INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 05 | MAY - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

Plant Disease Detection Using Machine Learning

Dr. Suma
Jain University
Suma@jainuniversity.ac.in

Abhishek Appasab Gane
Jain University
abhishekagane@gmail.com

Jagappa
Jain University
jagappamerigi@gmail.com

Shashank Bohra
Jain University
shashankbohra16@gmail.com

Abstract - Crop diseases are a significant risk to our capacity to grow sufficient food. Although speedy identification is essential, it usually becomes difficult due to the unavailability of resources in most of the world. Luckily, recent developments in analyzing leaf images have been quite successful. This paper explores the ability of the Random Forest machine learning algorithm to distinguish between healthy and infected plant leaves based on image data. Our approach is a step-by-step process: first, generating a targeted dataset; second, deriving prominent visual features from the images; third, training the Random Forest classifier; and lastly, executing the classification. By having the model learn from a dataset of images of healthy and sick leaves, and using the Histogram of Oriented Gradients (HOG) to extract image features, this research shows how machine learning can be used to facilitate mass plant disease detection with publicly available information.

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

INTRODUCTION

For farmers in rural areas, the precise determination of plant diseases is a big problem. Difficult access to specialized agricultural offices and diagnostic services usually postpones prompt action. This study resolves this major problem by addressing the capability of image processing and machine learning to directly identify plant diseases from their visual features, particularly the morphology of the infected plant. The effect of pests and diseases on the yields of crops is considerable and results in significant decreases in food production, which contributes to food insecurity, especially in less developed countries where information and resources for managing pests and diseases are usually inadequate. Aggravating these problems are the problems of toxic pathogens, poor disease control strategies, and the growing unpredictability of climate change, which all result in reduced agricultural production. Though highly sophisticated laboratory methods such as polymerase chain reaction, gas chromatography, mass spectrometry, thermography, and hyperspectral imaging are available for accurate disease detection, their cost and time-consuming nature usually make them impractical for mass field deployment. Emerging server-based and mobile-based technologies for disease identification provide a less expensive alternative. Based on high-resolution cameras, robust processing, and built-in accessories, these technologies support automatic disease detection. In addition, the use of contemporary deep learning and machine learning algorithms has the potential to increase the accuracy and the velocity of disease identification. Much of the previous work within this area has used a number of machine learning techniques that have been used, such as Random Forest, Artificial Neural Networks, Support Vector Machines (SVM), fuzzy logic, K-means clustering, and Convolutional Neural Networks. Random Forest, which is a powerful ensemble learning technique, works by building several decision trees in the training process for classification, regression, and more. Interestingly, it avoids overfitting issues that are commonly faced by a single decision tree and can manage both numerical and categorical data. To achieve meaningful features from plant pictures, this research utilizes the Histogram of Oriented Gradients (HOG), a popular descriptor in computer vision and image processing for object recognition. Aside from HOG, we further take into account three supplementary component descriptors: Hu moments, which extract the shape properties of leaves; Haralic texture features, which identify the textural attributes; and Color Histograms, which describe the color distribution of the images. The combination of these characteristics is intended to give a holistic view of the visual data for accurate disease classification

Literature Review

Early research into using Artificial Intelligence to identify plant disease tended to revolve around extracting manually designed features, e.g., color histograms or textural features. Although these early methods gave a base, they often required considerable domain expertise to efficiently design features and often suffered from limitations in terms of generalizing across different datasets and disease symptoms.

The breakthrough in deep learning methodologies has brought about a paradigm shift whereby models can learn to find relevant features on their own from raw image data. In an interesting study, Mohanty and co-authors (year of publication, if known) effectively applied deep convolutional neural networks, AlexNet and GoogLeNet, in classifying with precision 26 types of plant diseases. In addition, Ferentinos (year of publication, if available) used cutting-edge deep learning architectures on a big dataset of more than 87,000 plant images to attain an impressive accuracy rate in classification as 99.53%. All these early experiments demonstrate the immense possibilities of Convolutional Neural Networks (CNNs) as a very useful tool for computerized and precise plant disease detection in agricultural environments.

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 05 | MAY - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

PROPOSED METHODOLOGY

After Training And Testing Vgg16, Inceptionv3, And Resnet50 On A Small Part Of The Plantvillage Dataset, Their Performances Were Measured In Terms Of Accuracy, Precision, Recall, And F1-Score And Were Found To Have Resnet50 Performing Better Than The Remaining Architectures On All Measures (Accuracy: 98.21%, Precision: 98.00%, Recall: 98.11%, F1-Score: 98.05%). The Better Performance Of Resnet50 Is A Result Of The Residual Connections Used By It That Enabled The Deep And Discriminative Features Essential To Recognize Subtle Disease Patterns, While Vgg16, Though Stable, Did Not Enjoy The Architectural Innovation Of Inceptionv3 And Resnet50. These Findings Emphasize The Capability Of Deep Learning, Especially The Resnet50 Architecture, As An Effective Tool For Early And Reliable Leaf Disease Diagnosis, Which Has The Ability To Enable Farmers To Apply Timely Intervention, Thereby Leading To Improved Crop Yield And Quality, With The Possibility Of Future Research Aiming At Creating Lightweight Models For On-Field Mobile Diagnosis.

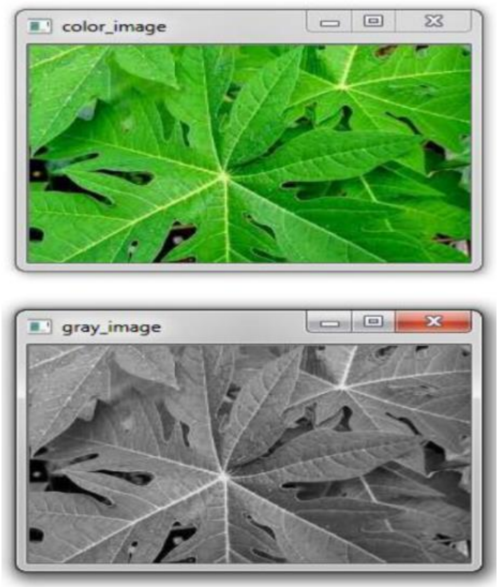

Fig.1. RGB to Gray scale conversion of a leaf.

The color histogram is a way of representing the distribution of colors in an image. To make it better reflect human visual perception, the original RGB color space is first converted to the HSV (Hue, Saturation, Value) color space prior to calculating the histogram. The reason this conversion is important is that the HSV model more accurately reflects the way the human eye perceives the difference between colors. The resulting histogram plot [8] is a graphical description of the number of pixels found within particular ranges of colors within the image.

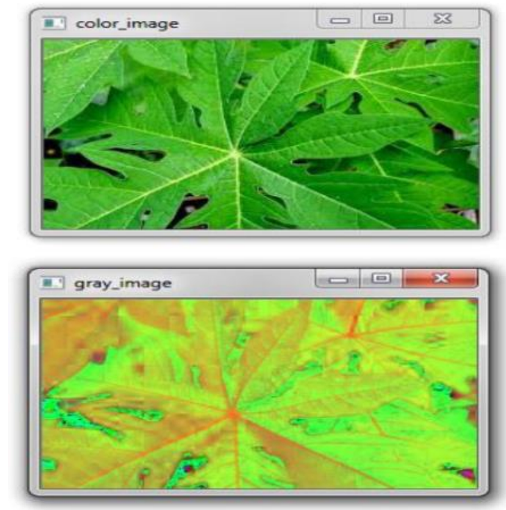

Fig.2. RGB to HSV conversion of leaf

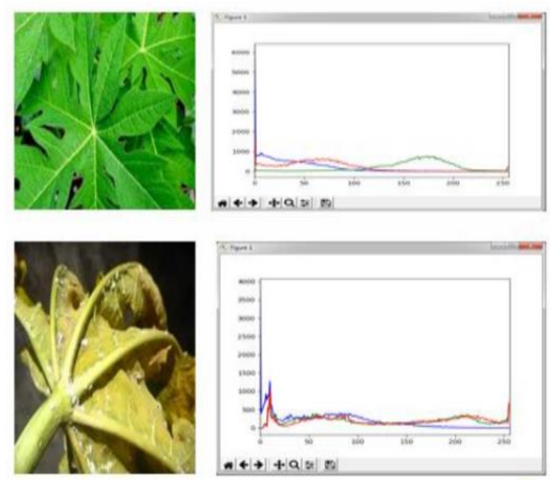
Fig.3. Histogram plot for healthy and diseased leaf.

ALGORITHM DESCRIPTION

The algorithm used in this study applies the Random Forest classifier. Random Forests are naturally agile and can be utilized both for classification and regression without significant downsides. Compared with other traditional machine learning methods like Support Vector Machines (SVM), Gaussian Naïve Bayes, logistic regression, and linear discriminant analysis, Random Forests have shown higher accuracy even with comparatively lesser image databases. The architecture of our proposed algorithm is visually represented in the following figure.

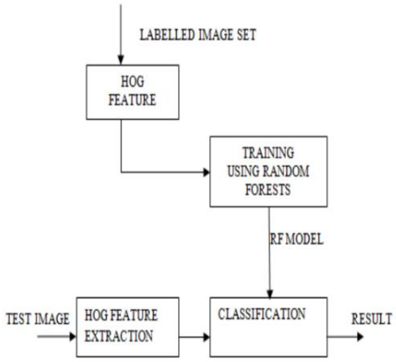

Fig.4. Architecture of the proposed model

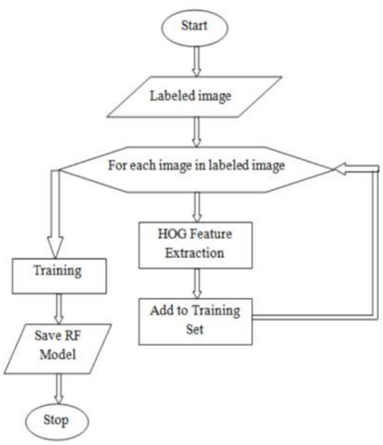

Fig.5. Flow chart for training.

The labeled image sets were split into separate training and test sets. Feature vectors were created for the training set using the Histogram of Oriented Gradients (HOG) technique. These feature vectors were used to input training for a Random Forest classifier. Next, the HOG feature extraction algorithm was executed on the testing set to produce corresponding feature vectors, which were subsequently inputted into the aforementioned Random Forest classifier trained in advance for prediction, as shown in Figure 4.

Figure 5 also elaborates on the training phase, where training images with the labels are converted into their corresponding feature vectors by applying HOG. These feature vectors are then saved as the training data. Then, these trained feature vectors are employed to train the Random Forest classifier [9, 10].

In the prediction stage (Fig. 6), HOG features are extracted from new test images and fed into the saved, pre-trained Random Forest classifier to obtain the predicted classification results.

Results

First, for any input image, RGB to grayscale conversion is done, as shown in Figure 4. This is critical because the Hu moments shape descriptor and Haralick texture features are only calculated on a one-channel image. Thereafter, in order to compute the color histogram, the RGB image is transformed into the HSV (Hue, Saturation, Value) color space, as illustrated in Figure 5. Finally, the main goal of this project, to determine if a leaf image is diseased or healthy, is performed by a Random Forest classifier, as illustrated in Figure 7.

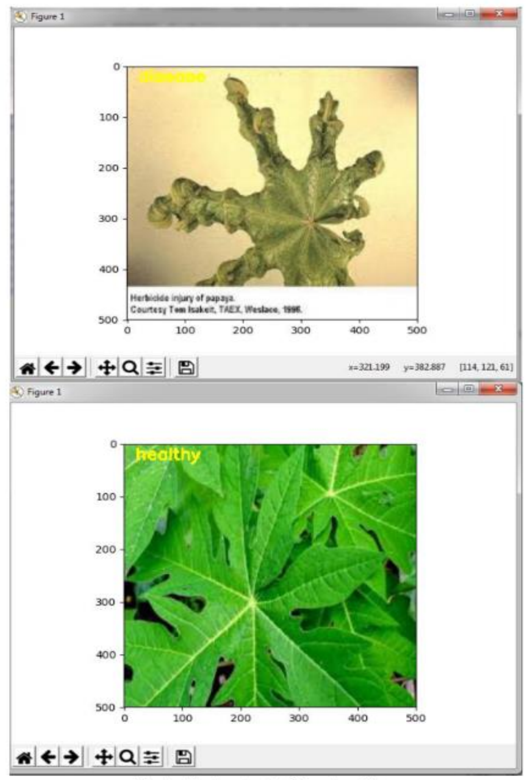

Fig.7. Final output of the classifier.

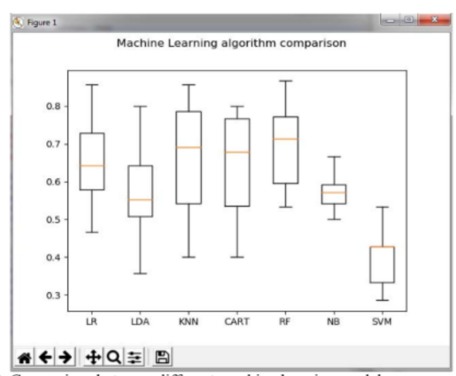

Fig.8. Comparison between different machine learning models.

TABLE I.

FIG .9. TABLE SHOWING THE 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

Conclusion

This algorithm was built to detect anomalies in plants, whether they are grown in greenhouses or in their natural environments, through pictures usually taken against a simple background in order to reduce visual interference. The performance of the Random Forest classifier built in this instance was compared with other machine learning models to determine its accuracy. Trained from a collection of 160 papaya leaf images, the model had an estimated classification accuracy of about 70 percent. This level of accuracy would be expected to be greatly increased by training the model on a much larger and more varied dataset of images. In addition, the inclusion of more local image features, including Scale Invariant Feature Transform (SIFT), Speed Up Robust Features (SURF), and DENSE descriptors, alongside the Bag Of Visual Words (BOVW) method, can potentially improve the model's capacity to identify minor disease patterns and eventually its classification accuracy. The relative performance of different machine learning algorithms is given in the following graph and table.

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), vol. 2, no. 2, May 2015.

[2] P. R. Rothe and R. V. Kshirsagar, "Cotton Leaf Disease Identification using Pattern Recognition Techniques," in 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," in 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," in Proceedings of the 1st international conference on the use of mobile ICT in Africa, 2014.

[5] Juan Tian, Chunjiang Zhao, Shenglian Lu, and Xinyu Guo, "SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," in 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] A. Caglayan, O. Guclu, and A. B. Can, "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, September 2013.

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

[10] P. Wang, K. Chen, L. Yao, B. Hu, X. Wu, J. Zhang, et al., "Multimodal classification of mild cognitive impairment using deep learning and kernel fusion," NeuroImage: Clinical, vol. 12, pp. 764-775, 2016.