INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1458-1462
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

CROP DISEASE VISUAL IDENTIFICATION USING MACHINE LEARNING

Bakka Rajeev Gandhi[1], Jonnakuti Rakesh Babu[2]

[1]Assistant Professor, Department of Computer Science & Engineering, Chalapathi Institute of Technology, Guntur, A.P, 522016

[2]Assistant Professor, Department of Computer Science & Engineering, Chalapathi Institute of Technology, Guntur, A.P, 522016

DOI: https://www.doi.org/10.58257/IJPREMS37861

ABSTRACT

A vital part of human civilization, agricultural biodiversity is necessary to supply humans with food and raw materials. The crop disease is mainly caused by pathogenic organisms such as fungi, bacteria, and nematodes; soil pH,temperature extremes,variations in the amount of moisture and humidity in the air and other factors. Research in crop disease is aimed towards an increase of productivity and quality of food.In this paper an automatic identification and classification of crop diseases using machine learning methods based on Image Processing, Support Vector Machines (SVMs), and Random Forest.Using these techniques, crop disease identification made it possible to get notable accuracies.

Keywords: Image Processing (IP), Machine Learning(ML), Random Forest, Classification and Plant Disease

1. INTRODUCTION

The agriculture industry contributes significantly to the economy of most countries and is the primary source of revenue in many developing countries.As the world's population grows, so does the food demand. To meet the challenges, food production must be prioritized. Unfortunately, pests, weeds, animals, and plant diseases and The impact of rapid weather change not only decreases the lifetime of a plant but also causes a problem with the quick diagnosis of infected plants .have a considerable impact on crop production quality and quantity worldwide as they are responsible for crop output losses, leading to a huge loss.

Crop leaf diseases vary in shape, size, and color. Certain diseases might have the identical color, but dissimilar shapes; while some have dissimilar colors but identical shapes. The model can be developed by capturing the diseased leaves and recognizing the patterns about the disease is helpful to get free of crop loss due to disease spread or increase [1] . In this approach, the images are often sent to a core crop leaf disease system for analysis; the system can recognize. The system generates information about the crop leaf disease.

In order to explore the use of Image processing for classifying citrus leaf diseases, a research [2] on 2006, was conducted. For this analysis, four distinct citrus leaf disease groups were used, namely greasy spot, melanose, healthy and scab. For feature extraction and classification the proposed algorithms based on image processing techniques were developed. The process of extraction of features used the technique of color co-occurrence, which uses an image's color and texture to achieve specific characteristics that reflect the image On all data models using intensity features, SAS discriminant analysis, hue and saturation features, hyperspectral image (HSI) features presented the results above 81 percent, and above 95.8 percent respectively.

Modern approaches such as machine learning and deep learning algorithms have 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 as the machine learning approach being random forest, support vector machine(SVM), etc. Random forests are, as a whole, a 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 forests 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

2. METHODOLOGY

Computer vision is a subdomain of AI that allows machines to counterfeit the human visual system and precisely draw out, inspect, and recognize real-world images in the same way that humans do [3].ML techniques have been used to detect and classify crop diseases, but with advancements in a subset of ML, DL, this area of research appears to have considerable potential in terms of increasing accuracy. Many developed DL architectures were used, along with various visualization techniques, to detect and classify plant disease symptoms accordingly [4].

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 [5] 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 advantages 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.

2.1 Hu moments

Image moments which have the important characteristics of the image pixels help in describing the objects. Here Hu moments help in describing the outline of a particular leaf. Hu moments are calculated over a 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.

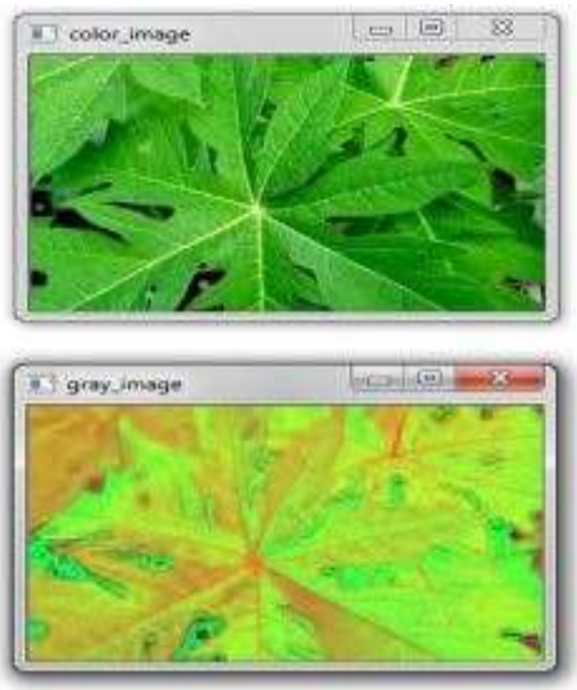
2.2 Haralick Texture

Usually the healthy leaves and diseased leaves have different textures. Here we use Haralick texture features 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 [6] 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.

2.3 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 the HSV model aligns closely with how human eye discerns the colors in an image. Histogram plot [7] provides the description about the number of pixels available in the given color ranges



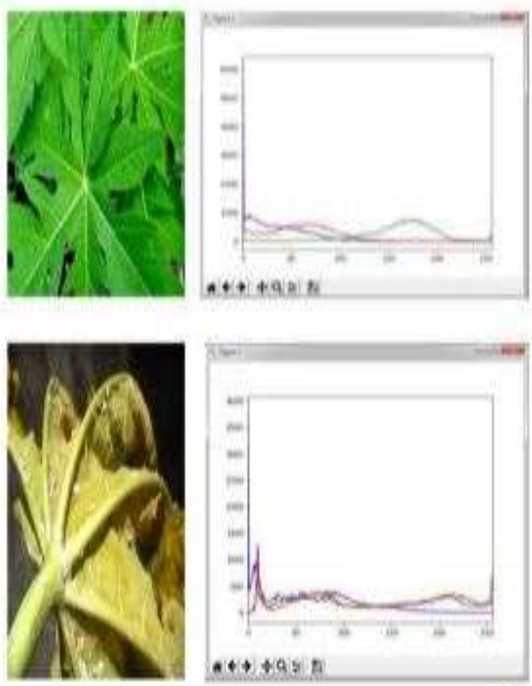


Figure 1: RGB to HSV conversion of leaf **Figure 1.1:** Histogram plot for healthy and diseased leaf.

3. MODELING AND ANALYSIS

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.

Vol. 04, Issue 12, Decembaer 2024, pp : 1458-1462
e-ISSN : 2583-1062
Impact Factor : 7.001

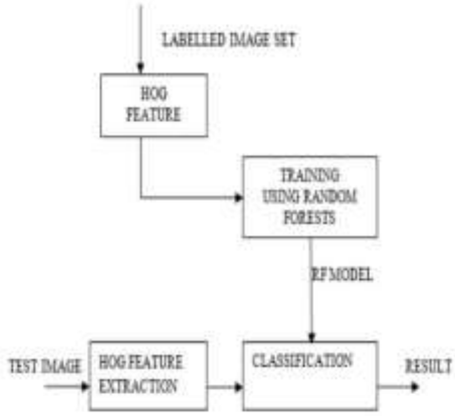


Figure 2:.Proposed model

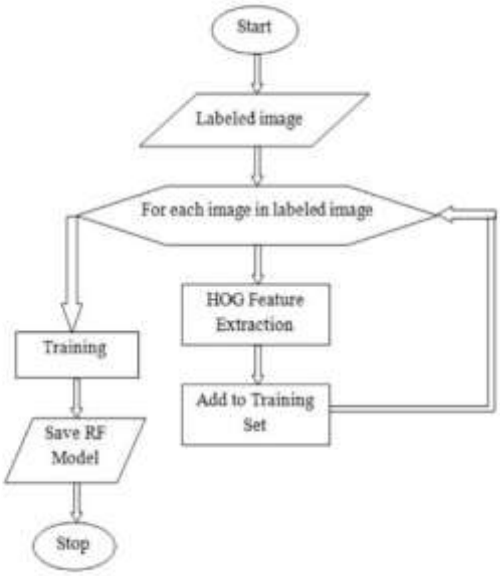


Figure 3 :.Flow Chart

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.2”.

As shown in the ‘Fig.3.” 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 [8, 9].As depicted in “Fig.4.” 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. 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.3.1.”

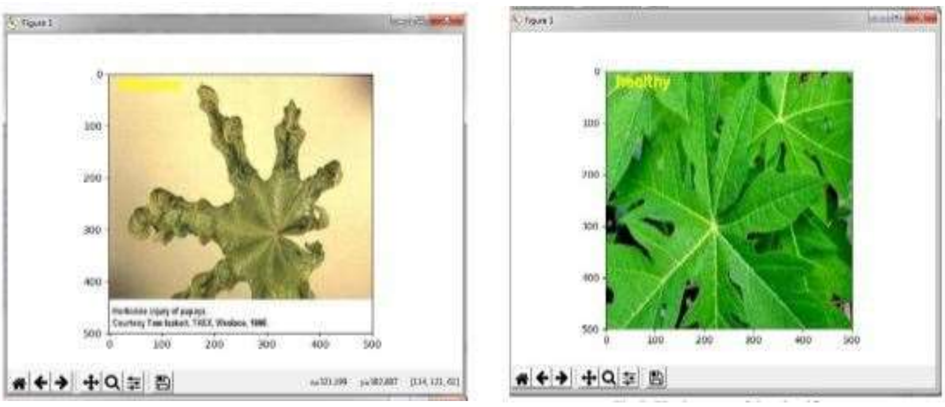


Figure 3.1: Proposed model

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1458-1462
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

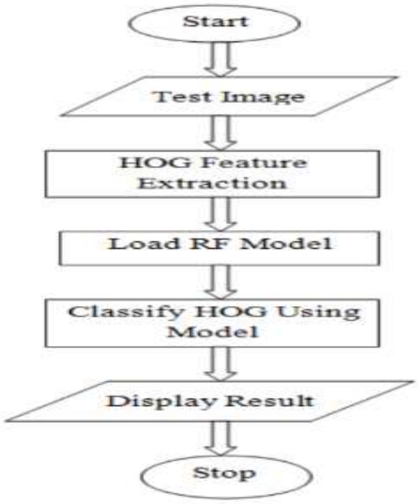


Figure 4 :.Flow Chart

4. RESULTS AND DISCUSSION

First for any image we need to convert the RGB image into a gray scale image. This is done just because Hu moments' shape descriptor and Haralick features can be calculated over a 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 1.To calculate histogram the image first must be converted to HSV (hue, saturation and value), so we are converting the RGB image to an HSV image as shown the figure1.1. Finally, the main aim of our project is to detect whether it is diseased or healthy leaf with the help of a Random forest

Table 1. Comparison

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

5. 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

6. REFERENCES

[1] T. Youwen, L. Tianlai, and N. Yan, "The recognition of cucumber disease based on image processing and support vector machine," in 2008 Congress on Image and Signal Processing, 2008, pp. 262–267, doi: 10.1109/CISP.2008.29

[2] R. Pydipati, T. F. Burks, and W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," Computers and Electronics in Agriculture, vol. 52, no. 1–2, pp. 49–59, Jun. 2006, doi: 10.1016/j.compag.2006.01.004.

[3] Kumar R, Chug A, Singh AP, Singh D. A systematic analysis of machine learning and deep learning based approaches for plant leaf disease classifcation: a Review. J Sensors. 2022. https://doi.org/10.1155/2022/3287561.

[4] Saleem MH, Potgieter J, Arif KM. Plant disease classifcation: a comparative evaluation of convolutional neural networks and deep learning optimizers. Plants. 2020;9(10):1–17. https://doi.org/10.3390/plants9101319.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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”.

[10] J.Rakesh Babu.,” Road sign intimation through voice alert system using deep Learning”, International Journal for Advanced Research in science & Technology, ISSN 2457 – 0362

[11] Bakka Rajeev Gandhi “Prediction And Classification Of Alzhimer’s Disease Using Machine Learning Techniques In 3d Mr Images”, at the international conference on sustainable computing and smart systems(ICSCSS 2023)

[12] J.Rakesh Babu.,” Road Sign Intimation Through Voice Alert System Using Deep Learning ”, International Journal for Advanced Research in science & Technology, ISSN 2457 – 0362

[13] Jonnakuti Rakesh Babu,Bakka Rajeev Gandhi ”Anomaly Attack Identification Security System Using Artificial Intelligence and Deep Learning”International Journal of Innovative Research in Computer and Communication Engineering,e-ISSN: 2320-9801, p-ISSN: 2320-9798

[14] J Rakesh Babu,”Neural Network Approach For Prediction Of Crop Yield Using RNN, Feed Forward And LSTM Algorithms”International Journal of Engineering Science and Advanced Technology (IJESAT),ISSN No: 2250-3676,Vol24 Issue 02, 2024