Educational Administration: Theory and Practice
2024, 30(4), 10311-10316
ISSN: 2148-2403
https://kuey.net/

Research Article

Study Of Plant Disease Detection Using Machine Learning And Deep Neural Network

Sarita [1*], Amit Sharma[2]

[1*]Student of M.Tech Computer Science and Engineering, SCRIET, Chaudhary Charan Singh University Campus, Meerut, India Email: nagarsarita1991@gmail.com
[2] Assistant Professor, Department of Computer Science and Engineering, SCRIET, Chaudhary Charan Singh University Campus, Meerut, India, Email: amitccsu.net@gmail.com

Citation: Sarita, et al. (2024), Study of Plant Disease Detection Using Machine Learning and deep Neural Network, *Educational Administration: Theory and Practice*, 30(4), 10311-10316
Doi: 10.53555/kuey.v30i4.6627

ARTICLE INFO

ABSTRACT

Crop diseases pose a significant risk to food security, yet their rapid identification remains challenging in many parts of the world due to the lack of necessary infrastructure. The emergence of accurate techniques in the field of leaf-based image classification has shown promising results. This paper utilizes Random Forest to distinguish between healthy and diseased leaves from datasets created for this purpose. Our proposed methodology includes various implementation phases: dataset creation, feature extraction, training the classifier, and classification. The datasets of diseased and healthy leaves are collectively trained using Random
Forest to classify the images accurately. For feature extraction, we use Histogram of Oriented Gradient (HOG). Overall, employing machine learning to train the publicly available large datasets provides a clear method to detect plant diseases on a large scale.

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

I. INTRODUCTION

Farmers in rural areas often find it difficult to identify diseases present in their crops. It is not always feasible for them to visit agricultural offices to diagnose crop diseases. Our main objective is to identify the disease in a plant by observing its morphology through image processing and machine learning.
Pests and diseases lead to crop destruction or damage, resulting in decreased food production and food insecurity. Knowledge about pest management and disease control is limited in various less developed countries. Toxic pathogens, poor disease control, and drastic climate changes are key factors leading to reduced food production.
Various modern technologies have emerged to minimize postharvest processing, fortify agricultural sustainability, and maximize productivity. Laboratory-based approaches such as polymerase chain reaction, gas chromatography, mass spectrometry, thermography, and hyperspectral techniques have been employed for disease identification. However, these techniques are costly and time-consuming.
Recently, server-based and mobile-based approaches for disease identification have been developed, featuring high-resolution cameras, high-performance processing, and extensive builtin accessories, resulting in automatic disease recognition. Modern approaches using machine learning and deep learning algorithms have increased recognition rates and accuracy. Research in machine learning for plant disease detection includes methods like Random Forest, artificial neural networks, support vector machines (SVM), fuzzy logic, K-means clustering, and convolutional neural networks (CNNs).
Random Forest is a versatile learning method for classification, regression, and other tasks, constructing a forest of decision trees during training. Unlike decision trees, Random Forests overcome the disadvantage of overfitting their training dataset and handle both numeric and categorical data.
The Histogram of Oriented Gradients (HOG) is an element descriptor used in computer vision and image processing for object detection. We utilize three feature descriptors:

1. Hu moments

2. Haralick texture
3. Color Histogram

Hu moments extract the shape of the leaves. Haralick texture distinguishes between the textures of healthy and diseased leaves. The color histogram represents the distribution of colors in an image.

II. LITERATURE REVIEW

Several studies have explored various techniques for leaf disease detection:

1. S. S. Sannakki and V. S. Rajpurohit (2015) used a Back Propagation Neural Network for classifying pomegranate diseases, achieving 97.30% accuracy.
2. P. R. Rothe and R. V. Kshirsagar (2015) employed pattern recognition techniques and a BPNN classifier for cotton leaf disease identification, with an average classification accuracy of 85.52%.
3. Aakanksha Rastogi et al. (2015) used K-means clustering, GLCM for texture features, and fuzzy logic for disease grading, with ANN as the classifier.
4. Godliver Owomugisha et al. (2014) proposed automated vision-based diagnosis using color histograms and multiple classifiers, with extremely randomized trees yielding high accuracy.
5. Juan Tian et al. (2010) used SVM classifiers for wheat leaf disease recognition, using color features and GLCM for shape parameters.

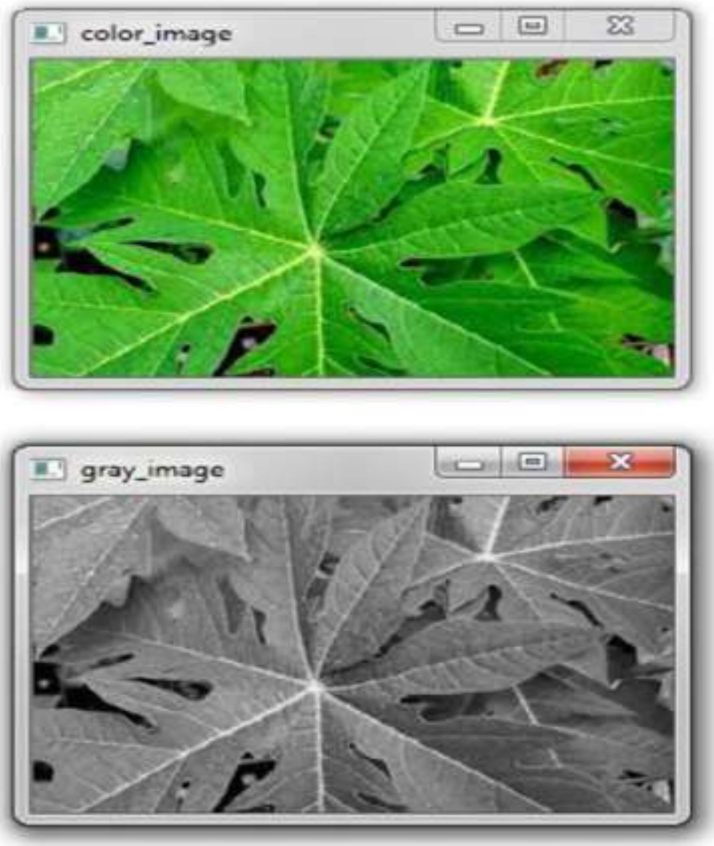


Fig.1. RGB to Gray scale conversion of a leaf.

III. PROPOSED METHODOLOGY

To determine if a leaf is diseased or healthy, we follow these steps: preprocessing, feature extraction, training the classifier, and classification. Preprocessing involves resizing all images to a uniform size. Feature extraction uses HOG, which describes the appearance and outline of the object by its intensity gradients.

Three feature descriptors are used:

- Hu moments: Calculated over a single channel after converting RGB to grayscale, describing the shape of the leaves.
- Haralick Texture: Distinguishes textures based on the frequency of pixel adjacency, calculated after converting the image to grayscale.
- Color Histogram: Represents the colors in an image, calculated after converting RGB to HSV color space.

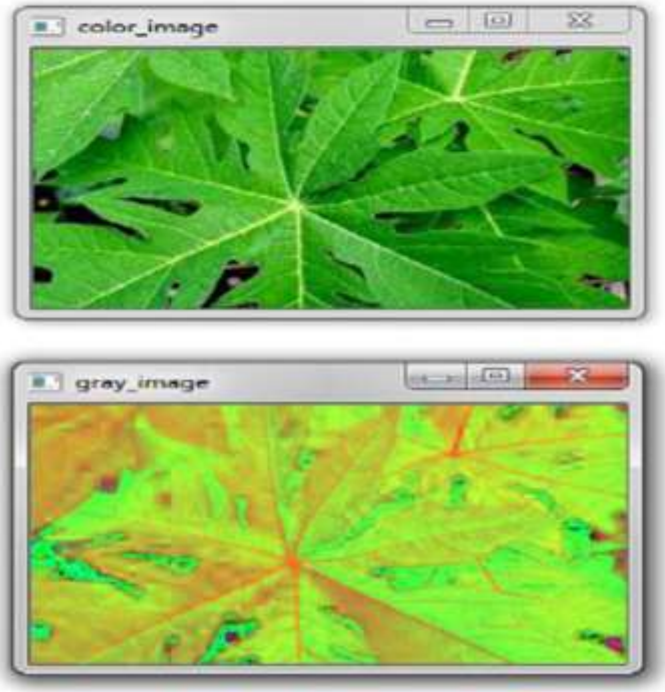


Fig.2. RGB to HSV conversion of leaf

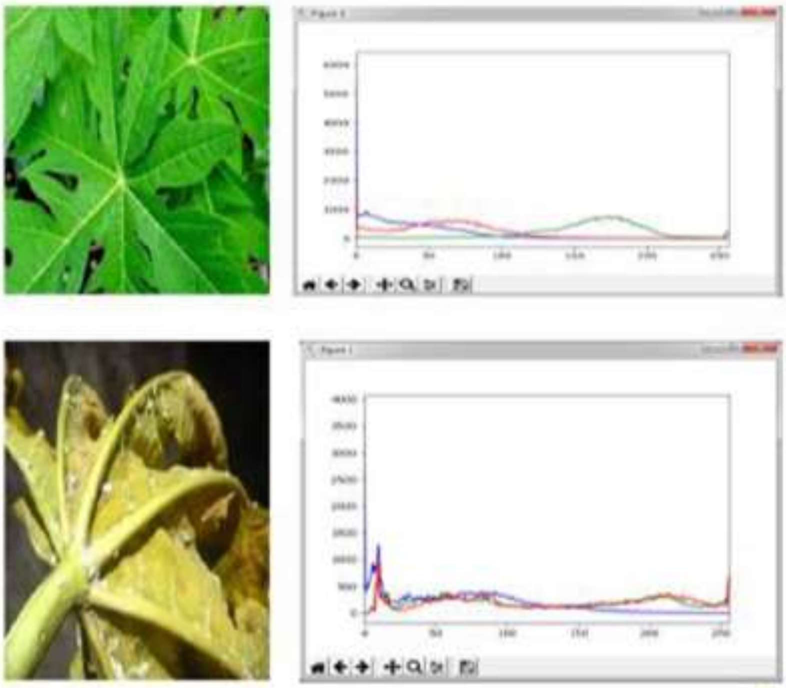
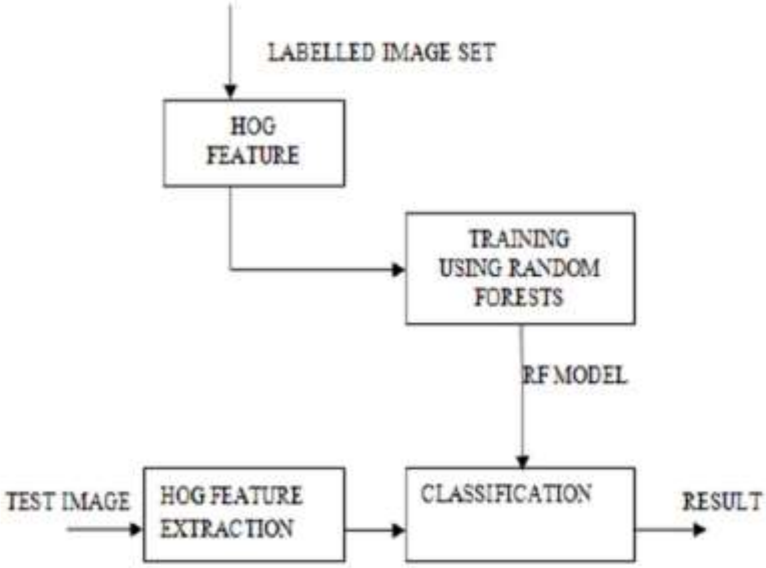


Fig.3. Histogram plot for healthy and diseased leaf.

IV. ALGORITHM DESCRIPTION

The algorithm uses a Random Forest classifier for its flexibility and accuracy with smaller datasets. The architecture involves:

- Segregating labeled datasets into training and testing data.
- Generating feature vectors for the training dataset using HOG feature extraction.
- Training the Random Forest classifier with the generated feature vectors.
- Generating feature vectors for the test images and using the trained classifier for prediction.



Using Random Forest, the model was trained with 160 images of papaya leaves, achieving approximately 70% accuracy. This accuracy can be improved by increasing the dataset size and incorporating other features such as SIFT, SURF, and BOVW.

Fig.4. Architecture of the proposed model

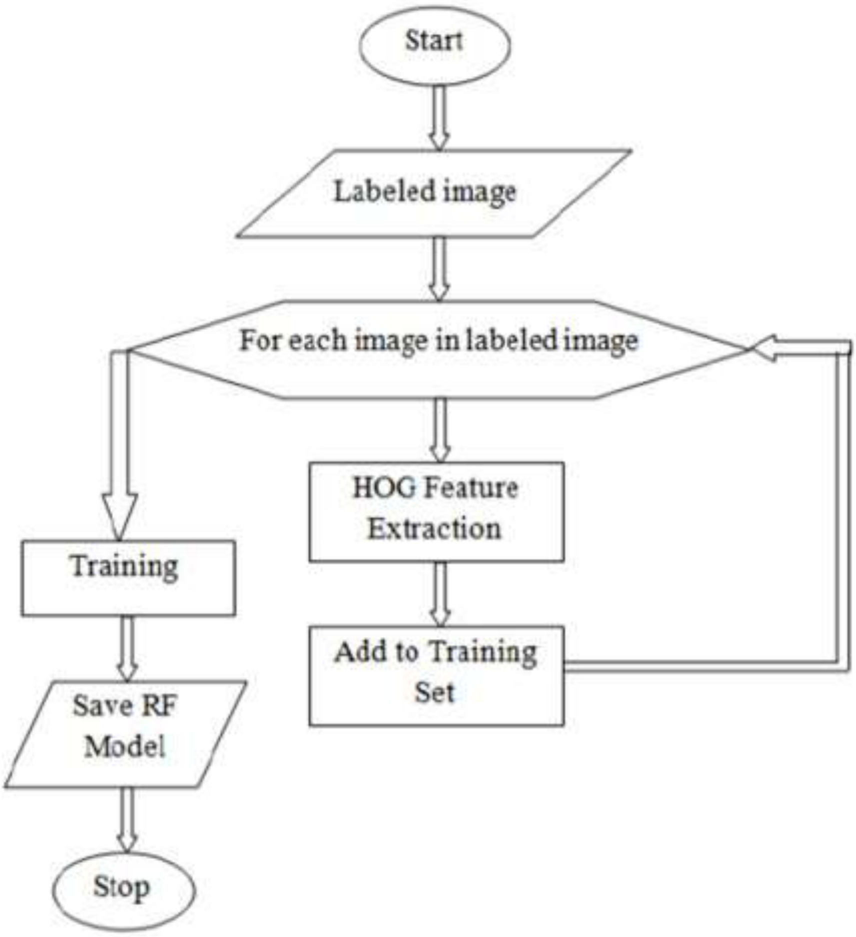


Fig.5. Flow chart for training.

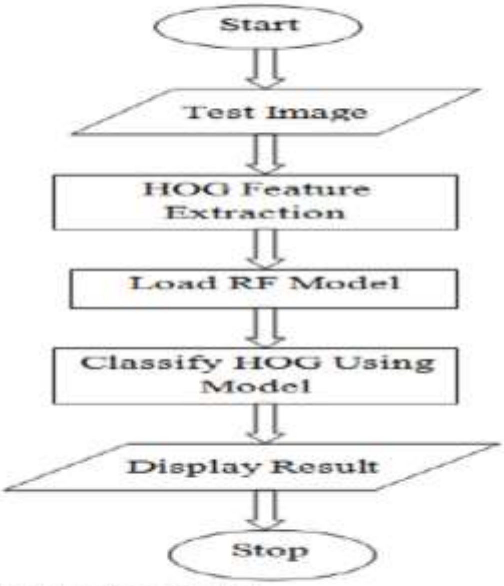


Fig.6. Flow chart for classification

V. RESULT

To calculate Hu moments and Haralick features, images are converted from RGB to grayscale. For color histogram calculation, images are converted from RGB to HSV. The Random Forest classifier then predicts whether a leaf is diseased or healthy, achieving around 70% accuracy with the potential for improvement with more extensive datasets and additional feature descriptors.

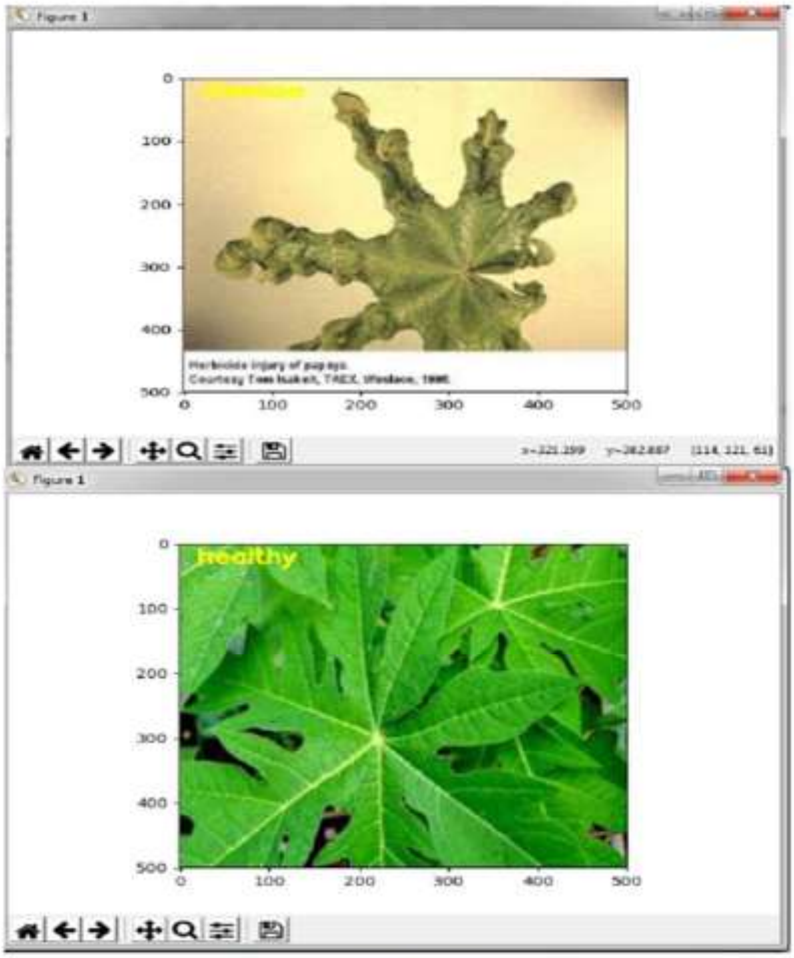


Fig.7. Final output of the classifier.

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.

REFERENCES

1. S. S. Sannakki and V. S. Rajpurohit, "Classification of Pomegranate Diseases Based on Back Propagation Neural Network," IRJET, Vol.2 Issue: 02, May-2015.
2. P. R. Rothe and R. V. Kshirsagar, "Cotton Leaf Disease Identification using Pattern Recognition Techniques," ICPC, 2015.
3. Aakanksha Rastogi et al., "Leaf Disease Detection and Grading using Computer Vision Technology & Fuzzy Logic," SPIN, 2015.
4. Godliver Owomugisha et al., "Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease," ICT in Africa, 2014.
5. Juan Tian et al., "SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," CDC, 2010.
6. S. Yun et al., "PNN based crop disease recognition with leaf image features and meteorological data," IJABE, 2015.
7. J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," Springer Plus, 2013.
8. A. Caglayan et al., "A plant recognition approach using shape and color features in leaf images," ICIAP, 2013.
9. Zhen X. et al., "Direct estimation of cardiac bi-ventricular volumes with regression forests," MICCAI, 2014.
10. Wang P. et al., "Multimodal classification of mild cognitive impairment based on partial least squares," 2016.