School of Electronics and Communication Engineering

Academic Year 2021 - 2022

T. Y. Mini Project - I
Trimester VIII

Report

Title of the Project: Leaf Disease Detection using ML

Domain of Project: Machine learning and Image Processing

Guide Name: Mrs. Vibha Patel

Batch Coordinator Name: Mrs. Trushita Chaware

Name of Students:

Sr. No	PRN No.	Name of Student	Contact No.	Email ID	Basket
1.	1032190782	Shreyash Butley	9325750600	shreyashbutley888pb@gmail.com	AI
2.	1032190814	Govind Varma	932522105	govindvarmaagroya105@gmail.com	Comm
3.	1032191036	Soham Vaidya	9284415508	vsovaidya12@gmail.com	AI

School of Electronics and Communication Engineering

MIT World Peace University

Pune

CERTIFICATE

This is to certify that the B. Tech. Mini Project entitled

Leaf Disease Detection

work has been carried out successfully
by

Shreyash Butley 1032190782
Govind Varma 1032190814
Soham Vaidya 1032191036

during the Academic Year 2021 – 2022 in partial fulfilment of their course of Mini Project for Third Year Electronics and Communication Engineering as per the guidelines prescribed by the MIT World Peace University, Pune

Project Guide	Batch Coordinator	Mini Project Coordinator	Head of School
Prof. Vibha Patel	Dr. Trushita S. Chaware	Dr. Trushita S. Chaware	Dr. VinayaGohokar

MIT World Peace University, Pune

School of ECE

Project Report Contents

Abstract:

Identification of the plant diseases is the key to prevent the losses in the yield and quantity of the agricultural product. The studies of the plant diseases mean the studies of visually observable patterns seen on the plant. Health monitoring and disease detection on plants is very critical for sustainable agriculture. It is very difficult to monitor the plant diseases manually.
Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

Agricultural productivity is something on which the economy highly depends. This is one of the reasons that disease detection in plants plays an important role in the agriculture field, as having disease in plants is quite natural. If proper care is not taken in this area, then it causes serious effects on plants and due to which respective product quality, quantity or productivity is affected. For instance, a disease named little leaf disease is a hazardous disease found in pine trees in the United States. Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops, and at a very early stage itself it detects the symptoms of diseases i.e., when they appear on plant leaves. This paper presents an algorithm for image segmentation technique which is used for automatic detection and classification of plant leaf diseases. It also covers surveys on different disease classification techniques that can be used for plant leaf disease detection. Image segmentation, which is an important aspect for disease detection in plant leaf disease, is done by using genetic algorithms.

Introduction:

India is a fast-developing country and agriculture is the backbone for the country's development in the early stages. Due to industrialization and globalization concepts the field is facing hurdles. On top of that the awareness and the necessity of the cultivation need to be instilled in the minds of the younger generation. Nowadays technology plays a vital role in all the fields but till today we are using some old methodologies in agriculture.

Modern technologies have given human society the ability to produce enough food to meet the demand of more than 7 billion people. However, food security remains threatened by a number of factors including climate change , the decline in pollinators (Report of the Plenary of the Intergovernmental Science-Policy Platform on Biodiversity Ecosystem and Services on the work of fourth session, 2016), plant diseases (Strange and Scott, 2005), and others. Plant diseases are not only a threat to food security at the global scale, but can also have disastrous consequences for smallholder farmers whose livelihoods depend on healthy crops. In the developing world, more than 80 percent of the agricultural production is generated by smallholder farmers (UNEP, 2013), and reports of yield loss of more than 50% due to pests and diseases are common (Harvey et al., 2014). Furthermore, the largest fraction of hungry people (50%) live in smallholder farming households (Sanchez and Swaminathan, 2005), making smallholder farmers a group that's particularly vulnerable to pathogen-derived disruptions in food supply.

Various efforts have been developed to prevent crop loss due to diseases. Historical approaches of widespread application of pesticides have in the past decade increasingly been supplemented by integrated pest management (IPM) approaches (Ehler, 2006). Independent of the approach, identifying a disease correctly when it first appears is a crucial step for efficient disease management. Historically, disease identification has been supported by agricultural extension organizations or other institutions, such as local plant clinics. In more recent times, such efforts have additionally been supported by providing information for disease diagnosis online, leveraging the increasing Internet penetration worldwide. Even more recently, tools based on mobile phones have proliferated, taking advantage of the historically unparalleled rapid uptake of mobile phone technology in all parts of the world (ITU, 2015).

Smartphones in particular offer very novel approaches to help identify diseases because of their computing power, high-resolution displays, and extensive built-in sets of accessories, such as advanced HD cameras. It is widely estimated that there will be between 5 and 6 billion smartphones on the globe by 2020. At the end of 2015, already 69% of the world's population had access to mobile broadband coverage, and mobile broadband penetration reached 47% in 2015, a 12-fold increase since 2007 (ITU, 2015). The combined factors of widespread smartphone penetration, HD cameras, and high-performance processors in mobile devices lead to a situation where disease diagnosis based on automated image recognition, if technically feasible, can be made available at an unprecedented scale. Here, we demonstrate the technical feasibility using a machine learning approach utilizing 10,000 images of 14 crop species with 26 diseases (or healthy) made openly available through the project Plant Village [7].

Identifying plant disease wrongly leads to huge loss of yield, time, money, and quality of product. Identifying the condition of the plant plays an important role for successful cultivation. In olden days identification is done manually by the experienced people but due to the so many environmental changes the prediction is becoming tough. So, we can use image processing techniques for identification of plant disease.

Generally, we can observe the symptoms of disease on leaves, stems, flowers etc. So here we use leaves for identification of disease affected plants. We are using ML algorithms for identifying the disease and we have used both supervised and unsupervised algorithms like K-means and SVM (Support Vector Machine) respectively.

Dataset:

We analyse 54,306 images of plant leaves, which have a spread of 38 class labels assigned to them. Each class label is a crop-disease pair, and we make an attempt to predict the crop-disease pair given just the image of the plant leaf. In all the approaches described in this paper, we resize the images to 256 × 256 pixels, and we perform both the model optimization and predictions on these downscaled images.

Across all our experiments, we use three different versions of the whole PlantVillage dataset. We start with the PlantVillage dataset as it is, in color; then we experiment with a gray-scaled version of the PlantVillage dataset, and finally we run all the experiments on a version of the PlantVillage dataset where the leaves were segmented, hence removing all the extra background information which might have the potential to introduce some inherent bias in the dataset due to the regularized process of data collection in case of PlantVillage dataset. Segmentation was automated by the means of a script tuned to perform well on our particular dataset. We chose a technique based on a set of masks generated by analysis of the color, lightness and saturation components of different parts of the images in several color spaces (Lab and HSB). One of the steps of that processing also allowed us to easily fix color casts, which happened to be very strong in some of the subsets of the dataset, thus removing another potential bias.

Objective/Aim:

- To detect the disease in plants, for that we have used different methodologies.
- With image processing applications we have converted and derived meaningful information from leaves.
- With the help of algorithms such as K-means, we have identified the disease part of the corresponding leaf and separated the disease from the rest of the leaf with the help of clustering.
- SVM is used to gain the accuracy of the model.

Literature Survey:

A new automatic method for disease symptom segmentation in digital photographs of plant leaves. The diseases of different plant species have been mentioned. Classification is done for a few of the disease names in this system. In this section, various methods of image processing for plant disease detection are discussed. The vegetation indices from hyper spectral data have been shown for indirect monitoring of plant diseases. But they cannot distinguish different diseases on crops. Wenjiang Huang et al developed the new spectral indices for identifying the winter wheat disease. They consider three different pests (Powdery mildew, yellow rust, and aphids) in winter wheat for their study. The most and the least relevant wavelengths for different diseases were extracted using the RELIEF-F algorithm. The classification accuracies of these new indices for healthy and infected leaves with powdery mildew, yellow rust and aphids were 86.5%, 85.2%, 91.6% and 93.5% respectively [1].

Enhanced images have higher quality and clarity than the original image. Color images have primary colors red, green and blue. It is difficult to implement the applications using RGB because of their range i.e., 0 to 255. Hence, they convert the RGB images into the grey images. Then the histogram equalization which distributes the intensities of the images is applied on the image to enhance the plant disease images.

Monica Jhuria et al uses image processing for detection of disease and the fruit grading in [3]. They have used artificial neural networks for detection of disease. They have created two separate databases, one for the training of already stored disease images and other for the implementation of the query images. Back propagation is used for the weight adjustment of training databases. They consider three feature vectors, namely, color, textures, and morphology [3]. They have found that the morphological feature gives better results than the other two features. Zulkifli Bin Husin et al, in their paper [4], they captured the chilli plant leaf image and processed it to determine the health status of the chilli plant. Their technique is ensuring that the chemicals should apply to the diseased chilli plant only.

They used the MATLAB for feature extraction and image recognition. In this paper preprocessing is done using the Fourier filtering, edge detection and morphological operations. Computer vision extends the image processing paradigm for object classification. Here digital camera is used for the image capturing and LABVIEW software tool to build the GUI. The segmentation of leaf image is important while extracting the feature from that image. Mrunalini R. Badnakhe, Prashant R. Deshmukh compare the Otsu threshold and the k-means clustering algorithm used for infected leaf analysis in [5] . They have concluded that the extracted values of the features are less for k-means clustering. The clarity of k-means clustering is more accurate than other methods. The RGB image is used for the identification of The RGB image is used for the identification of disease. After applying k-means clustering techniques, the green pixels are identified and then using otsu's method, a varying threshold value is obtained. For the feature extraction, color co-occurrence method is used. The RGB image is converted into the HSI translation. For the texture statistics computation the SGDM matrix is generated and using the GLCM function the feature is calculated [6].

The FPGA and DSP based system is developed by Chunxia Zhang, Xiuqing Wang and Xudong Li, for monitoring and control of plant diseases [7]. The FPGA is used to get the field plant image or video data for monitoring and diagnosis.

The DSP TMS320DM642 is used to process and encode the video or image data. The nRF24L01 single chip 2.4 GHz radio transmitter is used for data transfer. It has two data compress and transmission methods to meet user’s different needs and uses multi-channel wireless communication to lower the whole system cost. Shantanu Phadikar and Jaya Sil use pattern recognition techniques for the identification of rice disease in [9].

This paper describes a software prototype for rice disease detection based on an infected image of a rice plant. They used HIS model for segmentation of the image after getting the interested region, then the boundary and spot detection is done to identify infected part of the leaf.

Proposed Methodology:

Digital cameras or similar devices are used to take images of leaves of different types, and then those are fed into the model (SVM Model) to identify the affected area in the leaf. Then different types of image processing techniques are applied on them like color based , feature based , texture based to process those images, and to get different and useful features needed for the purpose of analysing later. The images are fed into the model using the os library which is used to link the system directory with the model and then popping the output image with the class name printing on it.

Algorithm written below illustrated the step-by-step approach for the proposed image recognition and segmentation processes:

Step 1: Image acquisition is the very first step that requires capturing an image with the help of a digital camera.

Step 2: Pre-processing of input image to improve the quality of image and to remove the undesired distortion from the image. Clipping of the leaf image is performed to get the interested image region and then image smoothing is done using the smoothing filter. To increase the contrast, Image enhancement is also done.

Step 3: Mostly green colored pixels, in this step, are masked. In this, we computed a threshold value that is used for these pixels. Then in the following way mostly green pixels are masked: if pixel intensity of the green component is less than the pre-computed threshold value, then zero value is assigned to the red, green, and blue components of this pixel.

Step 4: In the infected clusters, inside the boundaries, remove the masked cells.

Step 5: Obtain the useful segments to classify the leaf diseases. Segment the components using the ML algorithm.

Limitations:

Limitation of existing work:

- The implementation still lacks accuracy of result in some cases. More optimization is needed.
- Priori information is needed for segmentation.
- Database extension is needed in order to achieve more accuracy.
- Very few diseases have been covered. So, work needs to be extended to cover more diseases.
- The possible reasons that can lead to misclassifications can be as follows: disease symptoms vary from one plant to another, features optimization is

needed, more training samples are needed in order to cover more cases and to predict the disease more accurately.

To remove these research gaps a new methodology for automatic detection as well as classification of plant leaf diseases using image segmentation has been proposed. The advantages of proposed algorithm are as follows:

1. Use of estimators for automatic Initialization of cluster centres so there is no need for user input at the time of segmentation.
2. The detection accuracy is enhanced with the proposed algorithm.
3. Proposed method is fully automatic while existing methods require user input to select the best segmentation of input image.
4. It also provides environment-friendly recovery measures of the identified disease.

Block Diagram:

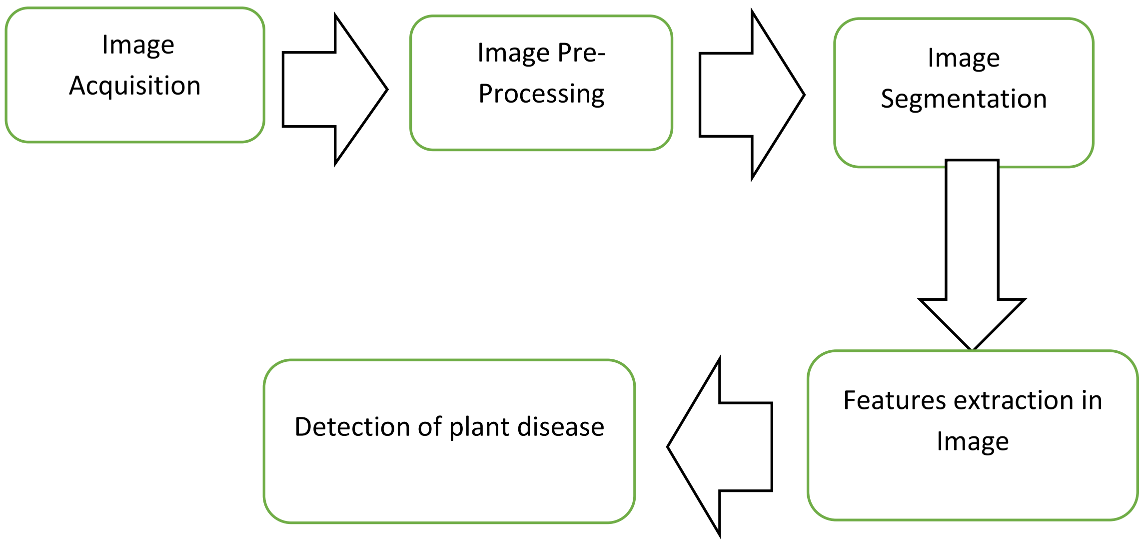


Figure 1: Block diagram of the flow of code

Methodology:

- **Image Acquisition**: First we need to select the plant which is affected by the disease and then collect the leaf of the plant and load the leaf image into the dataset and use it in our code. The images of the plant leaf are captured through the camera. This image is in RGB (Red, Green and Blue) form. color transformation structure for the RGB leaf image is created, and then, a device-independent color space transformation for the color transformation structure is applied [6]

- **Image Segmentation**: Segmentation means partitioning of image into various parts of the same features or having some similarity. The segmentation can be done using various methods like Otsu' method, k-means clustering, converting RGB image into HIS model etc.

  [1] Segmentation using Boundary and spot detection algorithm: The RGB image is converted into the HIS model for segmenting. Boundary detection and spot detection helps to find the infected part of the leaf as discussed in [9]. For boundary detection the 8 connectivity of pixels is consider and boundary detection algorithm is applied [9]

- **Image Pre-processing**: To remove noise in image or other object removal, different pre-processing technique region. Image smoothing is done using the smoothing filter. Image enhancement is carried out for increasing the contrast.

- **Feature Extraction**: Feature extraction plays an important role for identification of an object. In many applications of image processing feature extraction is used. Color, texture, morphology, edges etc. are the features which can be used in plant disease detection.

K-Nearest Neighbour Algorithm:

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- The K-NN algorithm assumes the similarity between the new case/data and available cases and puts the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using K- NN algorithm.

Step-1: Select the number K of the neighbors

Step-2: Calculate the Euclidean distance of **K number of neighbors**

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.

Step-6: Our model is ready.

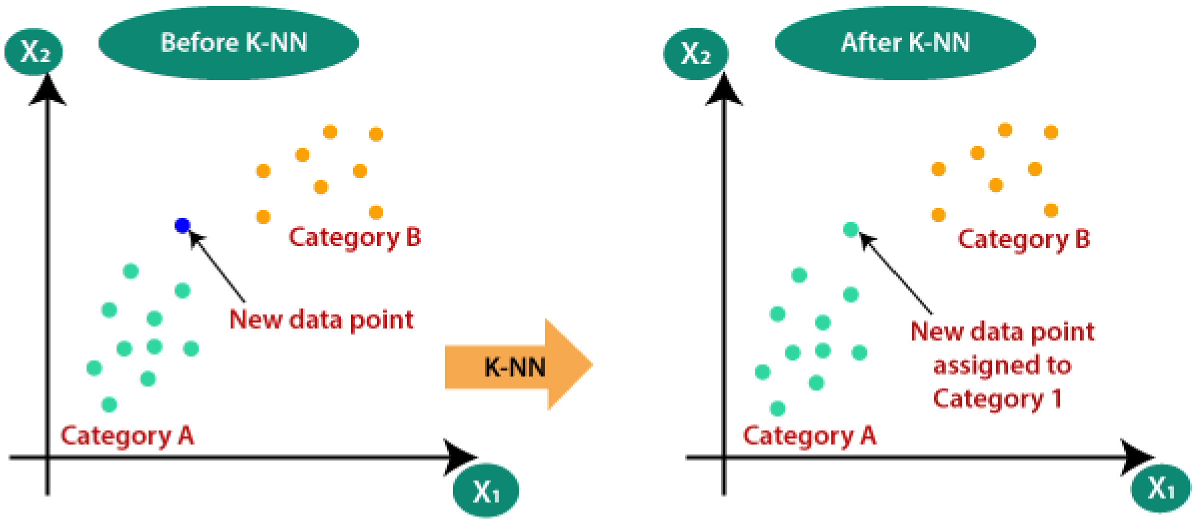


Fig 2: KNN Algorithm

Support Vector Machine (SVM): SVM is a statistical learning-based solver. Statistical is a mathematics of uncertainty.it aims at gaining knowledge, making decisions from a set of data. It is a simple classification problem given in two-dimensional input space. The two types of patterns indicate the images of crescent-shaped and oval objects. We can draw a line separating the two classes and many such possibilities exist.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

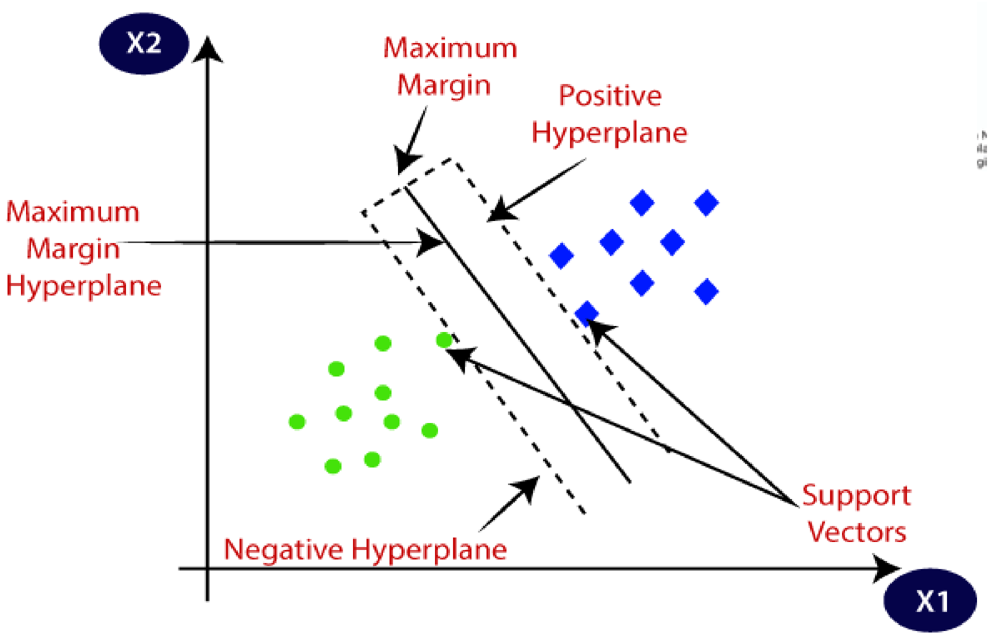


Fig 3:

SVM Algorithm

Output/Result:

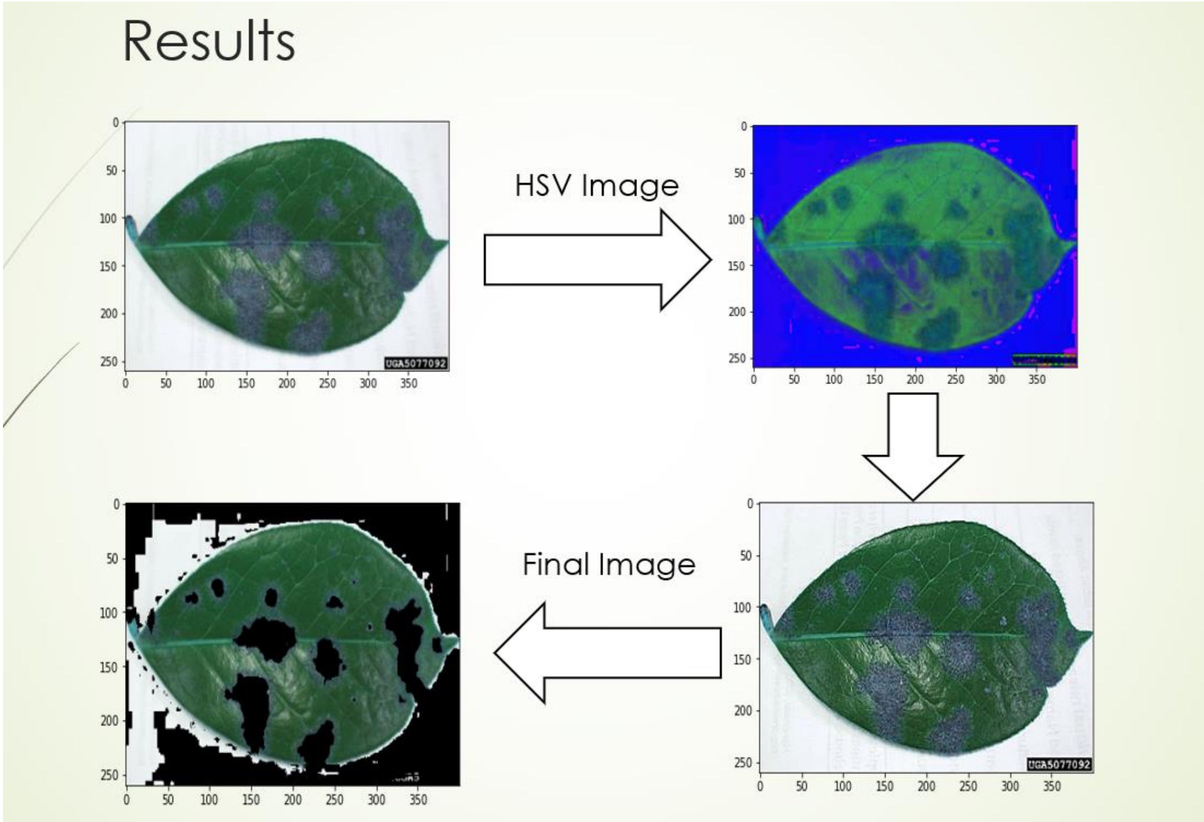


Fig 4: Steps of Image Processing

```
In [16]: # counting the number of pixels
         number_of_white_pix = np.sum(binary == 255)
         number_of_black_pix = np.sum(binary == 0)

         print('Number of white pixels:', number_of_white_pix)
         print('Number of black pixels:', number_of_black_pix)

         Number of white pixels: 73566
         Number of black pixels: 78054

In [17]: #total affected part of the leaf is
         ap=121078-120006
         print("Total no.of diseased pixel in the leaf will be =",ap)

         Total no.of diseased pixel in the leaf will be = 1072

In [18]: #percent affected part will be
         A = (ap*6.25)/100
         print("Affected part of the leaf will be =",A,"%")

         Affected part of the leaf will be = 67.0 %
```

Fig 5: Code for counting number of Pixels in image

Results of another image :

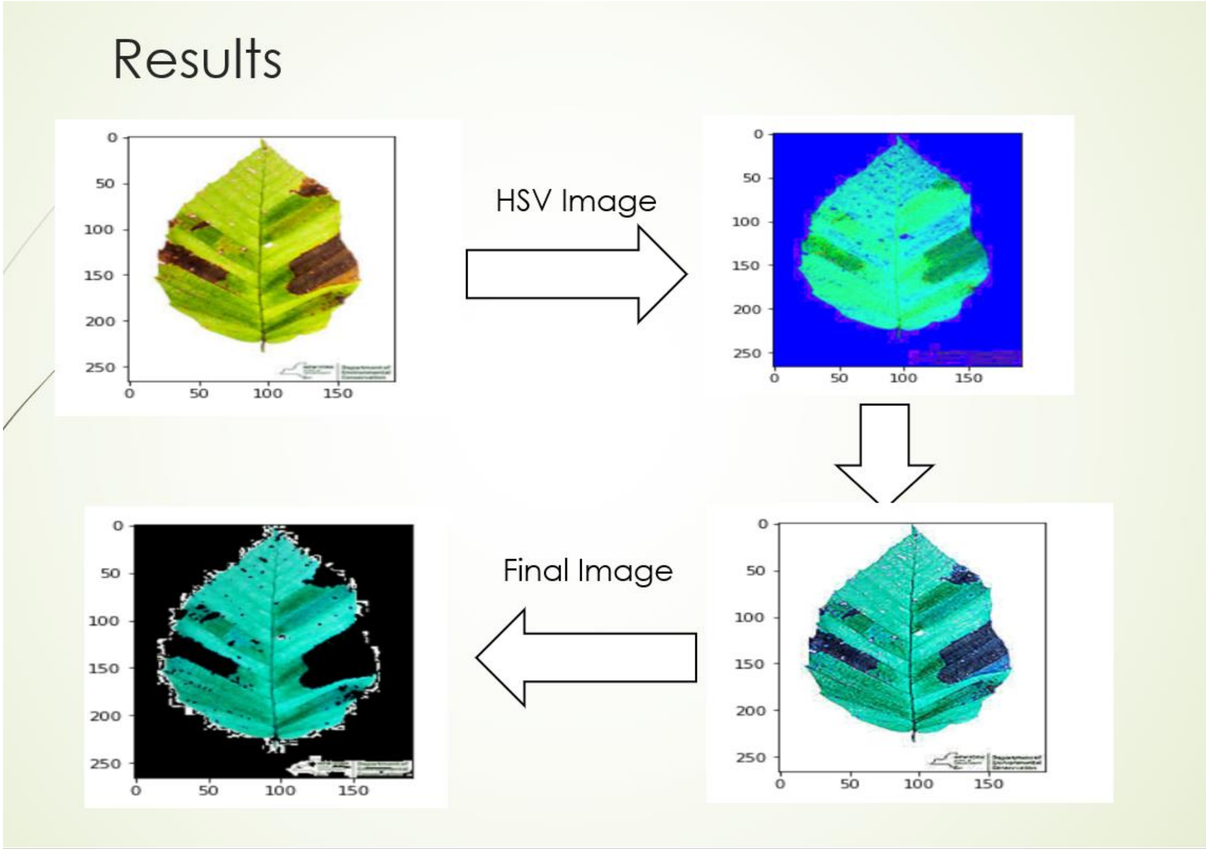
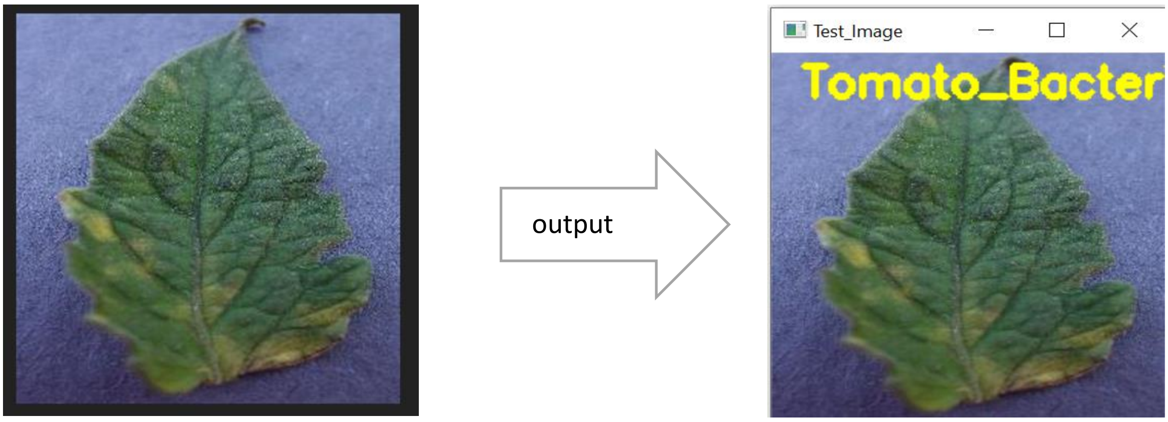


Figure 6: Result of another image

```
In [15]: # counting the number of pixels
         number_of_white_pix = np.sum(binary == 255)
         number_of_black_pix = np.sum(binary == 0)
         bp1=number_of_black_pix
         print('Number of white pixels:', number_of_white_pix)
         print('Number of black pixels:', number_of_black_pix)

         Number of white pixels: 201936
         Number of black pixels: 110064

In [16]: #total affected part of the leaf is
         ap=abs(bp1-bp2)
         print("Total no.of diseased pixel in the leaf will be =",ap)

         Total no.of diseased pixel in the leaf will be = 8787

In [17]: #percent affected part will be
         A = (ap*6.25)/1000
         print("Affected part of the leaf will be =",A,"%")

         Affected part of the leaf will be = 54.91875 %
```

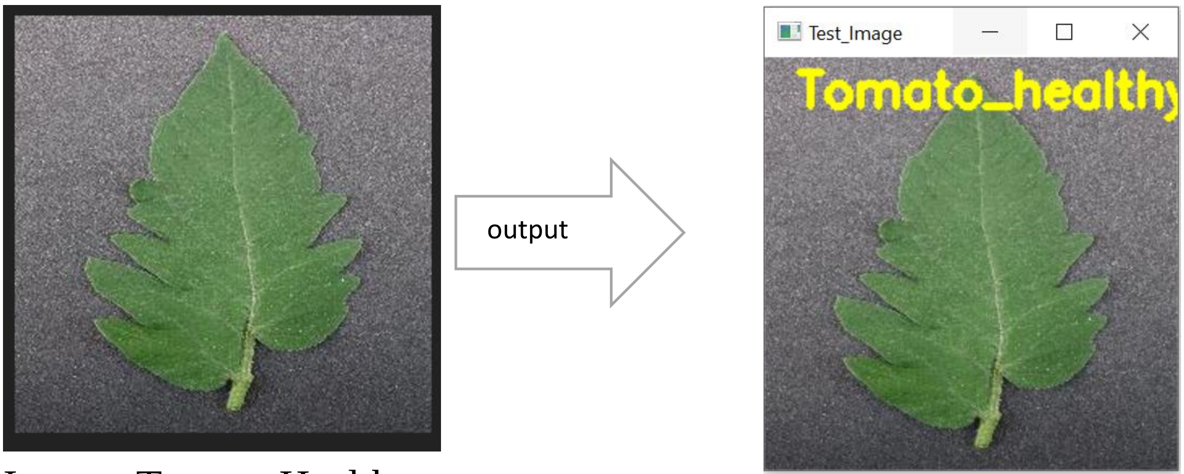
Figure 7: Calculation of accuracy of affected part of the leaf



#Figure :Input image with resulted output image

```
In [*]: # loop over the test images
test_path = "dataset/test"
for file in glob.glob('C:/Users/LENOVO/Desktop/sem 8/Input' + "/*.jpg"):
        # read the input image
        image = cv2.imread(file)

        # convert to grayscale
        gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

        # extract haralick texture from the image
        features = extract_features(gray)

        # evaluate the model and predict label
        prediction = clf_svm.predict(features.reshape(1, -1))[0]

        # show the label
        cv2.putText(image, prediction, (20,30), cv2.FONT_HERSHEY_SIMPLEX, 1.0, (0,255,255), 3)

       # display the output image
        cv2.imshow("Test_Image", image)
        cv2.waitKey(0)
```


Figure 8: Output of one Leaf with Disease name



Input – Tomato Healthy

Accuracy using Algorithms –

SVM :

```
In [49]: from sklearn import svm
```

```
In [50]: clf = svm.SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma=0.01, kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False)
         clf.fit(X_train,y_train)

Out[50]: SVC(gamma=0.01)
```

```
In [51]: y_pred = clf.predict(X_test)
```

```
In [52]: from sklearn import metrics
         metrics.accuracy_score(y_test, y_pred)

Out[52]: 0.7338888888888889
```

KNN :

```
In [18]: from sklearn.metrics import f1_score
         poly_accuracy = accuracy_score(yi_test, poly_pred)
         poly_f1 = f1_score(yi_test, poly_pred, average='weighted')
         print('Accuracy (Polynomial Kernel): ', "%.2f" % (poly_accuracy*100))
         print('F1 (Polynomial Kernel): ', "%.2f" % (poly_f1*100))

         Accuracy (Polynomial Kernel):  62.59
         F1 (Polynomial Kernel):  48.18
```

```
In [19]: from sklearn.neighbors import KNeighborsClassifier
         classifier = KNeighborsClassifier(n_neighbors = 1, metric = 'minkowski', p = 2)
         classifier.fit(x_train, y_train)

Out[19]: KNeighborsClassifier(n_neighbors=1)
```

```
In [20]: from sklearn.metrics import confusion_matrix, accuracy_score
         pred = classifier.predict(Xi_test)
         cm = confusion_matrix(yi_test, pred)
         print(cm)
         accuracy_score(yi_test, pred)

         [[38 17]
          [31 61]]

Out[20]: 0.673469387755102
```

Algorithm	**Accuracy**
SVM	73.33%
KNN	67.33%

Table 1: Accuracy

SVM with Different kernel	**Accuracy**
RBF	73.33%
Polynomial	62.59%
F1	48.18%

Table 2: Table for SVM accuracy

Conclusion:

Thus, an application built for the identification of disease affected plants and healthy plants is done and this proposed work focuses on the accuracy values during the real field conditions, and this work is implemented by having several plant diseases images. Overall, this work is implemented from scratch and produces a decent accuracy. The future work is to increase the number of images present in the predefined database and to modify the architecture in accordance with the dataset for achieving better accuracy. By using this concept, disease identification is done for all kinds of leaves and the user can know the affected area of the leaf in percentage by identifying the disease properly. The user can rectify the problem very easily and with less cost.

There are many cases where farmers do not have a fully compact knowledge about the crops and the disease that can get affected to the crops. effectively used by farmers thereby increasing the yield rather than visiting the expert and getting their advice.

References:

[1] Wenjiang Huang, Qingsong Guan, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Dong Liang, 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] Dr.K. Thangadurai, K. Padmavathi, "Computer Vision image Enhancement for Plant Leaves Disease Detection", 2014 World Congress on Computing and Communication Technologies.

[3] Monica Jhuria, Ashwani Kumar, and Rushikesh Borse, "Image Processing for Smart Farming: Detection of Disease 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 Plant Chili Disease Detection Using Image Processing Techniques", 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

[5] Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.

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

[7] Chunxia Zhang, Xiuqing Wang, Xudong Li, "Design of Monitoring and Control Plant Disease System Based on DSP FPGA", 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing.

[8] A. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol and A. Srikaew, "Grape leaf disease detection from color imagery using hybrid intelligent systems", Proceedings of ECTI-CON 2008. [9] Santanu Phadikar and Jaya Sil, "Rice Disease Identification using Pattern Recognition", Proceedings of 11th

International Conference on Computer and Information Technology (ICCIT 2008) 25-27 December 2008, Khulna, Bangladesh.

Plant Disease Detection Using Image Processing.
Plant Disease Detection Using Image Processing | IEEE Conference Publication | IEEE Xplore

Introduction to Image Segmentation
https://www.kdnuggets.com/2019/08/introduction-image-segmentation-k-means-clustering.html
Leaf Disease Detection using Image Processing.
https://www.researchgate.net/publication/318109025_Leaf_Disease_Detection_using_Image_Processing

Research on Plant Leaves.
(PDF) An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques (researchgate.net)