Plant Leaf Disease Using Machine Learning Algorithm

Samiksha Arjun Surywanshi
Department of computer engineering,
G H Raisoni college of engineering and management
Pune, India

Shivani gandhale
Department of computer engineering,
G H Raisoni college of engineering and management
Pune, India

Nutan navnathrao deshmukh
Department of computer engineering,
G H Raisoni college of engineering and management
Pune, India

Dipali santosh pandit
Department of computer engineering,
G H Raisoni college of engineering and management
Pune, India

Prof. Chhaya Nayak
Department of computer engineering,
G H Raisoni college of engineering and management
Pune, India

Abstract— Every country is having its economic source. The Indian economy depends on agricultural productivity. This technique is used to improve the performance of existing technology or to develop and design new technology for the growth of plants. The proposed agenda consists of four parts. They are Image preprocessing, Segmentation of the leaf, feature extraction, and classification of diseases. The first disease region is found using the segmentation technique, then both color and texture features are extracted. Finally, the classification technique is used to detect the type of leaf disease. The proposed system can effectively detect and categorize the examined disease with an accuracy of 80.05%.

***Keywords*—Leaf disease detection, Image segmentation, masking, feature extraction, gradient boosting classification**

I. INTRODUCTION

India is a cultivated country and about 70% of the population involved in agriculture. Farmers have a large range of diversity for selecting various suitable crops and finding suitable pesticide for the plant. The timely diagnosis of plant diseases is important as it is challenging. In the early days, the monitoring and analysis of plant diseases were done manually by the expertise person in that field. This requires a tremendous amount of work also requires excessive processing time. Although human sight and cognition are remarkably powerful in identifying and interpreting patterns, the visual assessment of plant diseases, being a subjective task, is subject to psychological and cognitive phenomena that may lead to bias, optical illusions and, ultimately, to error.But it is a very time-consuming process. To overcome this, Digital Processing Techniques can be used to detect the diseases of the plants.

Agriculture is the mother of all cultures. It has played a significant role in human life. That's why in India, agriculture is the backbone of the economy. 50% of the population is involved in farming activities directly or indirectly. Plants are everywhere we live. Lots of them carry important information for the growth of human society. Plant leaf disease is one of the key causes that reduce the superiority as well as the mass of the agricultural products. Nowadays chemicals are added to plants without knowing the capacity of the plants.

The purpose of agriculture is not only to feed the ever-growing population but also the solution for global warming and a vital source of energy [1]. Research is done in agriculture is for to aim towards the increase in productivity and food quality at a low rate and with increased profit. Common symptoms include abnormal color growth, color alteration, undersized growth, spoiled pods.

Although disease and insect pests can cause yield losses or death to the plants and indirectly to human health. These plants require cautious diagnosis and timely handling to protect the plants from heavy losses. In plants, generally, the disease detected in stem, leaves, and fruit.

Plant diseases affect production and financial losses in the agricultural industry. The main diseases of plants are viral, fungus, and bacterial disease. Environmental changes are the source of viral disease, the presence of fungus is the source of fungus disease presence of germs developed a bacterial disease in leaf or plants. The segmentation procedure is done on different appearances found in a picture such as color orientation, texture, boundaries, etc. A procedure of separating a picture into a different partition is called Image segmentation. In this paper, the segmentation of leaves is done using a Gradient Boosting Algorithm. If the image processing is applied for automatic disease detection then a low amount of effort will be needed, it will have low cost and on the other side, it will be less time consuming and will be more accurate. This paper discusses the importance of image processing techniques in the detection and classification of plant diseases in the earlier stages and hence the superiority of the product could be increased.

II. LITERATURE SURVEY

In this section, various method of image processing for the segmentation of plant disease detection is discussed. Probably the color images are having three colors generally red, green, and blue. It gives error while implementing the application using RGB because of their range 0 to 255. For that, they convert the

RGB images into the grey images. Then finally the histogram is calculated.

Monica Jhuria et. al. [3] proposed plant leaf disease detection and fruit grading system using image processing. Firstly, two separate databases created, one for the training which is of already stored images and others for the query images. The weight adjustment of training databases was done by backpropagation. As a result, they identified that the morphological features give a better result than the other two features.

Zulkifli Bin Husin et al. [4], captured the chili plant leaf image and processed it to determine the health status of the chili plant. Here the digital camera is used for image capturing and LABVIEW software tools to build the GUI. The RGB image is converted into the HSI translation. For the texture computation, the SDGM matrix is generated and using the GLCM function the features are calculated.

Shantanu Phadikar and Jaya Sil use pattern recognition techniques for the identification of rice diseases. In another approach, Chunxia Zhang, Xiuquing Wang, and Xudong Li developed a DSP and FPGA based system for monitoring and control of plant diseases.

III. PROPOSED SYSTEM

This system has five steps: Leaf Disease Image Dataset, Pre-processing, Feature extraction, Model training, Load Test Image, Gradient Boosting Model, Output Class.

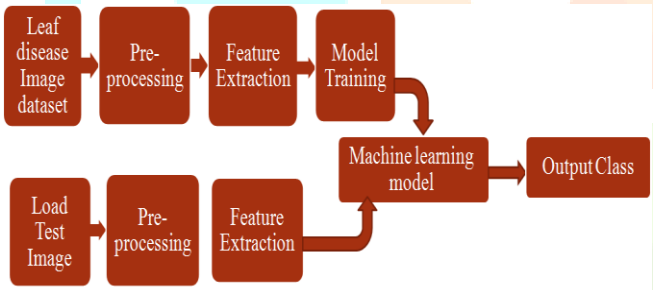


Fig. 1 Block diagram of the proposed system

A. Leaf Disease Image Database:

The database contains disease detected and healthy leaf images of apple, grapes, tomato, corn. 75% of the data is used for training purpose and the residual 25% used for testing purpose. The input image is then resized to 256*256 pixels. This database is carefully constructed because in that it usually decide the effectiveness of the classifier and performance of the proposed system.

B. Image Pre-processing:

Image pre-processing is used to improve the quality images for advance processing and analysis. The input images are first in RGB format. First RGB images are converted into greyscale form. The captured images are somewhat noisy in nature. The color transformation is done to determine the color and luminance of an image. The color space conversion is used for the enhancement of an image.

C. Image Segmentation:

Segmentation means partitioning of one image into another image. The parts having the same feature or somewhat similarity. The segmentation in this method can be done in various methods.

1) Segmentation using spot detection algorithm:

The RGB image is transformed into the HSV model for segmentation. Boundary detection and Spot detection helps to find the infected part of the leaf. For boundary detection, the 8 connectivity of pixel is checked and the boundary algorithm is applied.

2) Masking of green pixels:

Masking preserve the region of interest and remove the background and makes background pixels black. Masking means the pixel value in an image set to zero or some other value. Since the green region represents the healthy region of the leaf. When the intensity of a green pixel is greater than the previous one then values set to zero. In masking the disease portion of the leaf is identified by H and S plane value and value of '1' which is allocated to a particular portion. Rest of the regions '0'value is given. As a result, a binary image contains only ones and zeros. Thus, the diseased area of the leaf can be extracted.

3) Threshold based Segmentation:

Similarly based segmentation is adapted to segment input images on the basis of similarity in intensity or gray levels in an image. This method was usually used for light objects in a dark background or dark objects in a light background. The threshold-based algorithm which select the propoer value of threshold and segment the image into foreground and background pixels. Binary image contains only values of ones and zeros. Then this binary image is multiplied with the original RGB color image. In this way, the infected portion is extracted.

D. Feature Extraction:

Feature extraction is a technique to represent images in the feature set of the object of interest. The different types of features included in this system are color, texture, shape. The shape feature includes area, perimeter, and aspect ratio whereas color feature includes R, G, B values. The images can be better distinguished by texture feature. GLCM (Grey level co-occurrence Matrix) is generally used for the extraction of texture feature [10]. Texture classification produces a classified output of input images where each texture region is identified by different distances and different direction. The GLCM extract different features shown in fig. 1

GLCM is a statistical method which investigates the texture which considers the spatial relationship of pixels [10]. The GLCM function characterized the texture of the leaf image by calculating the occurrence of a pixel in an image with specific values and specified spatial relationship.

1) *Gray-level co-occurrence matrix (GLCM):*
It is the statistical method of identifying texture which considers some spatial relation with the pixel. The statistical features are then extracted from the matrix. The features like color, texture, morphology, homogeneity, contrast, correlation, energy are extracted from a given image. This extracted feature then give some value to the respected image. It can also be used for detection of infected plant areas.

2) *Shape Features:*
There are many features based on the geometry of the leaves. The length of the leaf is found by taking the Euclidean distance between the two points on either side of

the long axis whereas breadth corresponds to the length of the minor axis.

$$\text{Aspect ratio} = \frac{\text{Length of the leaf}}{\text{The breadth of the leaf}}$$

By using length and breadth of leaf aspect ratio is found.

- The area is calculated by initially finding the area of one pixel.

$$\text{Area} = \text{Area of one pixel} * \text{Total no. of pixels present in the leaf}$$

- The perimeter of the leaf is given by the count of pixels having the leaf margin.
- Rectangularity shows the similarity between a leaf and a rectangle.

$$\text{Rectangularity} = \frac{L*W}{A}$$

Where L is the length, W is the width and A is the area of the leaf.

E. Classification:

In this proposed work, images are classified by the machine learning algorithm. The classifier train on the training set applies it to the testing set and then measure the performance by comparing the predicted labels and give a decision.

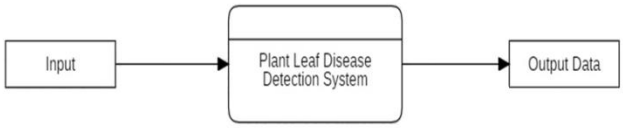
1) Gradient Boosting Algorithm:

Gradient boosting is a machine learning algorithm for regression and classification problems, which produces a prediction model in the form of ensemble typically binary trees. Gradient Boosting Algorithm supports both binary and multi-class classification. The proposed system gives better accuracy compared to other classification technique. After examining the result, the proposed system gives better accuracy with 80.02% in this system.

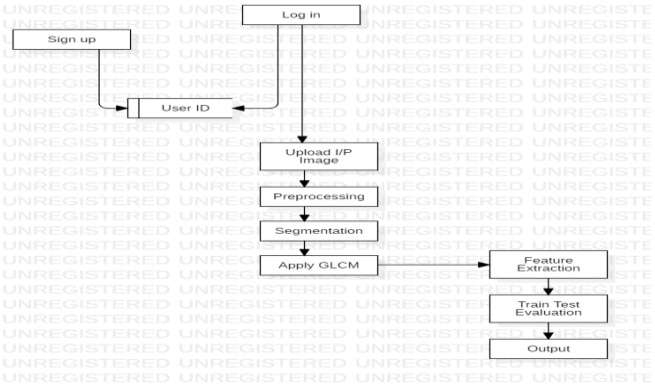
IV. Implementation

A. DFD 0

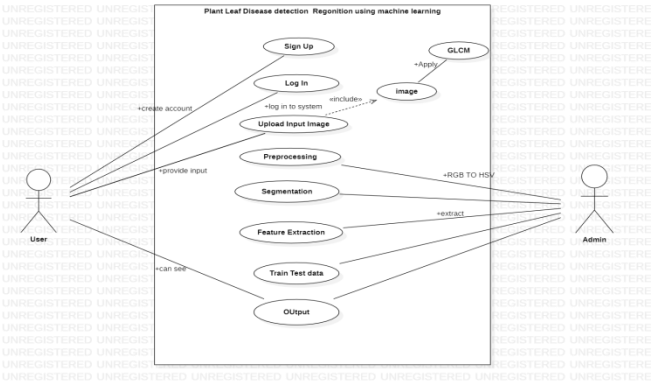
Data flow diagrams are used to graphically represent the flow of data in a business information system.



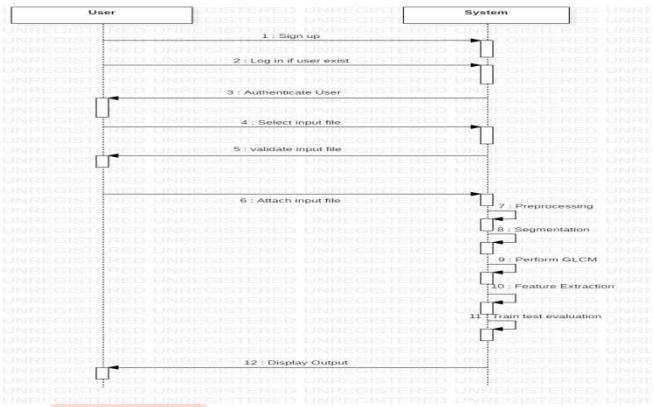
B. DFD Level 1



C. Use case Diagram



D. Sequence Diagram



V. RESULT

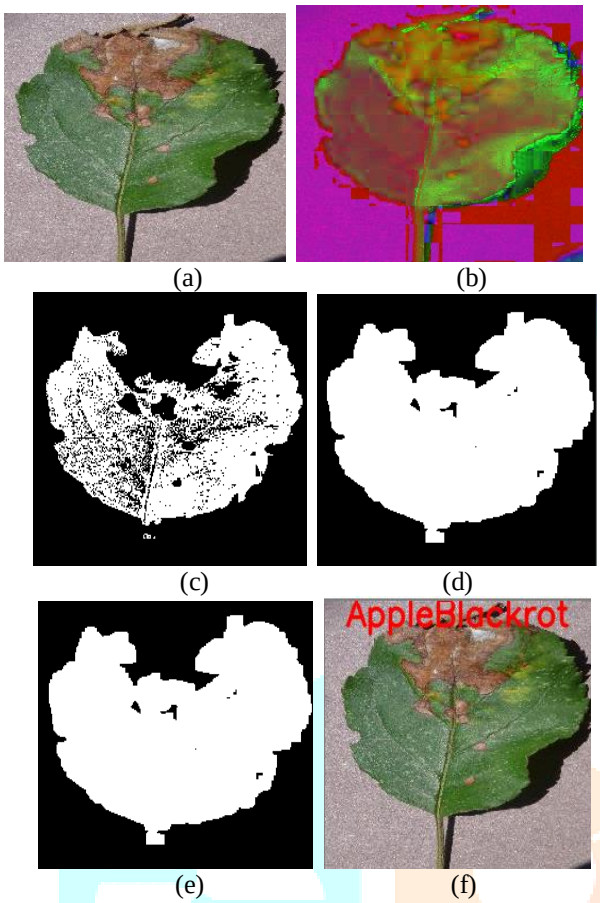
In this approach, twelve classes from 4 different plant species are considered for evaluation of performance of the system. The database is collected from the Plant Village-Database. The detailed Distribution of the database is tabulated in Table 1.

Table 1. Database Distribution

Database	Total images	Training	Testing
AppleBlackrot	621	497	124
AppleCedarapplerust	275	220	55
Applehealthy	1628	1299	329
Corn(maize) Grayleafspot	513	411	102
Corn(maize) Commonrust	1192	954	238
Corn(maize)healthy	1157	925	232
GrapeBlackrot	1180	944	236
GrapeEsca (BlackMeasles)	1383	1107	276
Grapehealthy	423	339	84
TomatoEarlyblight	1000	800	200
Tomatohealthy	1588	1270	318
TomatoLateblight	1902	1521	381

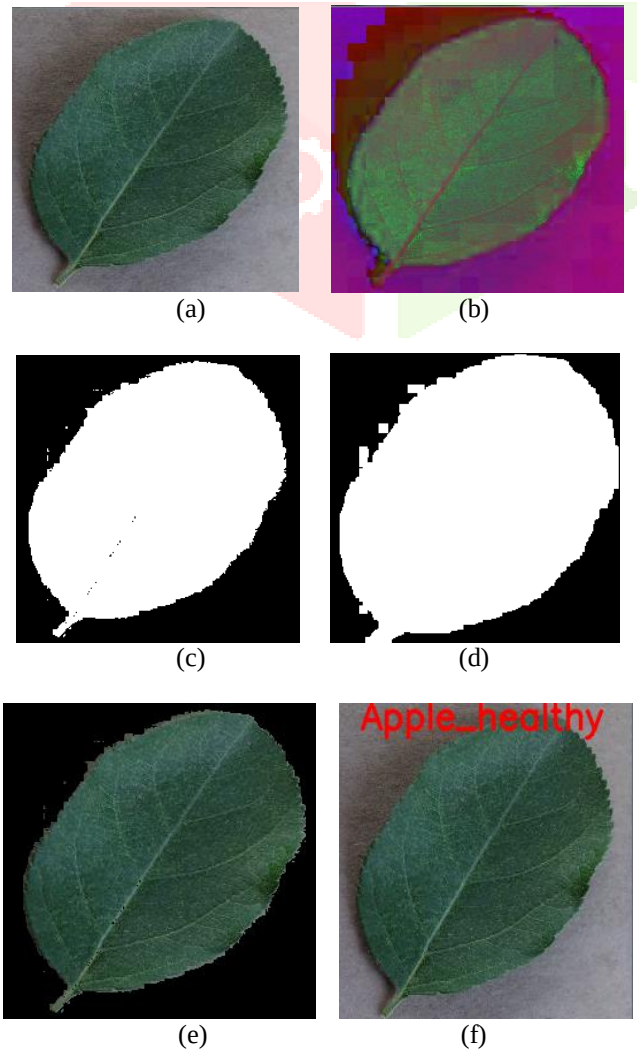
The analysis of the proposed system is performed on the basis of Qualitative and Quantitative analysis.

A. Qualitative Analysis



(a) (b) (c) (d) (e) (f)

Fig. 2) Qualitative Analysis on AppleBlackrot leaf (a) Input Image (b) HSV Image (c) mask (d) Processed Mask (e) Extracted Leaf (f) Classified Image



(a) (b) (c) (d) (e) (f)

Fig. 3) Qualitative Analysis on Apple Healthy leaf (a) Input Image (b) HSV Image (c) mask (d) Processed Mask (e) Extracted Leaf (f) Classified Image

B. Quantitative Analysis

Qualitative analysis was perform using different machine learning classifiers.

TABLE 1: Quantitative Analysis of proposed system

Classifier	Parameters	Combined
SVM	RBF	14.46
	Polynomial	14.31
	Sigmoid	14.31
Gaussian NB		44.71
Decision Tree		69.89
KNN	3	38.58
	5	41.44
	7	41.97
stochastic gradient descent (SGD)		21.64
Gradient Boosting		80.02
Multi Layer Perceptron		25.3
Logistic Regression		43.71
k-means Clustering		1.91
Adaboost Classifier		31.63

VI.CONCLUSION

This paper represents a review of the technical implementations in the research area of plant disease detection using image processing techniques. From the literature, it is evident that color, texture and morphological features are most suitable to identify and classify the diseases in plants. Most commonly used classification techniques for classifying plant diseases are artificial neural networks (ANN) and support vector machine (SVM). Automatic detection of plant diseases would solve the problem of an expensive domain expert. Detection of plant diseases in the early stage would help farmers to improve the crop yield, which in turn improves Indian gross domestic product (GDP).

REFERENCES

[1]. Wenjiang Haung, Quigsong Guan, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Dong Liang, Linsheng Huang, and Linsheng Huang, and Dongyan Zhang, "New Optimized Spectral Indices for identifying and Monitoring Winter Wheat Diseases", IEEE Journal of selected topics in applied earth observation and remote sensing, vol. 7, No. 6, June 2014.

[2]. R. Meena Prakash, G. P. Sarasathy, G. Ramalakshmi, K. H. Mangleshwari, T. Kaviya, "Detection of leaf Diseases and Classification using Digital Image Processing", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

[3]. Monica Jhuria, Ashwani Kumar, and RushikeshBorse, "Image Processing and Fruit Grading", Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013).

[4]. Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook, "Feasibility study on plants chili diseases using image processing Techniques", 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

[5]. Saradhambal G1, Dhivya.R2, Latha.S3, R. Rajesh4, "Plant Diseases Detection and its Solution using Image Classification", International Journal of Pure and Applied Mathematics Volume 119 No. 14 2018, 879-884.

[6]. H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. L Rahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Application (0975-8887) Volume 17-No. 1, March 2011.

[7]. Chunxia Zhang, Xiuquing Wang, Xudong Li, "Design of Monitoring and control Plant Disease System Based on DSP & FPGA", 2010 Second International Conference on Network Security.

[8]. D Venkatraman, Mangayarkarasi N, "Computer Vision-Based Feature Extraction of leaves for Identification of Medicinal Values of Plants", 2016 IEEE International Conference on Computational Intelligence and Computing Research.

[9]. Jitesh P. Shah, Harshadkumar B. Prajapati, Vipul K. Dabhi, "A Survey on Detection and Classification of Rice Plant Diseases", 2016 IEEE.

[10]. Fritz Albregtsen, "Statistical Texture Measure Computed from Gray Level Co-occurrence Matrices", Image Processing Laboratory Department of Informatics University of Oslo November 5, 2008

[11]. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610–621, 1973.