Doing Engineering -2
Project Report

on

LEAF DISEASE DETECTION USING IMAGE PROCESSING

In partial fulfillment of the requirements for the award of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering with Data Science(CSD)

N.SAKETH REDDY 21E51A6731

V.RAMESH 21E51A6762

V.ESHWAR 21E51A6764

Department of ET, HITAM

HYDERABAD INSTITUTE OF TECHNOLOGY AND MANAGEMENT

Autonomous, Approved by AICTE, Accredited by NAAC, NBA.

Gowdavelly (Village), Medchal (Mandal), Medchal - Malkajgiri (Dist.), Hyderabad, TS- 501401.

2023–2024

HYDERABAD INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(Affiliated to JNTUH, Approved by AICTE, Accredited by NAAC, NBA – TS 501401)

DEPARTMENT OF COMPUTER SCIENCE [DATA SCIENCE]

CERTIFICATE

This is to certify that the internship project work entitled **"LEAF DISEASE DETECTION USING IMAGE PROCESSING"** is being submitted by **N.SAKETH REDDY, V.RAMESH, V.ESHWAR** bearing Roll No. **21E51A6731, 21E51A6762, 21E51A6764** in partial fulfillment of the academic requirement, at Hyderabad Institute of Technology and Management, Hyderabad is a record of bonafide work carried out by them under our guidance. The matter contained in this document has not been submitted to any other University or institute.

Lab Incharge

Dr. G. Omprakash

Department of ECE, HITAM

Head of Department

Mrs. ILA CHANDANA KUMARI

Assistant Dean ET

Department of ET, HITAM

DECLARATION

We hereby declare that the internship project entitled **"LEAF DISEASE DETECTION USING IMAGE PROCESSING"** submitted to **Hyderabad Institute of Technology and Management affiliated to Jawaharlal Nehru Technological University**. Hyderabad (JNTUH) as part of academic requirement in the Doing Engineering-2 lab.

N.SAKETH REDDY 21E51A6731

V.RAMESH 21E51A6762

V.ESHWAR 21E51A6764

ABSTRACT

Plant diseases are one of the main reasons for the decline in the quality and quantity of agricultural produce. Farmers have a lot of trouble recognizing and controlling plant diseases. As a result, it is critical to diagnose plant diseases at an early stage so that farmers may take accurate and timely action to avert future losses. The study focuses on a method for identifying plant diseases based on image processing and machine learning. We would create a device in this project that would assist farmers in identifying plant illness by capturing a leaf image. A set of algorithms in the system can detect leaf disease. The input image given by the user to the system undergoes various image processing steps to detect the disease

1. Introduction

The revelation of disease is decisive in enriching the grade of the agricultural output and intercepting the overall depletion of the plants. The research on plant ailment is the mechanism of learning the detectable patterns visible on the plant. The plants being exposed to the outside environment get diseases from exposure. Plants are attacked by several diseases which usually target the specific plant parts like stem, fruit, leaf, seed etc. Leaf symptoms are an important origin of knowledge to detect the diseases in multiple types of plants and therefore it has to be contemplated in identifying the disease. This reports about the revelation of leaf disease using Image processing strategies, as it is a trendy process being used in agriculture, focused on five sorts of disease with training set of 20 images and testing of 20 images. They have inured K-means clustering algorithm for segmentation and used varied feature extraction strategies like skewness, contrast, etc.to pull out desired features. SVM classifiers were implemented for classification. Their model procured a recognition rate of 92.4%. In the proposed work, initially we take the picture of the leaf to identify which section is corrupted. Then image preprocessing is done where the picture is changed from RGB to Grayscale and Adaptive Histogram Equalization (AHE) is applied to enhance the contrast of the picture. There are many features in the image like texture, edges, morphology, colour, accounts color, texture and morphology. They have proved that morphological results are good than other results have employed K – nearest neighbor classifiers to recognize the plant disease. The outcomes of the features are matched and the results are classified using SVM classifiers and the resultants are presented to the user, have used Linear SVM classifier and K-means clustering to identify diseases in leaf.

The revelation of plant disease is a prominent research theme in the field of computer science. With the aid of Intelligent systems, the diseases can be detected effectively .The plant leaves are mainly affected by varied micro organisms. This paper centers on the discovery of disease in plants using the input picture. The disease identification concerns the steps like transfiguration of the picture format from RGB to Grayscale. Adaptive Histogram Equalization (AHE) is accustomed to improvise the contrast in the picture. The 13 prominent attributes are extracted by handling a feature extraction method called GLCM or Gray Level Co-occurrence Matrix .The standard benchmark images are trained using SVM classifier and the outcomes are displayed in the output screen.

2. Grayscale Conversion

A RGB photograph is basically a M*N*3 array of color pixel, in which each color pixel is a triplet which corresponds to red, blue and green shade issue of RGB image at a specific spatial location.Similarly, A Grayscale picture may be regarded as a single layered image. In MATLAB, a grayscale image is largely M*N array whose values have been scaled to represent intensities .In MATLAB, there is a function referred to as rgb2gray () is to be had to convert RGB picture to grayscale picture. I = rgb2gray (RGB) converts the true coloration image RGB to the grayscale image.

3. Adaptive Histogram Equalization

AHE is used for improvising the contrast and edges in the input picture. CLAHE is a modified part of AHE. In this approach enhancement characteristic is implemented on over all community pixels and transformation function is derived. This is differing from AHE because of its comparison limiting. The method used maximum value to clip the histogram and redistribute in grey stage image. Their algorithm implemented one at a time for heritage and foreground and restricts the noise, enhance the evaluation. Distribution parameter are used for shape of histogram equalization graph.

4. Feature Extraction

The majority of the diseases that corrupt the plant can be scrutinized by discoloration of leaves, or the presence of residue on the leaves thereby converting the natural texture of the leaves

4.1 Texture Feature Extraction

The texture can also be said as the regularity of patterns in the picture. The standard practice employed for is the concept of GLCM .It is a strategy of deriving second order statistical features. In a picture with pixels of different intensities, GLCM is a categorization of the frequency of combinations of gray levels that co-occurs in a section of the picture. The GLCM is a square matrix where the rows and columns are impartial to the quantity of gray levels in that image. The distances (i.e.), the similarities among the various histograms are used to evaluate the matrix by employing averages of four angles, 0, 45, 90, 135 degrees between the selected histogram and its encircling ones. Consider a pixel with a value x and one of its surrounding ones with value y.
Then mathematically, a GLCM is constructed for the image I of size N×N using the formula

$$G=\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\begin{matrix} 1 if\ I(i,j)=x\ and\ I\left(i+d_i,j+d_j\right)=y \\ 0 \qquad otherwise \end{matrix}\right\}$$

After devising the GLCM, the following features are accounted.

Table 1: Formulae for Texture Feature Extraction

Contrast	$\sum_i \sum_j (i-j)^2 G(i,j)$
Correlation	$\frac{\sum_i \sum_j (ij) G_{ij-\mu_i \mu_j}}{\sigma_i \sigma_j}$
Energy	$\sum_i \sum_j G\,(i-j)^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i+j)^2} G_{ij}.$
Mean	$\frac{1}{N} \sum_i \sum_j G_{ij}.$
Standard Deviation	$\sqrt{\left(\frac{\sum (x_i-\mu)^2}{N}\right)}.$
Entropy	$-\sum_{i=1}^{N} \sum_{j=1}^{N} G_{ij} \log[G_{ij}]$
Variance	$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(i-\mu)^2 G_{ij}}{\sum_{ij} G_{ij}}.$
Skewness	$\frac{\frac{1}{N} \sum_{i=1}^{N} (x_i - x')^3}{\left(\frac{1}{N} \sum_{i=1}^{N} (x_i - x')^2\right)^{\frac{3}{2}}}$
IDM	$\frac{\sum_{i=0}^{N} \sum_{j=0}^{N} G_{ij}}{1+(i-j)^2}.$

5. Support Vector Machines

In machine learning, SVM or Support Vector Networks shown in the figure 4 are supervised learning model which areassociated with learning algorithms. The ana\lysed data used for classification and anaysis. SVM models are the representation of the examples as points in space, mapped so that these examples of the different categories are divided by a clear gap that is as wide as possible. SVM can efficiently perform a non linear classification by using the algorithm called Kernel Trick in addition to performing the linear classification. A good separation is possible by using hyperplane which has the longest distance to the closest trained data point of any class by lowering the generalisation error of the classifier. SVM allows to classify data which are linearly separable for non-linear classification, Kernel Trick can also be used. In comparison with newer algorithms like neural networks, they may have two advantages: better performance with limited number of samples and higher speed with good efficiency. This algorithm is very suitable for text classification problems. It is most commonly and to access a dataset of tagged samples.

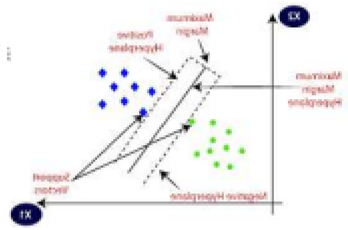


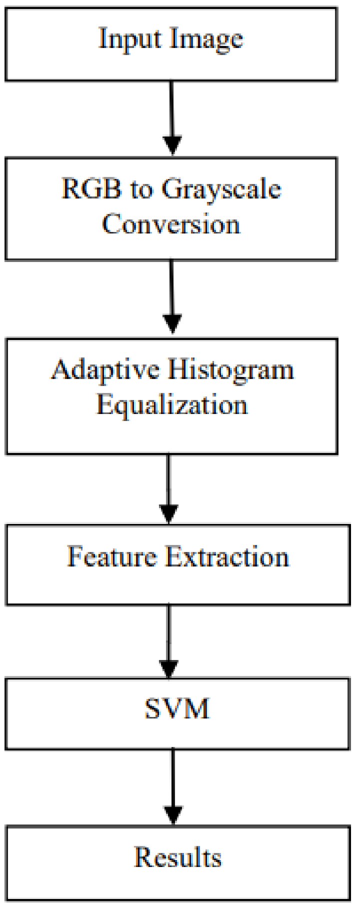
Fig -4: Support Vector Machine

Support vector machines are employed for classification of the disease. The SVM classifiers are trained using the images and disease types in the dataset. The features, dataset, and the type of the disease is given as input to the SVM classifier to train and test the data. [13]The kernel trick is employed to map the classes from input space to a kernel space. [14]The varied types of kernel functions are polynomial kernel function, Radial Basis Function (RBF), sigmoid kernel function. The RBF kernel function is propounded in this paper. If xi and xj are two samples then the Formula for

$$k\left(x_i - x_j\right) = e^{-\gamma(x_i - x_j)^2} \quad (2)$$

Where $\gamma = \frac{1}{2\sigma^2}.\sigma$ (3) is a free parameter.

6. Block Diagram



7. Methodology

Input: Standard benchmark images in RGB scale.
Output: Name of the disease
Step 1: The RGB color components of the image are transformed into Grayscale components
Step 2: Adaptive Histogram Equalization is applied to enhance the picture contrast.
Step 3: Application GLCM for texture extraction
Step 4: Classification of disease using SVM classifiers
Step 5: Display the results

8. Experimental Results

The following sample images are taken as the dataset of the project reported, and are used for the testing of the various operations that are done

Fig. 1 : Disease label : Anthracnose

Fig. 2 : Disease Label: Bacterial Blight

Fig. 3: Disease Label: Citrus Canker

Fig. 4: Disease Label: Gray Mould

The image of the leaf that is to be given as the input is put into the folder and then the GLCM values of the image has to be loaded into the database. Let us assume an input leaf as shown below.

Fig.5 Input Leaf Image

According to the flow of the project, the above input image is converted into its grayscale image. The result obtained upon converting is the figure shown below.

Fig.6: Grayscale Image

This grayscale image is then subjected to Adaptive histogram equalization to enhance the contrast of the image. The result of the adaptive histogram equalization is as follows.

Fig.7: Result of AHE

The value of the 8×8 GLCM matrix that is constructed for the above image of the adaptive histogram equalization is shown in the table.

Table2: GLCM values

640	613	103	28	10	1	0	0
578	17347	2444	141	53	20	15	1
148	2409	15950	2751	274	100	48	7
23	125	2881	13478	2933	331	141	33
1	28	217	3240	10735	2121	216	56
0	3	43	271	2388	6602	1112	123
0	0	6	53	239	1254	3056	537
0	0	0	1	23	113	563	1293

The values for the 13 statistical parameters calculated using the GLCM matrix is given below

Table3: GLCM parameters values

1	Contrast	0.4343
2	Correlation	0.9162
3	Energy	0.1010
4	Entropy	7.5866
5	Homogeneity	0.8471
6	IDM	255
7	Kurtosis	2.2498
8	Mean	77.0277
9	RMS	14.8734
10	Skewness	0.6318
11	Smoothness	1.0000
12	Standard Deviation	62.3032
13	Variance	2.3924e+03

And finally, the output is given as the disease infecting the leaf

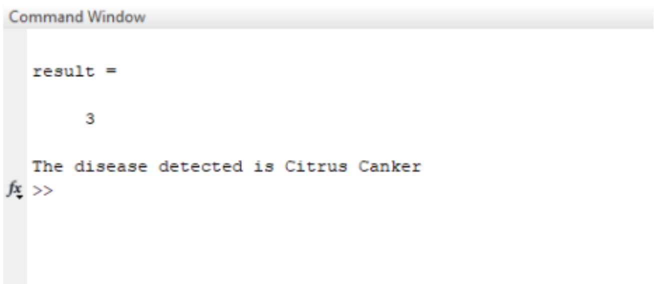


Fig.8: Output window snip

10. Conclusion

The color constituents of the image are deformed into grayscale. Adaptive Color Histogram is applied so as to reduce the overall contrast of the image. Textural feature extraction is done using GLCM. The dataset is trained using SVM classifiers, and thus the system identifies the disease that the leaf is affected. This project focuses on simple revelation of diseases in leaves at a relatively lower cost and time than conventional methods. In the future work, Convolutional Neural Networks can be used for the image classification. Also, Image Compression can be applied so that the processing time is lower than conventional methods.

ADVANTAGES

It is easy to implement and gives quick result.

Good potential with ability to detect plant leaf disease.

Prediction accuracy is high and having robust working when training example have error in

them.
Efficient and User friendly
Eco-friendly
Reliable
Low cost
Accuracy

APPLICATIONS
This project can be implemented at –
1) Bio-farm
2) Bio- Pesticides
Early recognition of plant diseases and pests
Detecting diseases on leaf of plant at early stages gives strength to overcome it and treat it appropriately by providing the details to the farmer that which prevention action should be taken.
Agriculture crop diseases detection for treatment.

Need of Project
From the start of humans are directly work in farms but from the start of 21st century many industries worked to reduce this human labor by making robots and machines. Now-a- days many industries are trying to reduce this human labor by making robots and machines. Now more and more chemicals applied to plants without knowing the requirement of plants. Hence productivity of agriculture decreases.

Objective of Project
To implement leaf diseases detection with improvement accuracy.
Segmentation of diseases part of leaf
To get the processed leaf image as an input.
To segment the image using K-Means clustering algorithm.
Finally provide the type of disease attacked in the leaf using SVM classifier and severity level.

code explanation(multisvm)

Explanation:

The function `multisvm` takes three inputs: training matrix (`T`), group labels (`C`), and testing matrix (`test`), and returns the resultant class (`itrfin`).

The code initializes some variables (`Cb`, `Tb`, `u`, `N`, `c4`, `c3`, `j`, `k`) and iterates over the testing samples in `test`.

The primary loop is used for multiclass SVM classification using a reduction approach. It trains an SVM for each class and reduces the training set and group labels based on the SVM predictions.

The while loop continues until either all classes are successfully classified or the maximum number of iterations is reached.

Inside the loop, it trains an SVM using the `fitcsvm` function with an RBF kernel and makes predictions on the testing sample.

It then reduces the training set and group labels based on the SVM predictions.

The process continues until either all classes are classified or the maximum number of iterations is reached.

The result (`val`) for each testing sample is stored in the `itrfin` matrix.

The function returns the `itrfin` matrix containing the resultant class for each testing sample.

code explanation(mainproject)

These lines close all open figures, clear the workspace, and clear the command window, respectively.

These lines open a dialog box for the user to select a leaf image file. It reads the selected image using `imread`.

It displays the original leaf image.

It resizes the image to 384x256 pixels and converts it to grayscale if it's a color image.

It displays the grayscale version of the leaf image.

It performs adaptive histogram equalization on the grayscale image to enhance contrast.

It creates the Gray Level Co-occurrence Matrices (GLCMs) from the preprocessed image.

It calculates various statistical features (Contrast, Correlation, Energy, Homogeneity) from the GLCMs.

It calculates additional statistical features (Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness).

It calculates the Inverse Difference Moment (IDM) as another texture feature.

It calculates the Inverse Difference Moment (IDM) as another texture feature.

It loads a pre-existing dataset (`dataset11.mat`) containing features (`dataset`) and corresponding labels (`diseasetype`).

It checks if the feature vector of the current leaf matches any in the dataset. If not, it throws an error.

It uses a custom `multisvm` function to classify the leaf based on the SVM model trained on the dataset.

It displays a message based on the classified result

The code appears to be a leaf disease classification system using texture and statistical features, where SVM is trained on a dataset to classify the leaf into different disease categories. The displayed messages provide recommendations for farmers based on the detected disease.

software requirements

- **MATLAB** is the core software for developing and running the image processing and machine learning algorithms.
- Ensure you have the Image Processing Toolbox and Statistics and Machine Learning Toolbox installed.

MATLAB provides an interactive environment with a command-line interface and a graphical user interface (MATLAB Desktop). The MATLAB Command Window allows for quick execution of commands and script development.

MATLAB provides a comprehensive set of tools for machine learning, including classification, regression, clustering, and deep learning. The Statistics and Machine Learning Toolbox facilitates the implementation of algorithms like Support Vector Machines (SVM) for classification.

Dataset:

- Collect or use an existing dataset of leaf images with annotations for disease labels. This might not be software, but it's a critical requirement for training and testing your model.

hardware requirements

Processor: A modern multicore processor (e.g., Intel Core i5 or AMD Ryzen 5) for smoother performance.

RAM: At least 8GB of RAM, but 16GB or more is recommended for seamless multitasking.

Hard Disk: A solid-state drive (SSD) with at least 256GB of storage for faster load times and efficient project management.

Monitor: A larger and high-resolution display (e.g., 24" Full HD or higher) for better design and development work.

Keyboard and Mouse: Standard keyboard and mouse should suffice, but consider investing in ergonomic peripherals for comfort during long work sessions.

3.Proposed Methodology: The proposed method for leaf disease detection using machine learning is a promising approach that can significantly improve the efficiency and effectiveness of plant disease detection. It has the potential to revolutionize the agricultural industry by enabling early detection and intervention of plant diseases, ultimately leading to increased productivity and profitability for farmers. Further research and development in this area can help to refine and improve the performance of the system, as well as expand its applications to other areas of plant health and agriculture. there are many areas of future research and development for the proposed method for leaf disease detection using machine learning. By continuing to refine and improve the system, we can help address a range of challenges faced by farmers in the agricultural industry and ultimately

improve the sustainability and productivity of the industry.

Proposed Methodology

To identify the affected area, the images of various leaves are taken Then to process those images, various image-processing techniques are applied on them to get different and useful features required for later analysing purpose.

The system‟s architecture is as follows:

Input Image

RGB to Grayscale Conversion

Adaptive Histogram Equalization

Feature Extraction

SVM

Results

Figure 1: Flowchart for the Proposed Concept

II. LITERATURE REVIEW

In paper [1],by Faye Mohameth, Chen Bingcai, Kane Amath Sada. In this paper authors implement a deep learning method to identify plant leaf disease, in which 3 deep learning models like VGG16, Google Net, ResNet 50 used to get good accuracy. (Faye Mohameth, Chen Bingcai, Kane Amath Sada, “Plant Disease Detection With Deep Learning And Feature Extraction Using Plant Village” Journal Of Computer And Communications,Doi: 10.4236/JCC. 2020.86002 Jun. 17, 2020.)

In paper [2],by S.S.Saranya, Nalluri Chandra Kiran, Komma Jyotheeswar Reddy. They proposed a project which increases the growth of farming outputs, their system is using image processing method (SVM) for detecting affected leaves and identifying using edge detection, texture, diseases spots compared with combined Healthy and Unhealthy leaves dataset and it gives the result as Disease name.(S.S. Saranya, Nalluri Chandra Kiran, Komma Jyotheeswar Reddy, “Identification Of Diseases In Plant Parts Using Image Processing”, International Journal Of Engineering & Technology, 7 (2.8) 461-463 , 2018.)

In paper [3],by Prof. Swati Pawar, Ms. Shweta Patil, Ms.Tejaswini Patil, Mr.Nasruddin Shaikh. The authors implemented a system which captures the image of the leaf and then uploads it to a dataset of diseases, in which using the Imageprocessing module and SVM classifier the disease is detected. They worked on the 3 basic leaves of the grapeplant. According to these considerations they worked on the project to detect the disease, they also implemented a hardware using microcontroller to inform farmers about the disease so famer can take preventive actions as soon as possible.(Prof. Swati Pawar Ms.Shweta Patil Ms .Tejaswini Patil Mr. Nasruddin Shaikh,” Detection And Classification Of Unhealthy Regions Of Grapes Plant Using Texture”, IJSRD, T Vol. 5, Issue 04, 2017.)