Field Management and Crop Disease Identification System

Navneet Mishra[1]
[1] Professor,
Electronics and Instrumentation Engineering,
Galgotias College of Engineering & Technology,
Greater Noida, India

Shivam Gupta[2]
[2]Professor,
Electronics and Instrumentation Engineering,
Galgotias College of Engineering & Technology,
Greater Noida, India

Vaibhav Jaiswal[3]
[3]B.Tech Student,
Electronics and Instrumentation Engineering,
Galgotias College of Engineering & Technology,
Greater Noida, India

Vivek Ranjan[4]
[4]B. Tech Student,
Electronics and Instrumentation Engineering,
Galgotias College of Engineering & Technology,
Greater Noida, India

Abstract-- India economy and food security in India is dependent on its agriculture. Diseases in crops is main constraint for farmers in India, diseases in plants initially appear as spots on the leaves. These diseases can be identified at initial stage and be treated to prevent from loss. Stranded method of disease identification requires human expert which is dependency and at times not affordable. Presented project is solution to this problem. Here Disease detection is done using digital image processing through MATLAB code written to identify various diseases. An application is to be installed by owner phone used for communication over Wi-Fi to receive notification and updates. Some feature seeming important like detection of intruder, temperature, humidity also added.

Keywords-- K-means clustering, SVM, Image Processing, IoT, Agriculture monitoring, Sensors.

I. INTRODUCTION

Diseases in plants can be caused by pathogens and microorganisms. Pathogens are biotic agents like virus, fungus etc which migrate from the infected leaf to the heathy leaf and infect it [1]. Microorganisms like fungi, bacteria and nematodes also cause diseases in plants. Identifying disease in plants is done by plant pathology experts with their eye observation and knowledge [2]. But now this creates dependency on experts visit to field and if disease identified lately then it may result in loss of yields. Hence a new method to detect plant disease based on MATLAB which is more fast, accurate and mainly removes dependency so realising this project will be beneficial to farmers facing these issues [3]. In this project we start with collecting set of samples of infected leaf or their images saved in folder name disease dataset. Our program has been tested on two diseases namely "BLACKSPOT"," ANTHRANOSE". Succeeding steps are performed manually and by program in order to know the disease name starting with step one which is selecting image and accepting it as an input or take fresh photo with camera interfaced with the system the step two is the image pre-processing, image segmentation, feature extraction of the selected image to be performed by MATLAB code to get set of values that represent the infected part of leaf as numerical values[4]-[6], step three is

matching the feature values, these are set of thirteen numerical values that represent the part of infected leaf image and these feature values are compared to values stored in diseaseset, a MATLAB file as a dictionary of feature values and disease name then at last name of disease is displayed on LCD display. Two different algorithms are used in this program namely k-means clustering and SVM (Support vector machine) [7]-[10]. The method for cure and prevention for the identified disease can also be implemented as part of this program. The system built for monitoring the crop field with the help of humidity and temperature sensors DHT 11 and IR sensor module is connected in network with internet through Wi-Fi module ESP 8266 and information is transferred to owner through this method [11]. Multiple IR sensors module is linked to the controller board through network and are placed at several positions in the field so that whole of the field is covered according to the range of sensors [12]. Being IoT-based system updates about field parameters can be obtained from anywhere and anytime.

II. LITERATURE REVIEW

A. *Sanjay B.* Patil and Shrikant K. Bodhe [4] proposed paper on Leaf disease severity measurement using image processing for diseases caused in sugarcane only that appears initially as generally brown spots on the leaves. They also did use of image processing using techniques Simple threshold and Triangle thresholding methods separate infected area. Lastly diseases are identified by calculating the quotient of lesion area and leaf area. Their program required outlining the leaf on white background paper that is an additional step.

B *Sanku Kumar Roy, Sudip Mishra* [12] *proposed a prototype system for intrusion detection based on multiple sensors around the boundaries of field connected all together wirelessly as a network.* This system is complete with alarms connected in the field owner house and also send a text message to the owner cell phone when an intruder enters into the field. All the monitoring is done by a microcontroller system. However, it doesn't mention the kind of intruder.

C. *Amritha A Joshi, B.D. Jadhav [14] proposed Monitoring and controlling of rice plant diseases using image processing techniques. His paper is for disease identification in rice plants also rice is a staple food of India hence this project can help rice growing farmers.* This project is also based on Image processing techniques which will help in accurate and fast detection. They focused on four diseases namely rice bacterial blight, rice blast, rice brown spot and rice sheath rot. Their algorithm uses different features like shape of infected leaf part, colour of a diseased portion of the leaf then the resulting features are used to classify disease using technique of Minimum Distance Classifier (MDC) and the k-Nearest Neighbour classifier also known as k-NN method.

III.   MATERIAL AND METHOD

Materials

Infected samples of leaf, White paper sheet, Arduino Board, IR Sensor, ESP 8266 WIFI Module, DHT 11 Temperature and Humidity Sensor, LCD Display, Breadboard and jumper wires, LED, Camera module, PC with MATLAB, Disease Identification Program.

Method

A. *Image Acquisition*: Image acquisition is a process of obtaining image either taking the saved image in a folder as the input to program or capturing image from the camera in real time .In this case the diseased leaf is taken in environment suitable for camera to take clear image. Infected leaf should be placed flat and on white background that can be a paper sheet. The image captured by camera be crosschecked so as to ensure that the picture taken must contain the clear image of infected part of leaf and further proceeded to next steps. The image can be read from the folder being of any extension. Mainly the format is jpeg, png , bmp containing image in RGB format .Here R represents red, G represents green,B represents blue plane and each plane has M x N pixels ie. M rows and N coloums of pixels see Fig 1. By the knowledge of how the images are processed in computer one can easily tell that they are merely numbers that represent the intensity of colour. Therefore RGB image is just a matrix of M x N x 3 numbers.
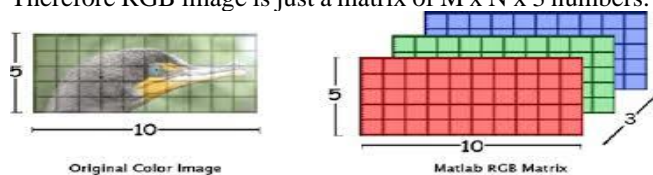

Fig 1: Representation of image as RGB Matrix.

There are other several ways of representing the image other than RGB like CIELAB ,HSV, CYMK etc.
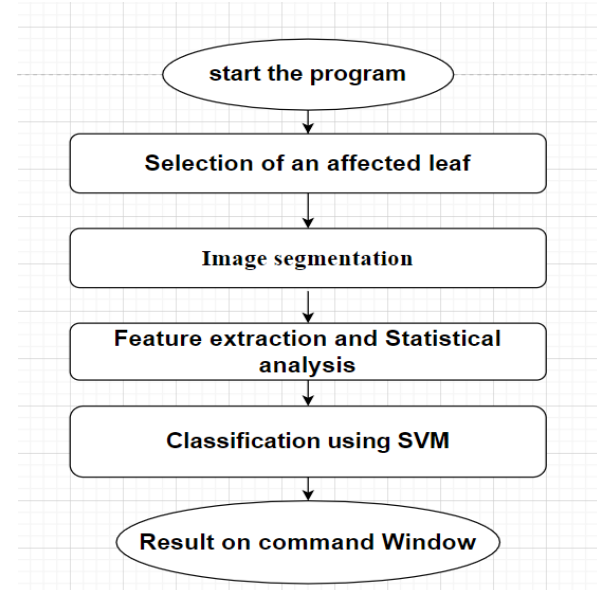

Fig 2: MATLAB program flowchart

B. *Image Segmentation*: In Image segmentation we divide the pixels of the image in smaller units based on the characteristics like colour values ,texture and edges in image[6].

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Fig 3: Mathematical Representation of output by k means clustering function(J)

The K-means clustering algorithm see Fig 3 that is used to make cluster of objects which are not explicitly labelled in training data set and guess the pattern that is used to take better decisions see Fig 4 [7][8]. It is frequently used as the pre-processing step before preceding to another algorithm. This algorithm has closely related to k-nearest neighbour classification algorithm which is also a popular machine learning algorithm used for classification and often they are confusing due to their similar names. Due to its easy available in many Open Source and Proprietary software its widely popular and used.
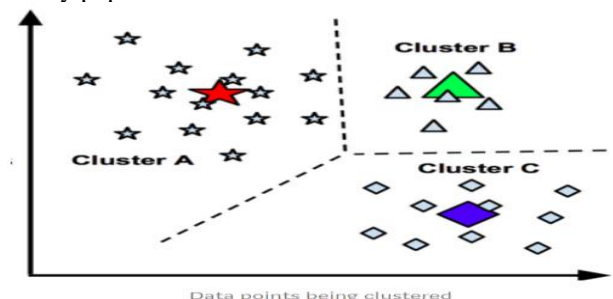

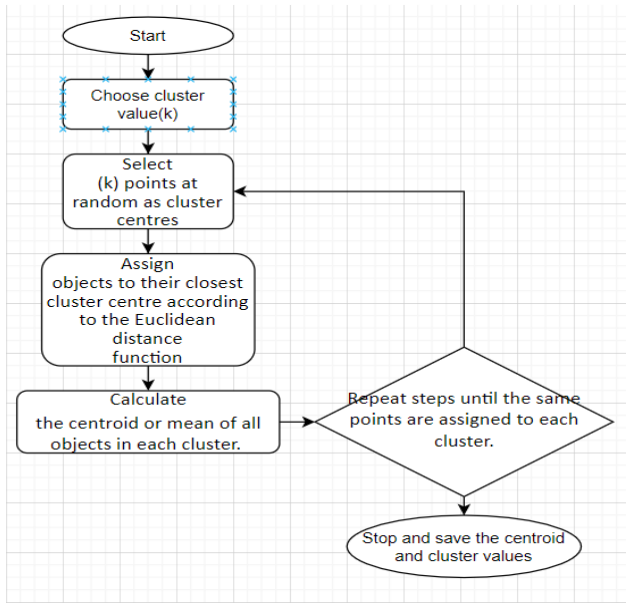Fig.4: k-means clustering, data points being grouped

Fig 5: Algorithm for k means clustering as flow chart.

Calculating the Euclidean distance between the pixel and corresponding colour marker then the resulting smallest distance helps to choose which pixel of the image most closely matches to which colour marker. As an example, is pixel x matches to blue colour then that pixel is marked as blue.

A practical way is to compare the results of the multiple executions with different k values and choose the best one based on the predefined conditions. In general, a large value of k will lead to decreases the error but at same time will increases the risk of overfitting.

Here in program we will use K Means clustering function from the MATLAB library that it stated as -:

$[idx,C] = kmeans(X,k)$

- Where idx is a M rows*N coloumns matrix .
- C returns the values of centroid .
- X is the reshaped array of image in CIELAB format and saved in X.
- k= number of cluster, here k =3.

Also kmeans() function will take some arguments which are distance criteria as sqeuclidean that means Squared Euclidean distance and replicates that represents the number of times clustering is to be repeated using the new initial cluster centroid positions.

The image is converted in lab format is converted into double values but only a* ,b* layers are only taken in consideration and the reshaped into matrix of M x N x 2. CIELAB expresses color in three axis L* for the luminosity or lightness that starts from black (0) to white (100), a* starts from green (−) to the red (+) on another axis, lastly b* starts from blue (−) to the yellow colour (+) on the last third axis see Fig 6. The same could be seen in the below figure. Reason Red/green and blue/yellow are choosen as pairs is because of the fact that they are opposite colour. CIELAB color were made to approximate about human vision.
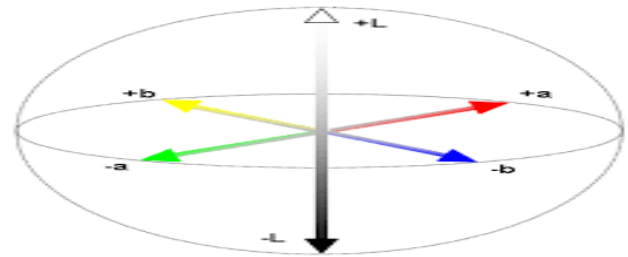
Fig 6: CIELAB colour space

Cluster value (k) is initialized to three thereafter implementation of Kmeans() with distance metric as squared eqcilidean and repetition values as arguments is taken. Kmeans() returns two matrices the first being idx , second is C . Then idx is reshaped M rows and N coloums.A blank cell array is created to store the images formed as a results of clustering. A new matrix is created by repeating reshaped idx matrix 3 times back to back to form a M rows x N columns x 3 so that it could be compared to original image matrix in RGB by for loop and resulting images are stored and displaying three images on screen with their position indicated by numbers one to three.

C. *Feature Extraction* : Process of Feature Extraction starts with selection of an image by user to give input into the dialogue box based on the closest match of part of disease affected leaf with the three cluster images, making sure that features are extracted only from the disease affected part of the leaf. It's made sure to convert images to grayscale if the selected image is RGB and store in variable. A Gray Level Cooccurrence Matrices (GLCM) is a method of calculating texture of the image by checking how frequently the pairs of pixels with certain values and with a specific spatial relationship are present ,resulting form is a GLCM matrix and its sometimes also known as gray-level spatial dependence matrix. By using GLCM matrix we can calculate certain statistical measures that would help in next steps of classification , which is the reason we consider this in our program .After you form the GLCMs, you can calculate the needful statistical values from them. These several statistics provide important data about the texture of an image. The following is the lists of the included statistics starting with Contrast, Correlation, Energy, Homogeneity. Many other features are calculated from the selected segmented image matrix like Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness Kurtosis, Skewness and lastly Inverse Difference Movement (IDM). We now put all the features (thirteen in number) in an array [Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation , Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM]. This array is used to train SVM in next step see Fig 7.

```
% Put the 13 features in an array
feat_disease = [Contrast,Correlation,Energy,Homogeneity,
                Mean, Standard_Deviation,Entropy, RMS,
                Variance, Smoothness, Kurtosis, Skewness, IDM];
% Classification Using SVM
```

Fig 7: Part of program showing the array of features set

IJERTV9IS060165
www.ijert.org
(This work is licensed under a Creative Commons Attribution 4.0 International License.)
278

D. *Classification Using SVM*: For image classification we first load the training set that contains two file the first being the matrix containing features of the disease affected leaves which is a matrix of M samples x 13(size of feature set) see Fig 8. The second contains the corresponding label which is the matrix of M rows x 1. We use SVM for classification. A support vector machine (SVM) is an algorithm of supervised learning type. Used for binary classification or regression. SVM are mostly popular for applications in natural language processing, speech and image recognition[9][10].
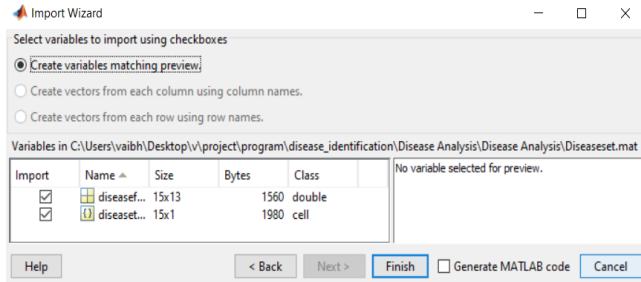
Fig 8: Feature values stored in diseaseset.mat file.

SVM is also called as kernel machines because they fall under a class of machine learning algorithms called kernel methods. Then we train the classifier with loaded training set using svmtrain().This produces a output that is saved and required for next function. Svmtrain-This function is used to train support vector machine classifier. Its syntax is as following SvmStruct=svmtrain (Training, Group); The Arguments being and explained as Training – It is matrix of the training set where each row corresponds to an observation and each column corresponds to a feature or variable.Group is a column vector and a character array or the cell array of strings for classifying data in the training into groups. It has same number of elements as there were number of rows in Training. Each element specifies the group to where the corresponding row in the Training belong. The next step is to classify the test leaf with the svmclassify() and the result is displayed on the command window. It takes two arguments that is svmstruct from svmtrain() and the array of thirteen feature values. Svmclassify is a function that classify the data using support vector machine algorithm. Its syntax is represented as Group = svmclassify (SVMStruct, Sample). This function classify each row of the data in the Sample using the information in a support vector machine classifier structure called SVMStruct, made using the svmtrain function. Sample must have the same number of columns as data that was used to train the classifier in svmtrain. Group indicates the group to which each of the row of Sample has been assigned for.

How to run disease identification program.
1. Place the 'Disease Detection' folder in your path which contains three required files namely my_program.m file, 'Disease Dataset' folder see Fig 9, Diseaseset.m file.
2. Run my_program.m
3. Select a leaf from the 'Disease Dataset' folder as shown in Fig.4, Disease Dataset folder containing image of infected leaves see Fig 10.
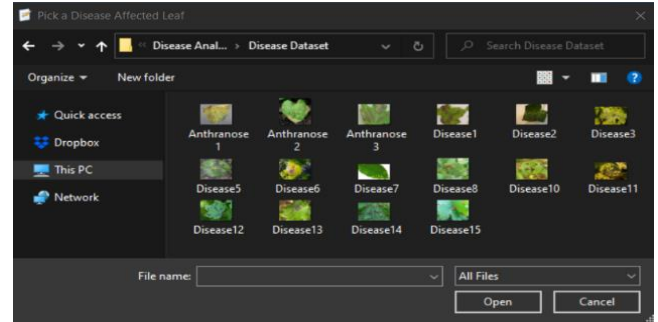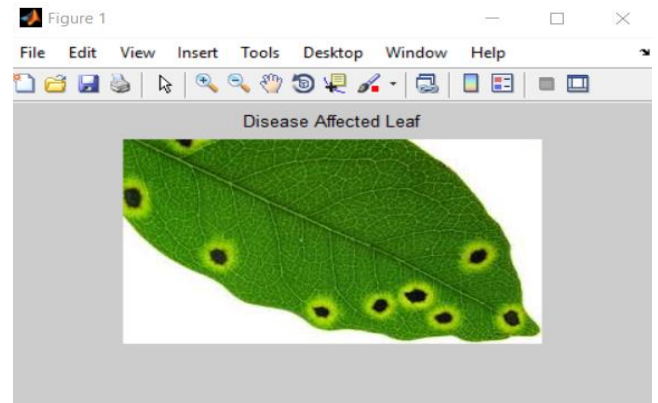
Fig 9: Disease dataset folder.

Fig 10: Selected Disease affected leaf.

4.Observe the results of K means clustering in the dialogue box as shown in Fig 11 and input the cluster image number that best matches the infected part of the leaf as shown in Fig.12. Here, in this case the first image best matches with disease affected leaf so we input value one.
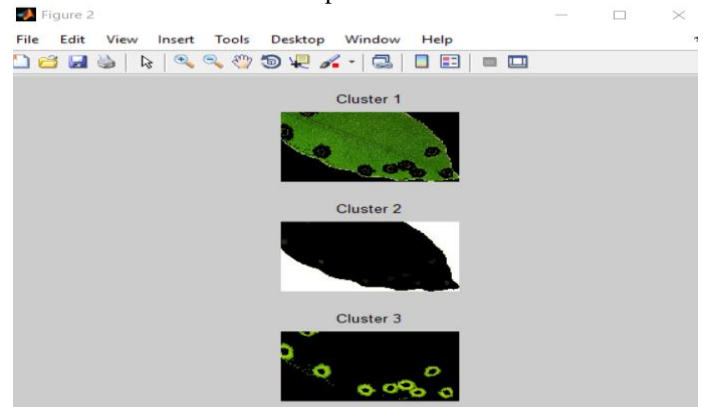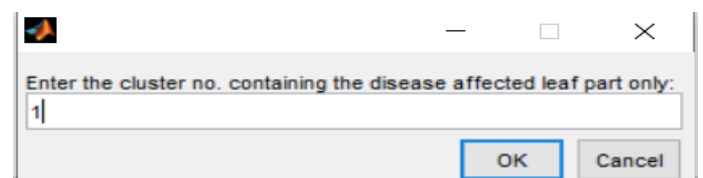
Fig 11: Images after k means clustering

Fig. 12: Entering matching cluster image.

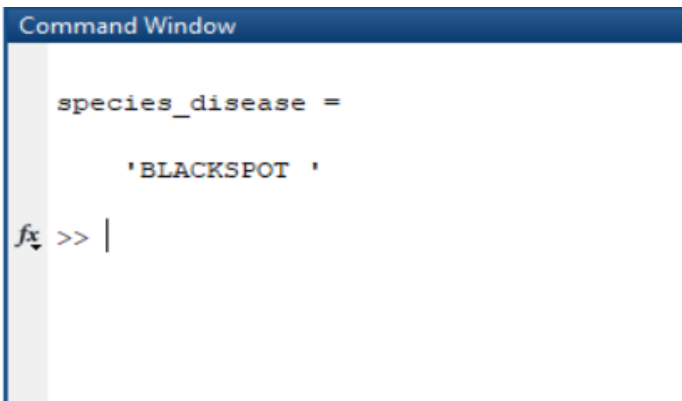5.Go on command window as shown in Fig 13 to know the disease name.

Fig 13: Result produced in command window.

Temperature and humidity sensor module

It's another part of system build with IR sensor, DHT11 sensor that is used to know temperature and humidity and a processing unit to operate the sensors and display result.

IV.   CONCLUSION

This project is a prototype design which implements the solution for irrigation system, field security, disease detection and distant monitoring. The IR sensor is used to detect any intruders. where the infrared rays emitting from the animals, humans etc. are received from the IR sensor and detects the intruder and the MATLAB code is used to detect the disease and all these applications will help the owner to monitor the field and owner gets the notification on his mobile through Wi-Fi which helps in distant monitoring. In future this prototype can also be programmed to predict preventive measures and cure for the diseased plant. The project which we have developed is just a prototype and which acts as a preliminary version of a device and where we can extend sensor for longer meters and a longer range of Wi-Fi module. So, this can be developed in real time application.

REFERNCES

[1]   "*Plant Pathogens*", By Waleed M. Abdulkhair and Mousa A. Alghuthaymi

[2]   "*Diseases Caused by Bacteria,Viruses, and Nematodes*", John P. Damicone, Lynn Brandenberger

[3]   Jennifer R. Aduwo, Ernest Mwebaze, "*Automated vision-based diagnosis of cassava mosaic diseases*", Virus Research, 2004.

[4]   Sanjay B. Patil and Shrikant K. Bodhe," *Leaf disease severity measurement using image processing*", IEEE.

[5]   R.C.Gonzalez, R.E.Woods, "*Colour Image Processing in Digital Image Processing*", 3rd ed., Pearson Education, Delhi, India, 2009.

[6]   "*Segmentation and Adaptive surface smoothing* ", IEEE, 2002

[7]   Ali, S. A, Sulaiman, N, Mustapha, "*K-means clustering to improve the accuracy of decision tree response classification*", Information Technology.

[8]   Deelers, S. and S. Auwatanamongkol, 2007, *Enhancing "K-means algorithm with initial cluster centres derived from data partitioning along the data axis with the highest variance*", Proc. World Acad. Sci. Eng. Technol., 26: 323-328.

[9]   Chang, C.-C. and C. J. Lin (2001), "*LIBSVM: a library for Support Vector Machines*". http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[10] Durgesh K. Srivastava, Lekha Bhambhu,"*Data classification using Support Vector Machine*".

[11] M Anupama Reddy, D Anjaneyulu, "*WSN in IOT environment interfacing with smart sensors using Arm7 with Zigbee for industries*", International Journal of Engineering Research and Development.

[12] Sanku Kumar Roy, Sudip Mishra, "A Prototype for agricultural intrusion detection wireless sensor networ*k*", Conference Paper, June 2015.

[13] Shen Weizhong, Wu Yachun, "Grading method of leaf spot disease based on image processing'', IEEE, 2008.

[14] Amritha A Joshi, B.D. Jadhav, "*Monitoring and controlling rice diseases using image processing techniques*", conference paper, May 2017.

[15] Prasad Babu, M. S. and Srinivasa Rao, (2010), "*Leaves recognition using back-propagation neural network - advice for pest and disease control on crops*", Technical report, Department of Computer Science & Systems Engineering, Andhra University, India. Downloaded from www.indiakisan.net on May 2010.

[16] J. K. Jain, R Rastogi, "*Application of image processing in Biology and Agriculture*", Nuclear India,1998.