International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 6 ǁ 2021 ǁ PP. 50-57

Leaf Disease Detection Using Machine Learning

Prof. Rupali Tajanpure , Kishori Birari, Tejal Kuwar, Pooja Pawatekar, Devika Tatar
Department of Information Technology, MVPS's K.B.T. College of Engineering, Nasik, INDIA

ABSTRACT --: *Crop diseases are an interesting risk to sustenance security, however their quick distinguishing proof stays troublesome in numerous parts of the planet thanks to the non-attendance of the important foundation. Emergence of accurate techniques within 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 info 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 a picture we use Histogram of an Oriented Gradient (HOG). Overall, using machine learning to coach the massive data sets available publicly gives us a transparent thanks to detect the disease present in plants in an exceedingly colossal scale.*

KEYWORDS*: Random forest, Histogram of Oriented Gradients, Feature extraction, Training, Classification.*

Date of Submission: 03-06-2021 Date of acceptance: 17-06-2021

I. INTRODUCTION

Agriculture could be a backbone of our country. Farmers have enough crops for his or her farm. Regardless, growing crops to maximise known income and manufacturing is mostly medical. this could are developed with technical assistance. Supervising that routine plants always want to own excellent strength, especially to combat diseases which will have important consequences on the weather of production for gain. The imaging method is of the very best quality, which has driven activity in agricultural applications. In developing countries, over 80% of agricultural production is produced by smallholder farmers, and reports of crop losses of over 50% thanks to pests and diseases are common. Moreover, the bulk of individuals full of poverty and hunger (50%) board these productive areas, making smallholder farmers a really vulnerable group to food disruptions caused by pathogens. There are methods to see the diseases of any plant, like taking samples of vegetative tissue to a specialized laboratory or taking an expert agronomist to the cultivation site; in either of the 2 methods, the disadvantage lies within the time necessary to get the results. Random forests are as a full, learning method for classification, regression and other tasks that operate by constructing a forest of the choice trees during the training time. Unlike decision trees, Random forest 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 a part descriptor utilized as a component 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's moments is essentially accustomed extract the form of the leaves. Haralick texture is employed to induce the feel of the leaves and color Histogram is employed to represent the distribution of the colours in a picture.

II. LITERATURE REVIEW

[1] S. S. Sannakki and V. S. Rajpurohit, proposed a "Classification of Pomegranate Diseases supported Back Propagation Neural Network" which mainly works on the strategy of Segment the defected area and color and texture are used because the features. Here they used neural network classifier for the classification. the most 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 most disadvantage is that it's used just 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 wont to limit the vitality inside the infection spot, BPNN classifier tackles the many class problems. the common classification is found to be 85.52%.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma,” plant disease Detection and Grading using Computer Vision Technology &Fuzzy Logic”. K-means clustering accustomed segment the defected area; GLCM is employed for the extraction of texture features, symbolic logic is employed for disease grading. They used artificial neural network (ANN) as a classifier which mainly helps to test 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 accustomed create max tree, five shape attributes are used and area under the curve analysis is employed for classification. They used nearest neighbors, Decision tree, random forest, extremely randomized tree, Naïve bayes and SVM classifier. In seven classifiers extremely, randomized trees yield a really high score, provide real time information provide flexibility to the appliance.

[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 pictures size to a reduced uniform size. Then comes extracting features of a preprocessed image which is completed with the assistance of HOG. HoG could be a feature descriptor used for object detection. during this feature descriptor the looks of the item and therefore the outline of the image is described by its intensity gradients. one among 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 specific leaf. Hu moments are calculated over single channel only. the primary step involves converting RGB to Gray scale so 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 tell apart between the textures of healthy and diseased leaf. it's supported the adjacency matrix which stores the position of (I,J). Texture is calculated based on the frequency of the pixel I occupying the position next to pixel J. To calculate Haralick texture it's required that the image be converted to gray scale.

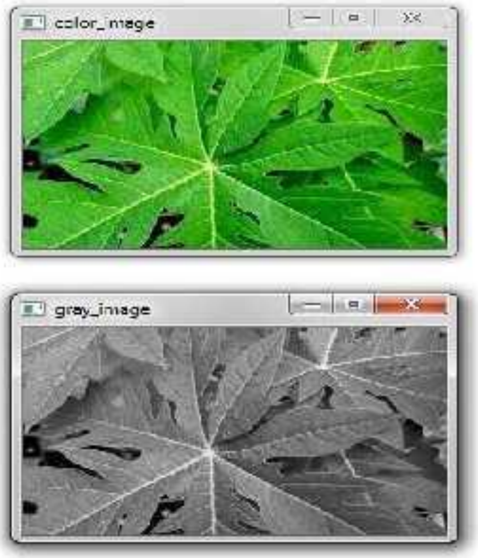

Fig.1. RGB to Gray scale conversion of a leaf.

Color Histogram: Color histogram gives the representation of the colours within the image. RGB is first converted to HSV color space and therefore the histogram is calculated for the identical. it's needed to convert the RGB image to HSV since HSV model aligns closely with how human eye discerns the colours in a picture. Histogram plot [8] provides the outline about the quantity of pixels available within the given color ranges

Fig.2. RGB to HSV conversion of leaf

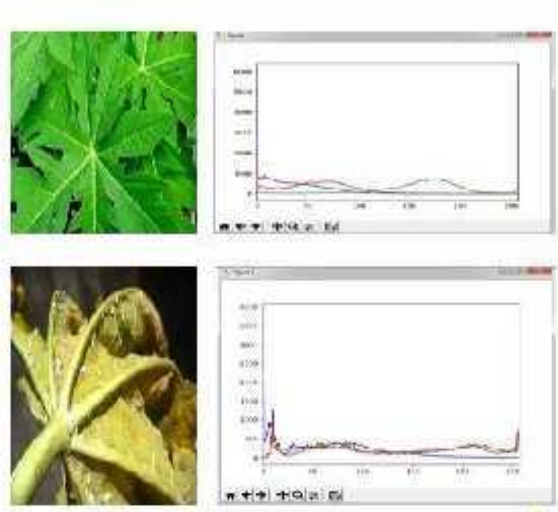

Fig.3. Histogram plot for healthy and diseased leaf

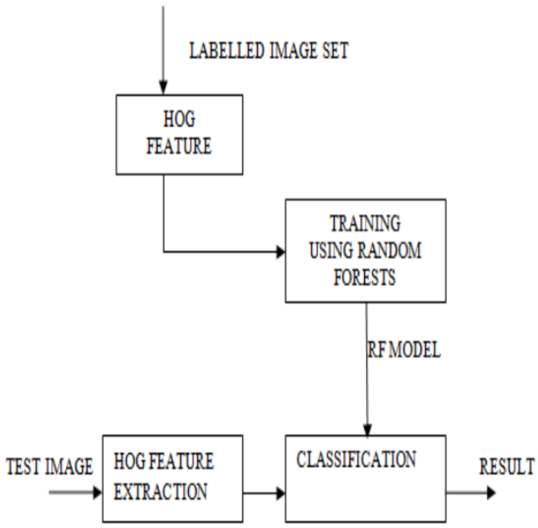

Fig.4. Architecture of the proposed model

In proposed model, we first provide labelled image set of infected and non-infected leafs then HOG extract that features from labelled imaged set and provides that features for training purpose to random forest model. Random forest makes classification supported that labelled imaged set. Then when user gives a leaf image as input then HOG extract the features from that leaf image then random forest model analyzed the features of that leaf then predict the disease of that leaf supported its stored labelled imaged set features. We get results of diseases of given leaf.

IV. ALGORITHM DESCRIPTION

The algorithm here is implemented using random forests classifier. they're flexible in nature and might 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 subsequent figure shows the architecture of our proposed algorithm.

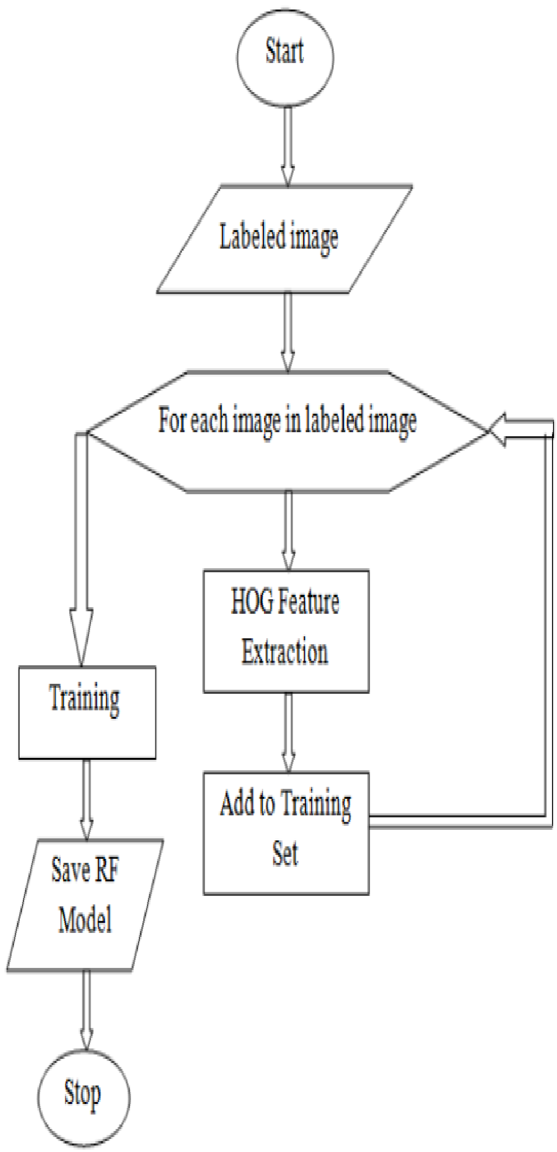

Fig.5. Flow chart for training.

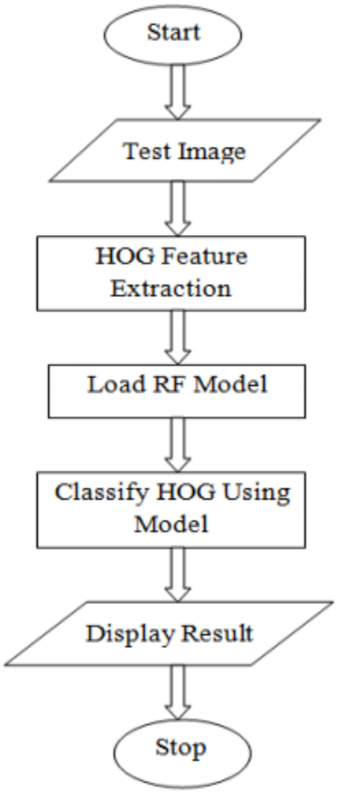

Fig.6. Flow chart for classification

he 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 noted in “Fig.4”. As shown within 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. 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.

Feature Extraction:

Histogram Oriented Gradient (HOG) Features. The Histogram of Oriented Gradient (HOG) may be a feature descriptor employed in computer vision and image processing for the aim of object detection. The procedure includes events of angle introduction confined bits of an image. this system is sort of a scale-invariant element change descriptors, however, vary therein it's processed on a thick framework of consistently divided cells and utilizations covering nearby different standard for enhanced precision.

1. Gradient Computation : so as to seek out the gradient, the grey scale image is filtered to urge the x and y derivatives of filters the foremost requires filtering the colour or intensity data of the image common method is to use the 1-D centered, point discrete derivative masking. Specifically, this method with the subsequent filter kernels:
Dx= [-1 0 +1] and Dy= [-1 0 +1] T
After finding x and y derivatives (Ix and Iy), the magnitude and orientation of the gradient are calculated: |G|= Ix^2 + Iy^2 and θ= arctan Iy/Ix
The orientation calculation method returns values between [-180◦,180◦]. Since unsigned orientations are desired for this implementation, the values which are but 0◦ is summed up with 180◦.
2. Orientation Binning : the subsequent step is to calculate the cell histogram for descriptor blocks. The 8x8 pixel size cells are computed with 9 orientation bins for [0◦,180◦] interval. for every pixel’s orientation, the

corresponding orientation bin is found and therefore the orientation's magnitude |G| is voted to the current bin.
3. Descriptor Blocks : To normalize the cells' orientation histograms, they should be combined into blocks. From the 2 main blocks, geometric, the implementation uses R-HOG geometry. Each R-HOG block has 2x2 cells and adjacent R-HOGs are overlapping one another for a magnitude of half-size of a block.
4.Block Normalization : Although there are three different methods for block normalization, L2-Norm normalization is implemented using norm (vec) method:
f=v/||v||2^2+e^2
5. Detection Window : Each R-HOG block has 2x2 cells within which each cell is 8x8 pixels, which also has 1x9 histogram vector each. therefore the overall size of R-HOG descriptor of a window is given by 144 values. Final vector size= 2 blocks horizontally x 2 blocks vertically x 4 cells per block x 9 bins per histogram = 144 values.

Classification by Random Forest Classifier Finally, after extracting the features of all the photographs, we proceeded to separate them into two classes: test and test, because for the training of the model, we'll work with the test class, and to verify the model, we are going to use the test class. Random forest classifier is used as the learning algorithms, explained as follows. Random Forests are supported a heuristic division of the outline space. the development of the decision structure of a tree is administered by recursive partitioning of this space, which makes the ultimate decisions strongly smitten by the upstream divisions. The search space of the possible decision tree structures is then strongly restricted by these dependencies. specifically, there are many techniques that benefit of the chance to make diversity in three communities. Breiman presented a proper framework for this category of methods called random forest (RF). A random forest may be a classifier consisting of a group of base classifiers like a choice tree show:

Random forests are composed of a set of binary decision trees within which randomness has been introduced. Random forests were introduced by Breiman (2001) by the subsequent very general definition [29]: Let a set of tre predictors, with random variables independent. The predictor of random forests is obtained aggregating this collection of random trees as follows:
• Average of individual tree predictions in regression
•A majority vote among individual prediction trees in classification

The term random forest comes from the actual fact that individual predictors are, here, explicitly predictors per tree which each tree depends on a further variate (that is, additionally to). In terms of precision, the results obtained with this approach are reminiscent of those obtained directly using the one against the remainder method.

V. RESULT

First for any image we want to convert RGB image into gray scale image. is} done simply because Hu moments shape descriptor and Haralick features can be calculated over single channel only. Therefore, it's necessary to convert RGB to gray scale before computing Hu moments and Haralick features. As depicted within 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 most aim of our project is to detect whether it's diseased or healthy leaf with the assistance of a Random forest classifier which is as depicted within the "Fig.7."

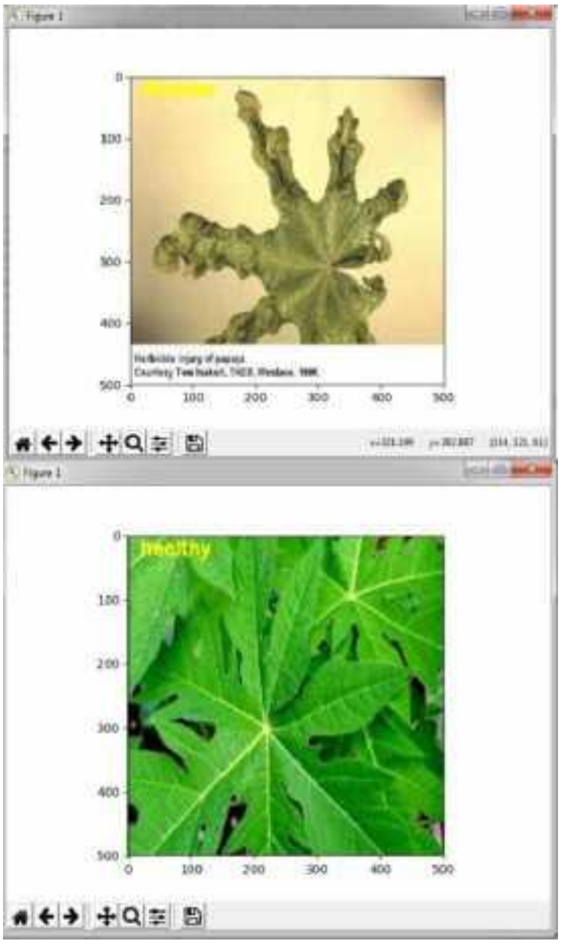

Fig.7. Final output of the classifier.

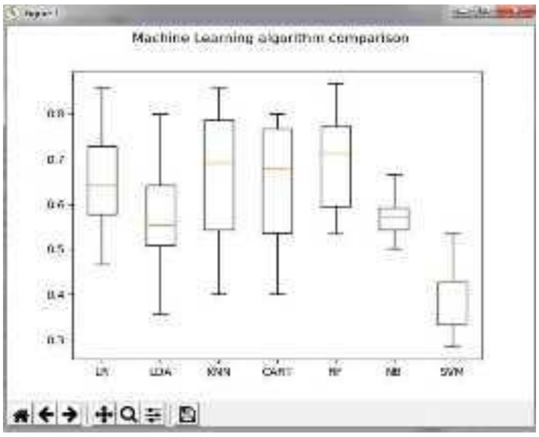

Fig.8. Comparison between different machine learning models.

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

VI. CONCLUSION:

The objective of this algorithm is to acknowledge abnormalities that occur on plants in their greenhouses or natural environment. The image captured is typically soft on an understandable 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 are often increased when trained with vast number of images and by using other local features along with the worldwide features like SIFT (Scale Invariant Feature Transform), SURF (Speed Up Robust Features) and DENSE together 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 supported Back Propagation Neural Network," International Research Journal of Engineering and Technology

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

[3]. Aakanksha Rastogi, Ritika Arora and Shanu Sharma," plant disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks

[4]. Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial plant disease and Black Sigatoka Disease ", Preceding of the 1'st international conference on the utilization 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.

[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 supported partial least squares".