

Leaf Diseases Detection of Medicinal Plants based on Image Processing and Machine Learning Processes

Payal Bose¹, Prof. Shawni Dutta², Prof. Vishal Goyal³ and Prof. Samir K. Bandyopadhyay^{4,*}

¹Research Scholar, Lincoln University College, Kota Bharu, Kelantan, Malaysia

²Department of Computer Science, The Bhawanipur Education Society College, Kolkata, India

³GLA University, Mathura-Delhi Road Mathura, Chaumuhan, Uttar Pradesh, India

⁴Distinguish Professor, Lincoln University College, Kota Bharu, Kelantan, Malaysia

* Corresponding author. (e-mail address:1954samir@gmail.com)

Abstract: On earth, plants play the most important part. Every organ of a plant plays a vital role in the ecological field as well as the medicinal field. But on the whole earth there are several species of plants are available. Different plants have different diseases. Therefore it is needed to identify the plants and their diseases to prevent loss. Now to identify the plants and their diseases manually is very time consuming. In this research an automatic plant and their disease detection system is proposed. For experimental purposes, high-quality leaf images are accepted for training and testing. For detecting the healthy and diseased area in a leaf, region-based and color-based region thresholding techniques were used. For feature selection Histogram Oriented Gradient (HOG) and Local Binary Pattern (LBP) method were applied. Finally for classification two-class and multi-class Support Vector Machine (SVM) was used. It is observed that both feature selection processes with SVM give 99% accuracy. Finally to understand the automated system a graphical user interface was created for all users.

Keywords: Image Processing, Automated Plant Diseases Detection, Histogram Oriented Gradient (HOG), Local Binary Pattern (LBP), Support Vector Machine (SVM)

I. INTRODUCTION

Among all the living things in this world, plants are one of the most important things. Horticulture, a part of agriculture used to deal with gardening of useful plants. Horticulture plants like flowers, fruits, vegetables are used to fulfill daily life necessities. So it is necessary to have complete knowledge about these plants as well as the diseases affecting these plants. Detection of plant diseases in the early stages can help to prevent leaf losses and poor production. Detecting plant diseases manually is a time-consuming process, but with the developing technology, today is quite easy to monitor and detect the plant diseases. Computer Vision technology is much more suitable and gives more accurate results in these cases. In this automatic plant diseases system image segmentation based on region-based thresholding and color based thresholding techniques are used. Based on these techniques it is much easier to detect the disease-affected and healthy leaf area. Finally a classification algorithm is used to classify all the diseased and healthy leaf categories.

II. LITERATURE SURVEY

Different researchers have done different research to detect various plant diseases based on different techniques. Traditionally the researchers used spectroscopic and imaging techniques to detect the diseases. But with the development of artificial intelligence, machine learning and deep learning based on computer vision study gives more than satisfying results. These techniques are widely used to detect plant and leaf diseases automatically.

In a research paper, the authors proposed a real-time grape disease detection process based on an improved convolution neural network [1]. First, they used digital image processing techniques to create a grape leaf disease database. Second, they apply a faster-RCNN and a deep learning-based faster DR-IACNN model to detect the disease. They have also shown that the faster DR-IACNN model achieved an 81.1% precision rate and detection speed reaches up to 15.01 FPS.

In another research paper, the authors used a convolution neural network with an image processing technique to detect plant diseases [2]. They mostly used common plants that are available all over the world specifically in Iraq. They classified 15 classes, among them 12 classes were diseased and 3 were healthy. They obtained more than 98% accuracy for both the training and testing dataset.

For all nations, economic growth mostly depends on agricultural products. But due to the various plants' diseases, the growth and productivity are decreased from time to time. In the early stage of diagnosis of these diseases can help to prevent the spreading and helps to better productivity. The authors survey various kinds of classification models based on image processing and machine learning techniques to detect and recognized the diseases of agricultural fields. They also show the proposed method and its accuracy for detailed analysis [3].

In another research paper, the authors also survey a detailed analysis of different classification techniques to detect the diseases of various agricultural plants [4].

To detect the apple leaf disease authors proposed a deep neural base improved convolution neural network model. They create an apple leaf diseases dataset based on the laboratory and real-life complex images. Then to detect the diseases they introduce GoogLeNet interception and Rainbow concatenation with a deep neural network. They used this model to train up to five apple leaf diseases. The detection performance is 78.80% with high detection speed is 23.13 FPS [5].

To identify the plant leaf diseases the author proposed a model based on image pre-processing and machine learning approaches. They use different image processing techniques to pre-process the image and the GLCM feature extraction method to extract the features of the leaf. Finally, apply the K Nearest Neighbour classifier to detect the diseases of the leaf. The proposed implementation predicts the 98% accuracy for disease detection [6].

To identify automatic crop diseases the authors survey 19 studies based on Convolution Neural Network (CNN). In this study, the authors describe diseases profile, implementation techniques of CNN, and analysis of the performance of the techniques. Finally, they provide the guidelines for an improved CNN for future research [7]. In another paper, the authors' overview different plant leaf disease detection based

on different machine learning classification techniques. They describe different algorithms and their performance [8].

III. PROPOSED METHODOLOGY

In this experiment to detect the leaf diseases from an input image, the following steps are involved. Figure-1 shows the overview structure of the proposed leaf diseases system.

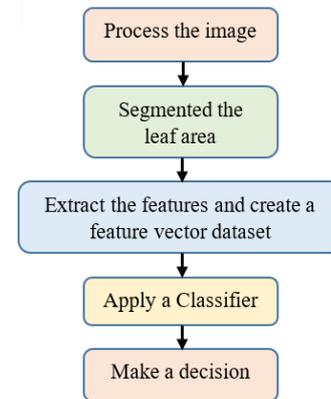


Figure-1: Overview Structure of the Proposed System

1. PROCESSING THE IMAGES

a. Database Details

The complete set of images consists of 12 plants they are named as “Mango”, “Arjun”, “Saptaporni (Alstonia Scholaris)”, “Guava”, “Bael”, “Jamun”, “Jatropha”, “Pongamia Pinnata”, “Basil”, “Pomegranate”, “Lemon”, and “Chinar”. All the plants have their own economical and environmental values. The entire database is divided among two main classes “Healthy” and “Diseased” and 22 subclasses. Figure 2 and 3 shows a detailed view of the database images. This database has a total of 4503 images, among them, 2278 images are for healthy leaf images and 2225 are for diseased images. All the images are collected from the Shri Mata Vaishno Devi University, Katra [9-10].

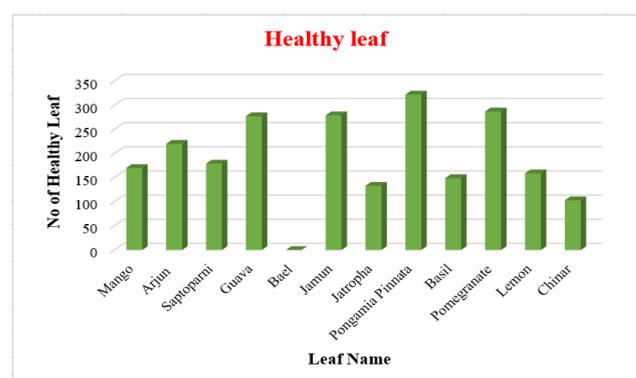


Figure-2: Healthy Leaf details in Leaf database

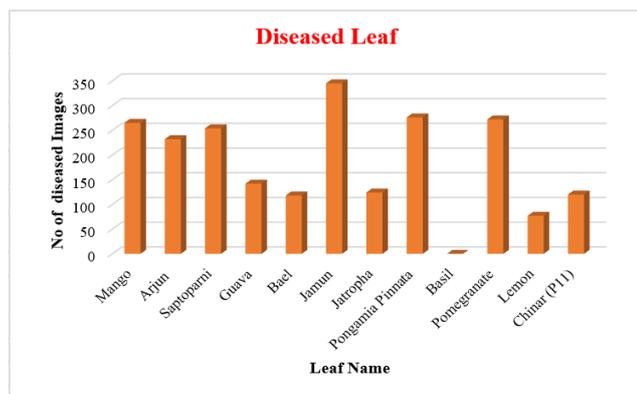


Figure-3: Diseased Leaf details in Leaf database

b. Pre-processing the images

Pre-Processing [11] the images means processing the images before the computation processes. This process commonly involves eliminating the background from the objects, reduces the background noises, resize the input images, etc. In this experiment, the images in the leaf database are very high in quality [dimension 6000×4000]. All the images are RGB in color and taken with a high-quality digital camera. Therefore for experimental usage, here the images reduce to dimensions 600×400 for display purposes and dimensions 150×100 for feature selection purposes. Figure 4 shows the sample leaf example.

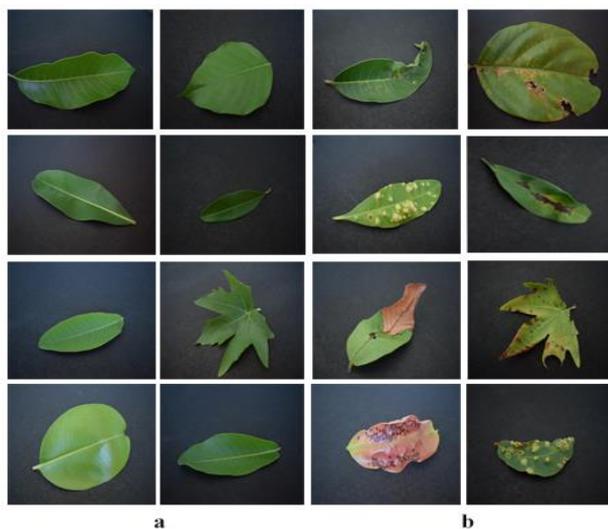


Figure-4: The Sample Leaf (a) Healthy Leaf, (b) Diseased Leaf

2. SEGMENTATION

Image segmentation [12-13] is one of the most important parts of image processing. The criteria of this process are to divide the input image into the same type of area and extract the region of interest. Threshold segmentation is a common

and simple image segmentation technique. It is a region-based segmentation technique. This method can be divided into two categories 1) Global thresholding and 2) Local thresholding. In the Global thresholding technique, the input image is divided into two main regions 1) background and 2) the region of interest. The global thresholding method uses a single thresholding level to segment the input image. The Local thresholding technique uses multiple thresholding levels to divide the input region and background.

In this experiment, the global thresholding technique is used. For segmentation, a global optimum threshold value is selected. In the leaf image database, the background of the input images is much darker therefore the segmentation effect is more effective. This procedure is faster and the calculation is much simple. Figure 5 shows the segmentation result of the input image.

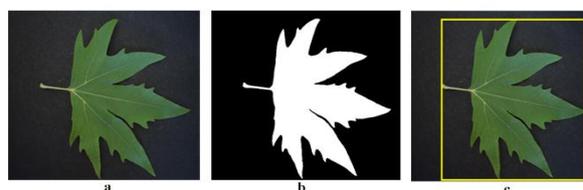


Figure-5: a) Original Image, b) Segmented Image, c) Region of Interest

For detecting the diseased affected area in this experiment a color-based region segmentation technique is used. To segment the green color region from the diseased defected regions first convert the input image into Lab color space. Lab color space is a 3-axis color channel. It splits the RGB color image into one lightness (L) and two chromaticity layers (a, b). After converting into Lab color space, equation 1 is applied to segment the green leaf and diseased affected area.

$$RGB_{segmentation} = [(0 \leq R \leq 80) \& (-55 \leq G \leq -10) \& (5 \leq B \leq 10)]$$

Figure 6 shows the segmentation of diseased and green leaf area.

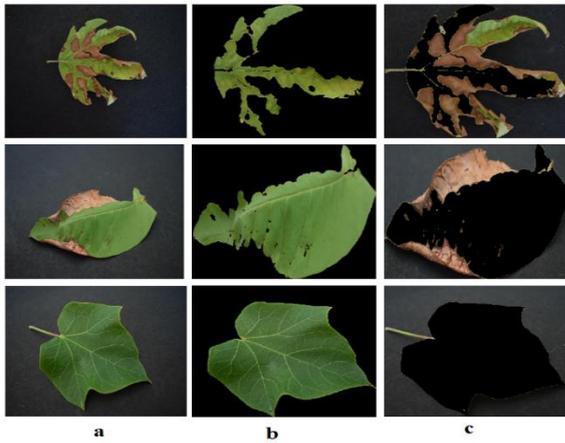


Figure-6: a) Original Image, b) Green Leaf area, c) Disease affected area

3. FEATURE SELECTION AND EXTRACTION

Feature selection is a technique in image processing. This technique helps to reduce a large set of features by selecting the best features from the original dataset. It abandons the redundant data from the original dataset. In computer vision object classification or object recognition is one of the most popular subjects. The aim of the object classification is to extract features from the input image and use a proper classifier to classify the feature class label Histogram Oriented Gradient (HOG) and Linear Binary Pattern (LBP) are the two efficient gradient-based feature selection techniques [14]. Their performance is much better than any other feature selection technique.

a. Histogram Oriented Gradient (HOG)

It is a gradient-based feature descriptor [15]. It calculates the existing gradient and orientation of the input image. Basically, this method broke the input image into a small piece of regions, then calculate the vertical and horizontal gradients of those blocks locally. Finally calculate the magnitude and orientations of those gradients.

b. Linear Binary Pattern (LBP)

It is a gradient-based effective texture descriptor [15]. This method divided the input image pixels into 3×3 matrices. Then it considers the central value of 3×3 matrices as a threshold value and set a binary value based on the threshold value. This process is continued for all pixel matrix and finally, it converts all the binary numbers into decimal and represents the input image in a better way.

4. CLASSIFICATION

In machine learning classification [16-17] is a process, used to understanding, recognizing, and grouping the same type of data based on their categories. The classification algorithms are used a pre-trained training dataset to predict the category of an unknown sample whose data fall into the predetermined categories.

For classification of an image support vector machine is one of the most popular and best classification method. It is a supervised learning method. This algorithm is divided the whole data space with a maximum margin to predict the class of the unknown data sample. Figure 6 shows the classification model of this experiment.

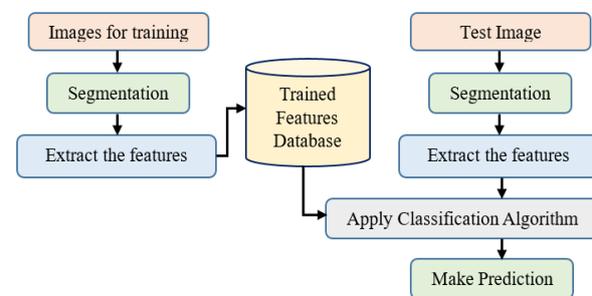


Figure-6: Classification Model

IV. EXPERIMENTAL RESULT AND DISCUSSIONS

The experimental database has 10 leaf classes and two main classes 'Healthy' and 'Diseased'. Table 1 shows the details about the experimental database.

Table-1: Database Details

Name of the plant	Healthy Leaf (No of images)	Diseased Leaf (No of images)	Disease Name
Mango	170	265	Anthracoese
Arjun	220	232	Leaf Spot
Saptoparni	179	254	Foliar Galls
Guava	277	142	Fungal Disease
Jamun	279	345	Fungal Disease
Jatropha	133	124	Leaf Spot
Pongamia Pinnata	322	276	Chlorotic lesions
Pomegranate	287	272	Cercospora spot
Lemon	159	77	Citrus Canker
Chinar	103	120	Leaf Spot

Figure 7a and 7b show the leaf disease detection and classification process.

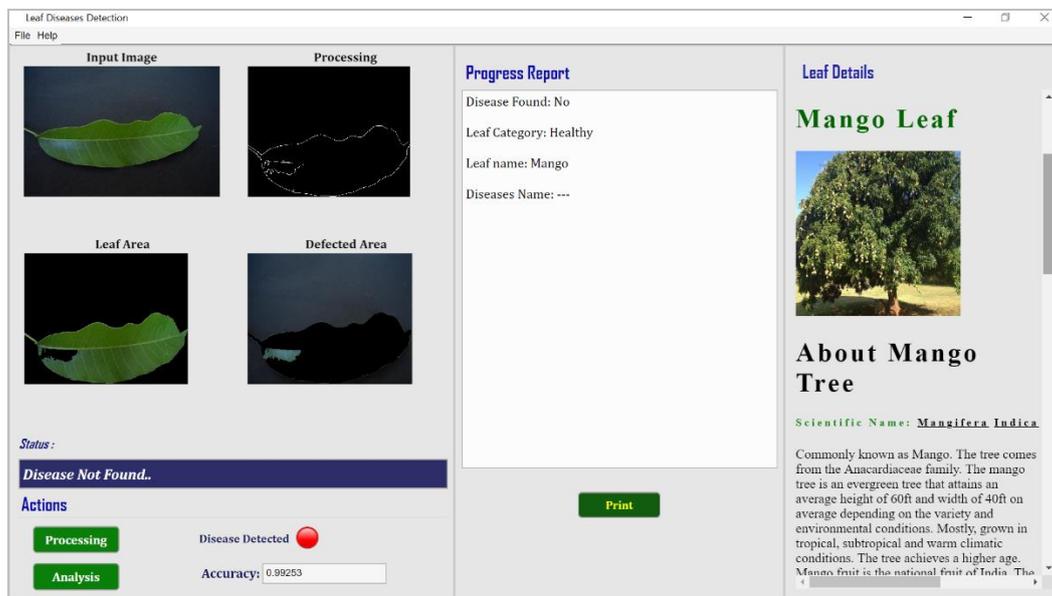


Figure-7a: Leaf diseases Detection and Classification (Healthy Leaf)

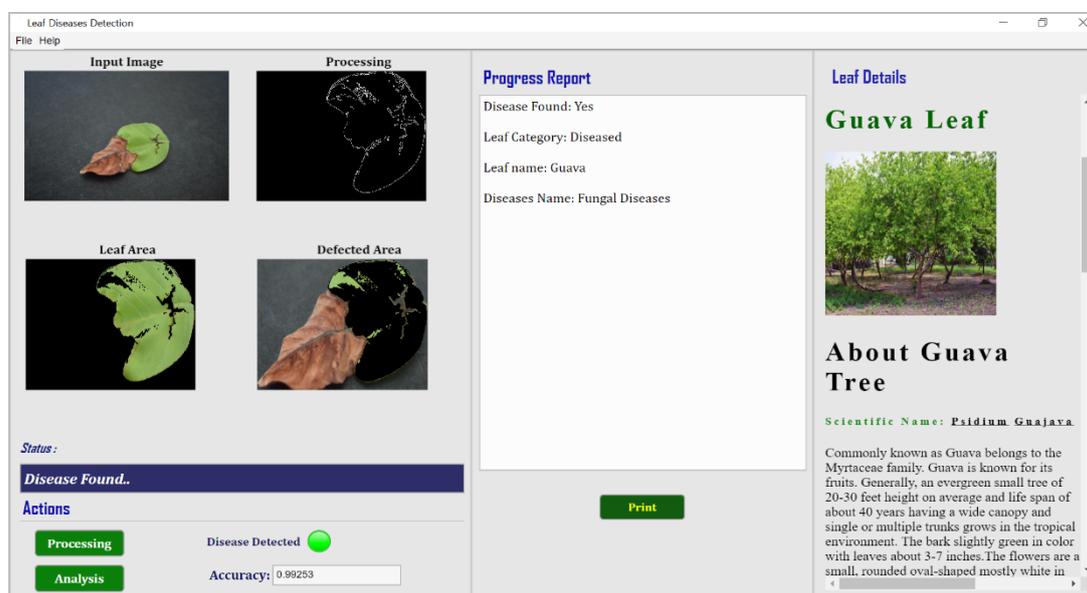


Figure-7b: Leaf diseases Detection and Classification (Diseased Leaf)

Table 2 shows the performance analysis of different classifiers with HOG and LBP features extraction.

Table-2: Performance Analysis

Using Histogram Oriented Gradient (HOG) Features			
Methods	Accuracy	Specificity	Sensitivity
SVM (Proposed Method)	0.99175	0.99447	0.98905
Decision Tree	0.98113	0.98419	0.97811
Linear Analysis	0.89623	0.91621	0.87647
Discriminant Analysis	0.98192	0.97866	0.98514

Using Linear Binary Pattern(LBP) Features			
Methods	Accuracy	Specificity	Sensitivity
SVM (Proposed Method)	0.98146	0.97695	0.99592
Decision Tree	0.98045	0.97695	0.98391
Linear Analysis	0.85305	0.95525	0.75201
Discriminant Analysis	0.96562	0.95932	0.97185

Figures 8a and 8b show the performance details of each main class and the subclasses using the HOG feature selection method. Figures 9a and 9b show the performance details of each main class and the subclasses using the LBP feature selection method.

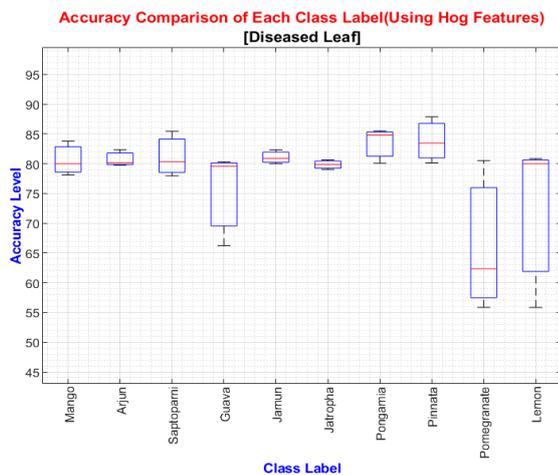


Figure-8a: Performance details each diseased leaf classes (Using HOG Features)

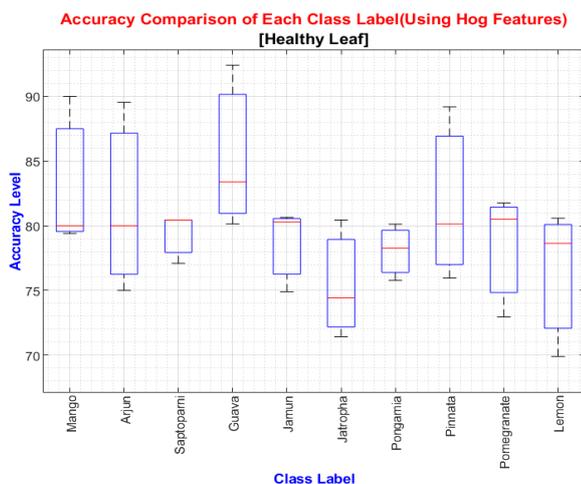


Figure-8b: Performance details each healthy leaf classes (Using HOG Features)

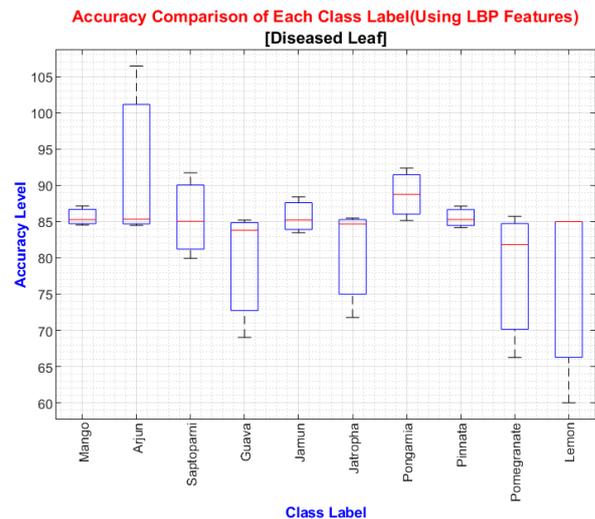


Figure-9a: Performance details each diseased leaf classes (Using LBP features)

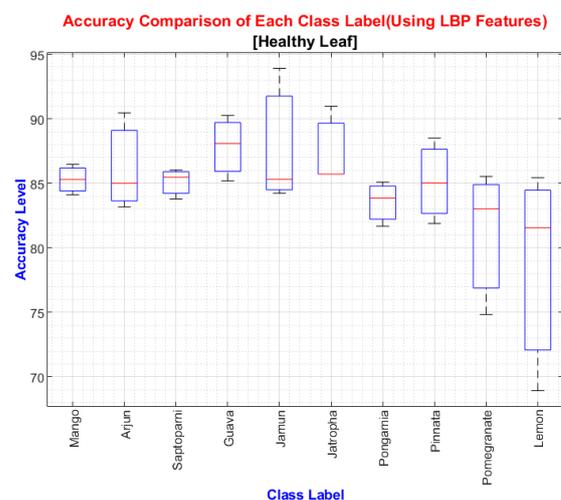


Figure-9b: Performance details each healthy leaf classes (Using LBP features)

V. CONCLUSION

In this paper, an automated leaf detection system is proposed. To detect the healthy and diseased leaf image region-based thresholding technique is used. Again to detect the particular diseased area of a diseased leaf color-based region thresholding method was used. For feature selection from the input images both HOG and LBP feature selection technique is used. To classify the category healthy and diseased leaf with subclasses, leaf name, and disease name the two-class, and multi-class SVM classifier is used. A detailed performance analysis was done between main classes and subclasses by using different classifiers. Finally, a graphical user interface is created for all users.

REFERENCES

- [1] Xie, X., Ma, Y., Liu, B., He, J., Li, S., & Wang, H. (2020). A Deep-Learning-Based Real-Time Detector for Grape Leaf Diseases Using Improved Convolutional Neural Networks. *Frontiers in Plant Science*, 11(June), 1–14. <https://doi.org/10.3389/fpls.2020.00751>
- [2] Jasim, M. A., & Al-Tuwaijari, J. M. (2020). Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques. *Proceedings of the 2020 International Conference on Computer Science and Software Engineering, CSASE 2020*, 259–265. <https://doi.org/10.1109/CSASE48920.2020.9142097>
- [3] Swain, S. (2020). A Review on Plant Leaf Diseases Detection and Classification Based on Machine Learning Models Page No : 5195. IX(Vi), 5195–5205.
- [4] Santhosh Kumar, S., & Raghavendra, B. K. (2019). Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 313–316. <https://doi.org/10.1109/ICACCS.2019.8728325>
- [5] [5] Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access*, 7, 59069–59080. <https://doi.org/10.1109/ACCESS.2019.2914929>
- [6] Tulshan, A. S., & Raul, N. (2019). Plant Leaf Disease Detection using Machine Learning. *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*, 1–6. <https://doi.org/10.1109/ICCCNT45670.2019.8944556>
- [7] Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Frontiers in Plant Science*, 10(July). <https://doi.org/10.3389/fpls.2019.00941>
- [8] Sherly Puspha Annabel, L., Annapoorani, T., & Deepalakshmi, P. (2019). Machine learning for plant leaf disease detection and classification - A review. *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, 538–542. <https://doi.org/10.1109/ICCSP.2019.8698004>
- [9] Chouhan, S. S., Singh, U. P., Kaul, A., & Jain, S. (2019). A Data Repository of Leaf Images: Practice towards Plant Conservation with Plant Pathology. *2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019*, 700–707. <https://doi.org/10.1109/ISCON47742.2019.9036158>
- [10] CHOUHAN, Siddharth Singh; Kaul, Ajay; SINGH, UDAY PRATAP; & Science, Madhav Institute of Technology (2020), “A Database of Leaf Images: Practice towards Plant Conservation with Plant Pathology”, Mendeley Data, V4, doi: 10.17632/hb74ynkjen.4
- [11] Koprowski, R. (2017). Image pre-processing. *Studies in Computational Intelligence*, 682, 21–38. https://doi.org/10.1007/978-3-319-50490-2_3
- [12] Zaitoun, N. M., & Aqel, M. J. (2015). Survey on Image Segmentation Techniques. *Procedia Computer Science*, 65(December), 797–806. <https://doi.org/10.1016/j.procs.2015.09.027>
- [13] Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), 41–49. <https://doi.org/10.1016/j.inpa.2016.10.005>
- [14] Greeshma, K. V., College, C., & Gripsy, J. V. (2020). Image Classification using HOG and LBP Feature Descriptors with SVM and CNN. 8(04), 4–7.
- [15] Islam, M. A., Billah, M., & Yousuf, S. I. (2019). Automatic Plant Detection Using HOG and LBP Features With SVM. *International Journal of Computer (IJC)*, 33(1), 26–38. <http://ijcjournal.org/>
- [16] Oo, Y. M., & Htun, N. C. (2018). Plant Leaf Disease Detection and Classification using Image Processing. *International Journal of Research and Engineering*, 5(9), 516–523. <https://doi.org/10.21276/ijre.2018.5.9.4>
- [17] Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019). A Review on Machine Learning Classification Techniques for Plant Disease Detection. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 281–284. <https://doi.org/10.1109/ICACCS.2019.8728415>