JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Plant Disease Detection Using Machine Learning

[1]Shiv Narain Gupta, [2]Sourabh Bailwal, [3]Rohit Rai

[1]Assistant Professor, [2]Student, [3]Student,
[1]Department of Electronic and Communication Engineering,
[1]Greater Noida Institute of Technology (GNIOT), Greater Noida, Uttar Pradesh, India

Abstract: Leaf disease pose a significant threat to food security; however their quick identification remains a problem in different region across the world due to the absence of necessary foundations. Advancements in the field of image-based techniques of plant leaf classification showed great results. In this research paper, we will use Random Forest technique to identify whether the leaf is healthy or it is infected by any disease based on results of datasets which is created. This research paper consists of several stages for the implementation such as creating datasets, Extraction of different features of leaf, classifier then training and the last step is classification. The datasets of the leaves which is infected by any disease and the healthy leaves are combined and is trained under Random Forest for the classification of the infected and the healthy leaves. A histogram of oriented gradient is used as a tool for extracting features from an image. Overall, training large data sets with machine learning enables accurate disease detection of plant leaf on large scale.

***Index Terms* - Random Forest Technique, Extraction of Features, Training of Model, Classification.**

I. INTRODUCTION

In provincial regions, it is very difficult for an agriculturist to determine what kind of disease is present in their harvests. It is not easy for them to get in touch with an agribusiness office and find out what the disease is. In this study we are primarily concerned with determining whether plants exhibit any symptoms of illness by observing their morphology using picture handling and machine learning. In less developed countries people have limited knowledge on how to control any disease that is occurring on leaf and the pest management, which results in declination of their production that cause food insecurity. Pests and diseases also damage the crops and the different part of the plants that result in reduction of food. One of the key reasons for decreased food production is toxic infectious agents, poor control of disease, and extreme changes in climate. Harmful pathogens such as bacteria viruses, lack of control on disease and the drastic change in climate are the main reasons which causes decline in the production of food.

To minimise post-harvest losses, to enhance sustainability, and to increase the productivity, different types of modern technologies have been used. For the diagnosis of any diseases, different laboratory-based approaches like gas chromatography, polymerase chain reactions, and mass spectrometry, thermography, and hyperspectral techniques have been used [1-5]. However these methods are quite time-consuming and not cost-effective. Mobile-based and internet-based approaches for the recognition of any disease are currently in use. There are various factors of these technologies, including High Resolution camera's extensive built-in accessories and, high performance processing that result in automatic recognition of disease. Various approaches such as deep learning and machine learning algorithm have been used to increase the accuracy and the recognition rate of the results. Various researches have been conducted in the field of machine learning for plant disease detection and diagnosis. Artificial Neural Network, Random forest, Fuzzy logic, Support Vector Machine, Convolutional Neural Network, K-means method are some of the traditional machine learning approaches.

In general Random Forest is a method widely used for classification and regression problem and the other task that operates by constructing a decision tree at the time of training. The decision tree has the disadvantage like over fitting of trained datasets. Random Forests have the advantages of handling both categorical and numerical data. The Histogram of Oriented Gradient [HOG] is a feature descriptor used in computer vision and image processing. The main purpose of using HOG is to extract the different features of leaf. That's why using three feature extractor techniques.

- Hu Moments Feature Extractor
- Haralick Texture Feature
- Color Histogram Feature Extractor

Hu Moments is a feature Extractor which is generally used for extracting the outline of the leaf. For extracting the texture of leaf Haralick texture feature is used and for extracting the distribution of different colours in image, Color Histogram is used.

II. METHODOLOGY

For determining whether the leaf is infected or healthy, we have to follow several steps. This includes Pre processing, Extracting features, Training and Classification. The main function of pre-processing of any image is to improve the image data that suppresses undesired distortion and then extracting all the features of the pre-processed image by using Histogram of Oriented

Gradient (HOG). An object is detected by using HOG [6] feature descriptor, the outline of an image as well as the appearance of the object, is described by its intensity gradient. This technique counts occurrences of gradient orientation in localised portion of an image that's the major advantage of HOG feature extraction. In this paper we are using three feature extractors that are:

A. Hu Moment

Hu Moments are used to describe, characterize and quantify the shape of an object. Hu Moments are generally extracted from silhouette or the outline of leaf. In step first the RGB image is converted into Gray scale so that Hu Moment can be calculated as illustrated in Fig. 1.

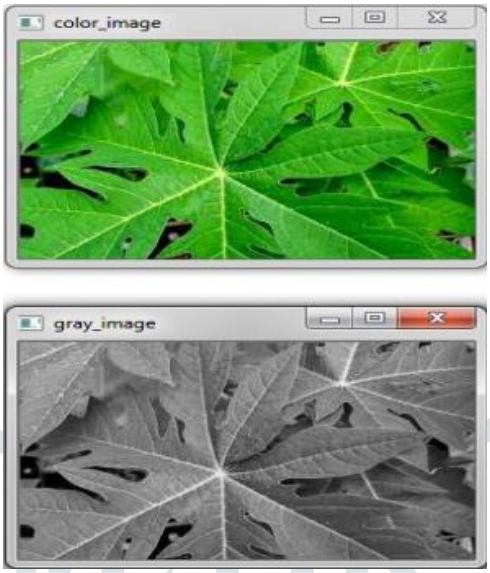

Figure 1 Converting RGB image to grayscale

B. Haralick Texture Feature

Generally texture of diseased leaf and healthy leaf are not similar to each other. So we are using Haralick Texture Feature to to identify infected leaf and the healthy leaf. Then we store the position of (i,j) on the basis of adjacency matrix. According to the frequency of pixels in the image, the texture [7] is calculated by placing pixel "i" next to pixel "j". Therfore it is mandatory to convert RGB image into grayscale for calculating Haralick texture.

C. Color Histogram

A color histogram represents the colour in an image. In first step RGB images are converted to HSV color spaces (shown in Fig. 3) and then the histogram is calculated. It is essential that RGB images are converted to HSV color spaces. Histogram plot diagram [8] provides us the description of the number of pixel available in the given color range as shown in Fig. 2 and Fig. 3.

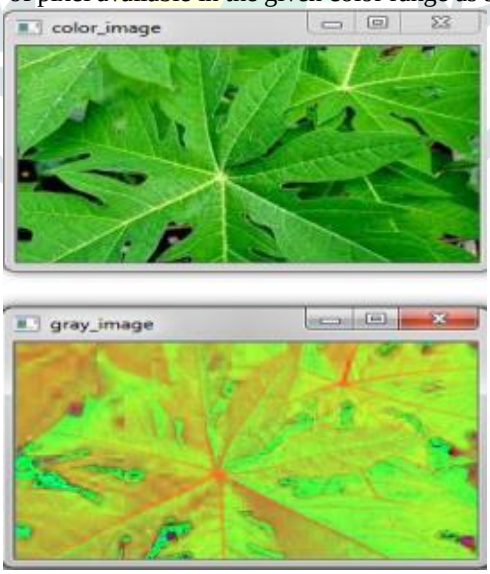

Figure 2 Converting RGB image to HSV

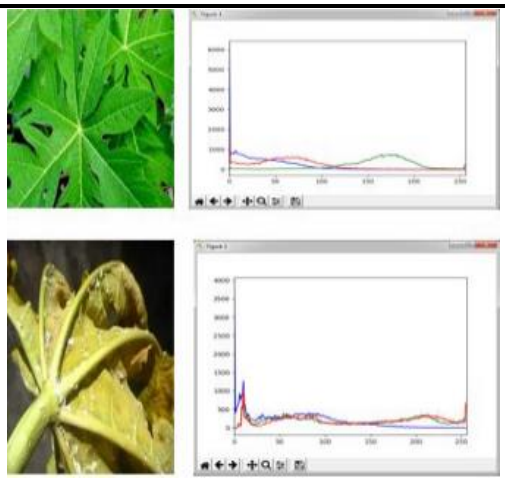

Figure 3 Infected and healthy leaf histogram plot diagram

III. Proposed Algorithm

The algorithm that is implemented by using random forests classifier and can be used for both classification and regression problems. On comparing it with various techniques like support vector machine, logistic regression and the accuracy of random forests is more with less datasets. The architecture and the flow charts of proposed algorithm are illustrated in Fig. 4 and Fig. 5, respectively. Training and testing datasets are separated from labeled datasets. The HOG feature extraction is used to generate a feature vectors for the training data. By using Random Forest Classifier this feature vector is trained and from HOG feature extraction the testing data obtained from HOG feature extraction the feature vector of testing data is obtained. It is then passed to the classifier for predicting as shown in Fig. 4. By using HOG feature extraction the labelled trained datasets are converted into feature vector as shown in Fig. 5.

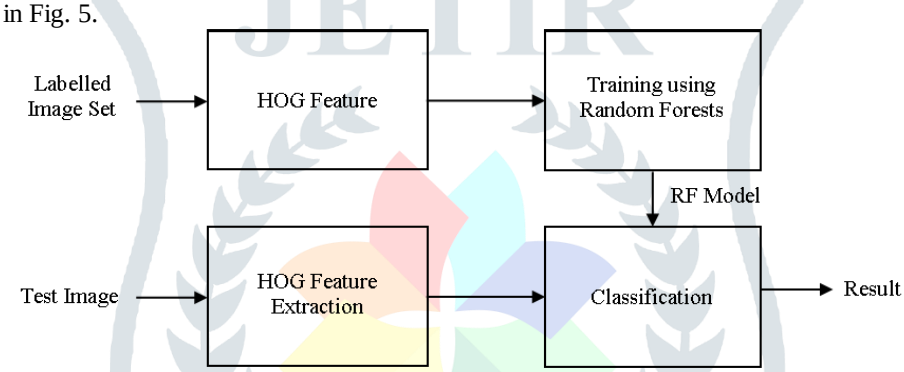

Figure 4 Proposed architecture model

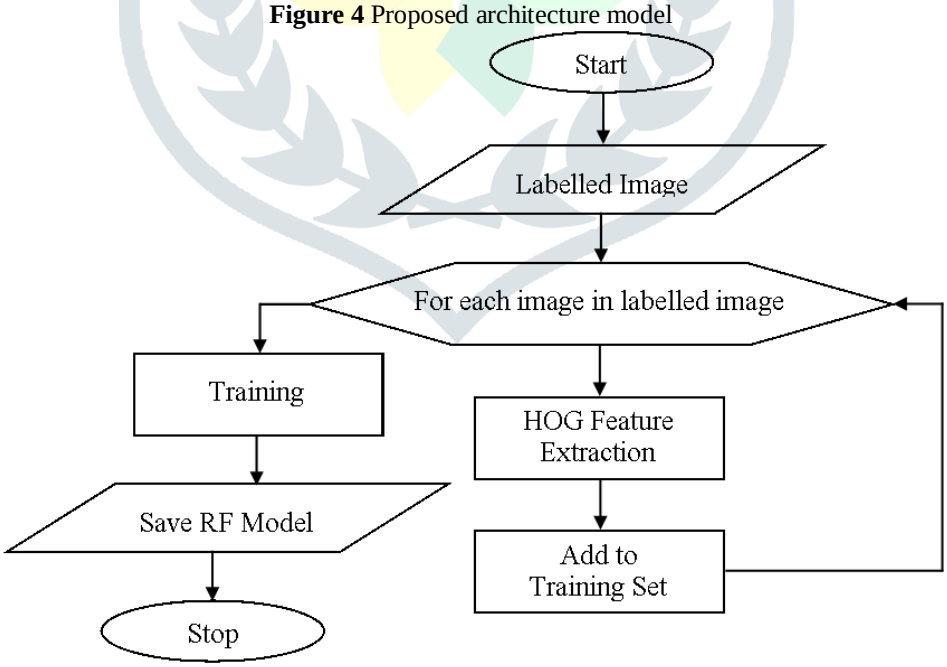

Figure 5 Training Model flow chart

Once the feature vector is extracted, it is then saved under training datasets which is trained under Random Forest Classifier [9, 10] and by using of HOG feature extraction the feature vectors are extracted for the test image as illustrated in Fig. 6. A saved and trained classifier is used to predict the result based on the generated feature vector.

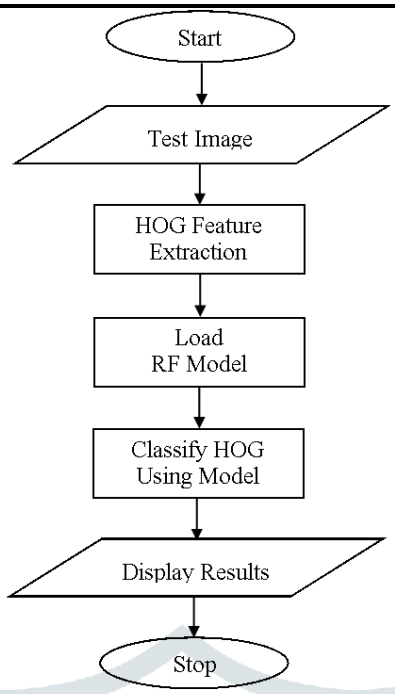

Figure 6 Classification Model flow chart

IV. RESULTS AND DISCUSSION

For the detection of disease in any image, first of all RGB images are converting to Grayscale images. This is done because Haralick texture feature and Hu Moments can only be calculated over a single channel. In the next step histograms are calculated by converting RGB images into HSV [Hue, Saturation, Value], so first we convert RGB images to HSV as shown in Fig. 5. The main aim of this project is to detect whether the leaf is healthy or a diseased by using random forest classifier as shown in Fig. 7.

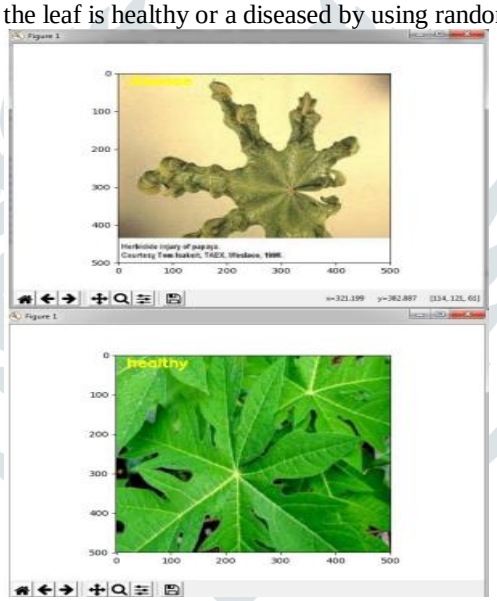

Figure 7 Classifier output

Table 1 Accuracy with different machine learning algorithms

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

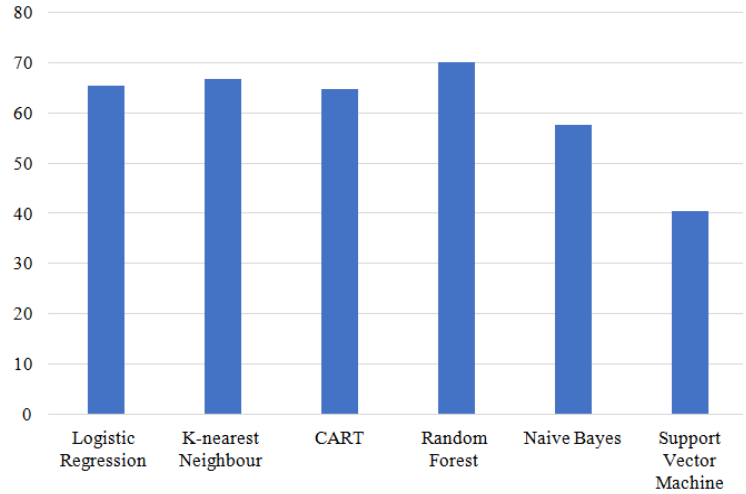

Figure 8 Accuracy Percentage of different Machine learning models

V. Conclusion

This algorithm is designed to identify abnormalities occurring on plants in their natural habitat. The images are taken with plain background to remove occlusion. We compared the results of the algorithms of different machine learning models. With Random Forest classifier we trained 160 papaya images. This model can easily classify image with nearly 70.14 percent accuracy. The accuracy of this model can be increased on training the model with large number of image dataset and by using local and global features such as BOVW (Bag of Visual Words), SURF (Speed Up Robust Features), SIFT (Scale-Invariant Feature Transform) etc. The comparisons of different machine learning algorithms are shown in Table 1 and graph (Fig. 8), respectively.

References

[1]Zhao, C., Lu, S. and Guo, X. 2010. SVM-based multiple classifier system for recognition of wheat leaf diseases. In Proceedings of 2010 Conference on Dependable Computing (CDC '2010).

[2] Rastogi, A., Arora, R. and Sharma, S. 2015. Leaf disease detection and grading using computer vision technology & fuzzy logic. In 2015 2nd international conference on signal processing and integrated networks (SPIN) (pp. 500-505). IEEE.

[3] Rothe, P. R., and Kshirsagar, R. V. 2015. Cotton leaf disease identification using pattern recognition techniques. In 2015 International conference on pervasive computing (ICPC) (pp. 1-6). IEEE.

[4] Owomugisha, G., Quinn, J. A., Mwebaze, E. and Lwasa, J. 2014. Automated vision-based diagnosis of banana bacterial wilt disease and black sigatoka disease. In International conference on the use of mobile ICT in Africa (pp. 1-5).

[5] Sannakki, S. S. and Rajpurohit, V. S. 2015. Classification of pomegranate diseases based on back propagation neural network. International Research Journal of Engineering and Technology (IRJET), 2(2).

[6] Yun, S., Xianfeng, W., Shanwen, Z. and Chuanlei, Z. 2015. PNN based crop disease recognition with leaf image features and meteorological data. International Journal of Agricultural and Biological Engineering, 8(4): 60-68.

[7] Arnal Barbedo, J. G. 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus, 2(1): 1-12.

[8] Caglayan, A., Guclu, O. and Can, A.B. 2013. 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., Bhaduri, M., Chan, I. and Li, S. 2014. Direct estimation of cardiac bi-ventricular volumes with regression forests. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 586-593). Springer, Cham.

[10] Wang, P., Chen, K., Yao, L., Hu, B., Wu, X., Zhang, J., Ye, Q. and Guo, X. 2016. Multimodal classification of mild cognitive impairment based on partial least squares. Journal of Alzheimer's Disease, 54(1): 359-371.