURE REVIEW

[1] S. S. Sannakki and V. S. Rajpurohit, proposed a "Classification of Pomegranate Diseases Based on Back Propagation Neural Network" which mainly works on the method of Segment the defected area and color and texture are used as the features. Here they used neural network classifier for the classification. The main advantage is it Converts to L*a*b to extract chromaticity layers of the image and Categorisation is found to be 97.30% accurate. The main disadvantage is that it is used only for the limited crops.

[2] P. R. Rothe and R. V. Kshirsagar introduced a" Cotton Leaf Disease Identification using Pattern Recognition Techniques" which Uses snake segmentation, here Hu's moments are used as distinctive attribute. Active contour model used to limit the vitality inside the infection spot, BPNN classifier tackles the numerous class problems. The average classification is found to be 85.52%.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic". K-means clustering used to segment the defected area; GLCM is used for the extraction of texture features, Fuzzy logic is used for disease grading. They used artificial neural network (ANN) as a classifier which mainly helps to check the severity of the diseased leaf.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa, proposed" Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease "Color histograms are extracted and transformed from RGB to HSV, RGB to L*a*b.Peak components are used to create max tree, five shape attributes are used and area under the curve analysis is used for classification. They used nearest neighbors, Decision tree, random forest, extremely randomized tree, Naïve bayes and SV classifier. In seven classifiers extremely, randomized trees yield a very high score, provide real time information provide flexibility to the application.

[5] uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Color features are represented in RGB to HIS, by using GLCM, seven invariant moment are taken as shape parameter. They used SVM classifier which has MCS, used for detecting disease in wheat plant offline.

III. PROPOSED METHODOLOGY

To find out whether the leaf is diseased or healthy, certain steps must be followed. i.e., Preprocessing, Feature extraction, Training of classifier and Classification. Preprocessing of image, is bringing all the images size to a reduced uniform size. Then comes extracting features of a preprocessed image which is done with the help of HOG . HoG [6] is a feature descriptor used for object detection. In this feature descriptor the appearance of the object and the outline of the image is described by its intensity gradients. One of the advantage of HoG feature extraction is that it operates on the cells created. Any transformations doesn't affect this.

Here we made use of three feature descriptors.

Hu moments: Image moments which have the important characteristics of the image pixels helps in describing the objects. Here Hu moments help in describing the outline of a particular leaf. Hu moments are calculated over single channel only. The first step involves converting RGB to Gray scale and then the Hu moments are calculated. This step gives an array of shape descriptors.

Haralick Texture: Usually the healthy leaves and diseased leaves have different textures. Here we use Haralick texture feature to distinguish between the textures of healthy and diseased leaf. It is based on the adjacency matrix which stores the position of (I,J). Texture [7] is calculated based on the frequency of the pixel I occupying the position next to pixel J. To calculate Haralick texture it is required that the image be converted to gray scale.

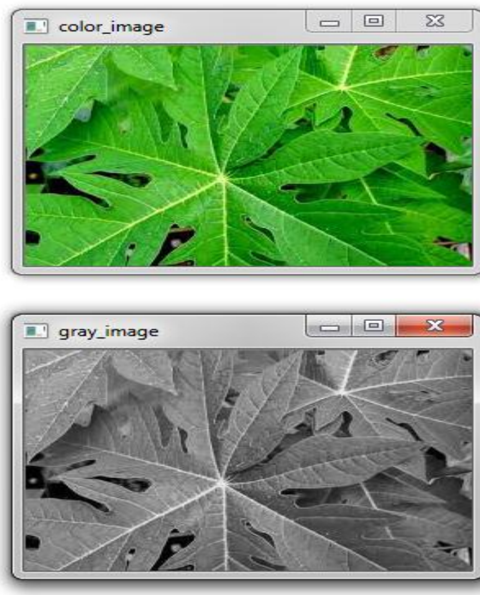

Fig.1. RGB to Gray scale conversion of a leaf.

Color Histogram: Color histogram gives the representation of the colors in the image. RGB is first converted to HSV color space and the histogram is calculated for the same. It is needed to convert the RGB image to HSV since HSV model aligns closely with how human eye discerns the colors in an image. Histogram plot [8] provides the description about the number of pixels available in the given color ranges

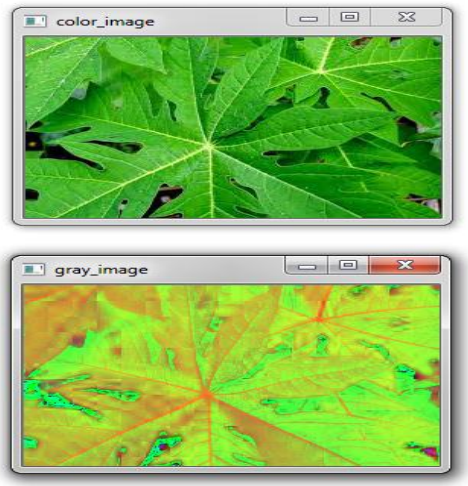

Fig.2. RGB to HSV conversion of leaf

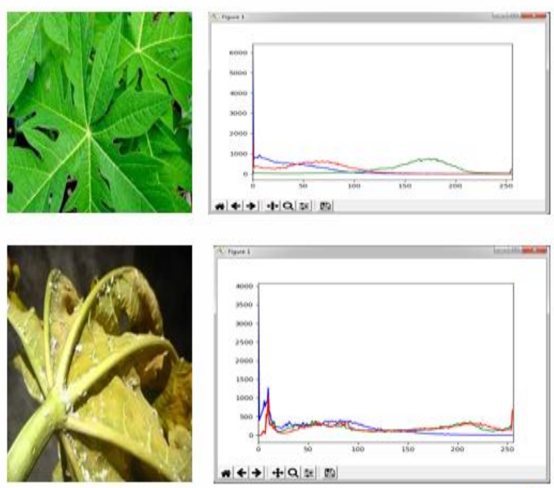

Fig.3. Histogram plot for healthy and diseased leaf.

IV. ALGORITHM DESCRIPTION

The algorithm here is implemented using random forests classifier. They are flexible in nature and can be used for both classification and regression techniques. Compared to other machine learning techniques like SVM, Gaussian Naïve bayes, logistic regression, linear discriminant analysis, Random forests gave more accuracy with less number of image data set. The following figure shows the architecture of our proposed algorithm.

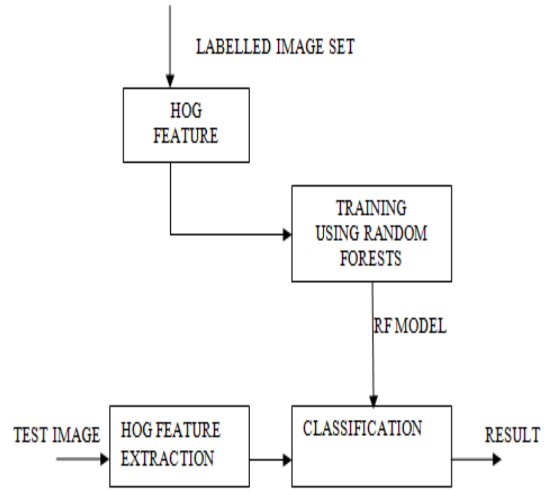

Fig.4. Architecture of the proposed model

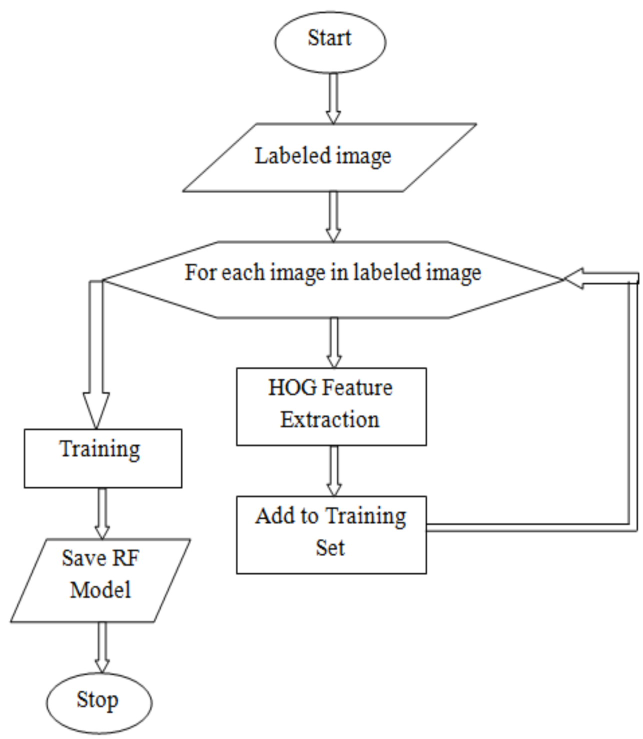

Fig.5. Flow chart for training.

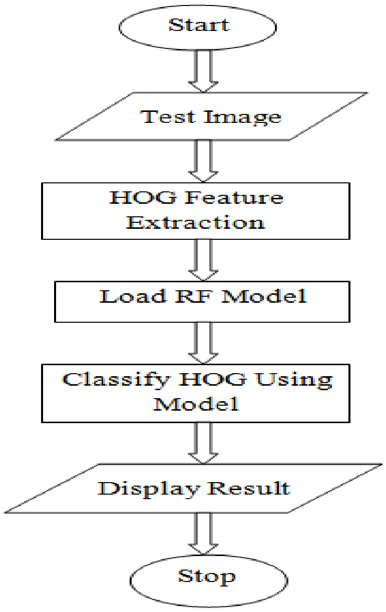

Fig.6. Flow chart for classification

The labeled datasets are segregated into training and testing data. The feature vector is generated for the training dataset using HoG feature extraction. The generated feature vector is trained under a Random forest classifier. Further the feature vector for the testing data generated through HoG feature extraction is given to the trained classifier for prediction as referred to in "Fig.4".

As shown in the 'Fig.5." labeled training datasets are converted into their respective feature vectors by HoG feature extraction. These extracted feature vectors are saved under the training datasets. Further the trained feature vectors are trained under Random forest classifier [9, 10].

As depicted in "Fig.6." the feature vectors are extracted for the test image using HoG feature extraction. These generated feature vectors are given to the saved and trained classifier for predicting the results.

V. RESULT

First for any image we need to convert RGB image into gray scale image. This is done just because Hu moments shape descriptor and Haralick features can be calculated over single channel only. Therefore, it is necessary to convert RGB to gray scale before computing Hu moments and Haralick features. As depicted in the figure 4.

To calculate histogram the image first must be converted to HSV (hue, saturation and value), so we are converting RGB image to an HSV image as shown the figure5.

Finally, the main aim of our project is to detect whether it is diseased or healthy leaf with the help of a Random forest classifier which is as depicted in the "Fig.7."

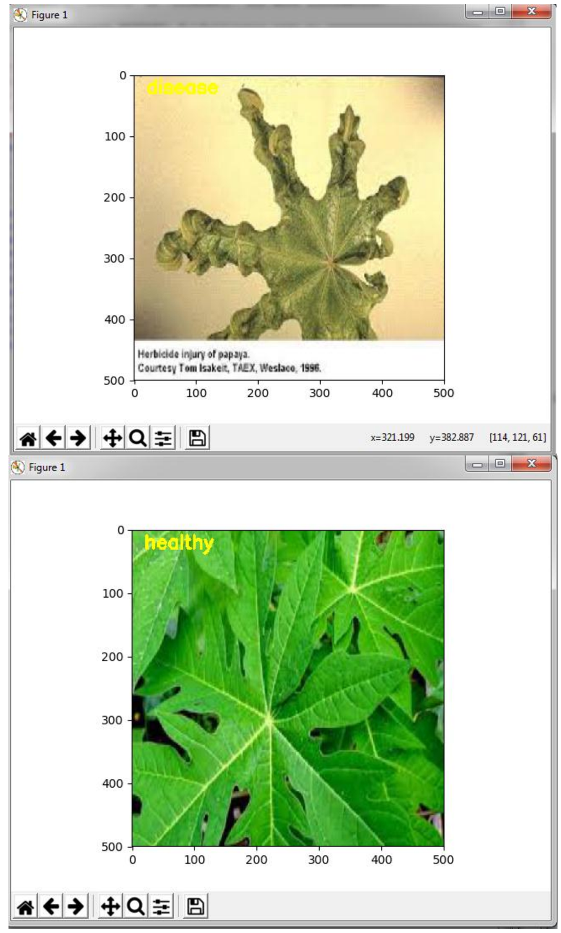

Fig.7. Final output of the classifier.

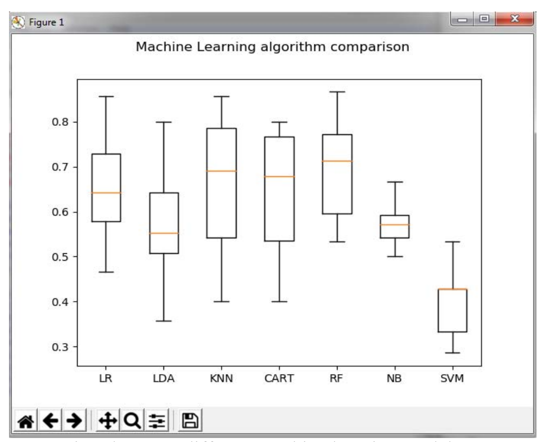

Fig.8. Comparison between different machine learning models.

TABLE I.

Various Machine learning models	Accuracy(percent)
Logistic regression	65.33
Support vector machine	40.33
k- nearest neighbor	66.76
CART	64.66
Random Forests	70.14
Naïve Bayes	57.61

\

Fig .9. Table showing the comparison.

conclusion

The objective of this algorithm is to recognize abnormalities that occur on plants in their greenhouses or natural environment. The image captured is usually taken with a plain background to eliminate occlusion. The algorithm was contrasted with other machine learning models for accuracy. Using Random forest classifier, the model was trained using 160 images of papaya leaves. The model could classify with approximate 70 percent accuracy. The accuracy can be increased when trained with vast number of images and by using other local features together with the global features such as SIFT (Scale Invariant Feature Transform), SURF (Speed Up Robust Features) and DENSE along with BOVW (Bag Of Visual Word)

The graph and table below gives the comparison of machine learning algorithms.

REFERENCES

[1] S. S. Sannakki and V. S. Rajpurohit," Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering and Technology (IRJET), Vol2 Issue: 02 | May-2015

[2] P. R. Rothe and R. V. Kshirsagar," Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC),2015.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease ", Preceding of the 1'st international conference on the use of mobile ICT in Africa ,2014.

[5] uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Proceedings of 2010 Conference on Dependable Computing (CDC'2010), November 20-22, 2010.

[6] S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, "Pnn based crop disease recognition with leaf image features and meteorological data," International Journal of Agricultural and Biological Engineering, vol. 8, no. 4, p. 60, 2015.

[7] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," Springer Plus, vol. 2, no.660, pp. 1–12, 2013.

[8] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). "A plant recognition approach using shape and color features in leaf images." In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.

[9] Zhen, X., Wang, Z., Islam, A., Chan, I., Li, S., 2014d. "Direct estimation of cardiac bi-ventricular volumes with regression forests." In: Accepted by Medical Image Com- puting and Computer-Assisted Intervention–MICCAI 2014.

[10] Wang P., Chen K., Yao L., Hu B., Wu X., Zhang J., et al. (2016)." Multimodal classification of mild cognitive impairment based on partial least squares".