Different Disease Detection in Cotton Leaf from Images taken in ungoverned environment

Honey Gadhiya, Ishika Agarwal, Nivedita Rao, Pooja Langhnoda

***Abstract*—The primary purpose of the project is to identify whether a particular cotton leaf is diseased or not and if it is diseased, identification of the disease is done. We have considered three main diseases - Alternaria Macrospora, Bacterial Blight and Grey Mildew. We are using Multi-SVM classification algorithm, to identify among different classes. The training data are images taken from web which are uploaded by ordinary humans. Assuming that the images are taken in uncontrolled environment, many difficulties are faced, making the machine-learning algorithm less effective. Therefore, some pre-processing like clustering, segmentation is done to improve the quality of learning. We have used various statistical features (mean, Standard deviation, variance etc.) as training features for our algorithm.**

***Index Terms*—Leaf Recognition, Disease detection, Feature Extraction, Classification, Disease Identification**

I. Introduction

In Developing countries like India, the economy is greatly dependent on agriculture. The quality and quantity of agricultural product is reduced because of plant disease.The naked eye observation of experts is the first approach for detection and identification of plant diseases. But, this requires constant monitoring of experts which might be expensive in large farms. And also farmers will have to go long to contact experts, this is time consuming and expensive for farmers. Therefore, automatic detection of plant diseases is an important topic to work upon, as it will reduce the farmer's work and will increase the quality and quantity of agricultural products.

There are various types of cotton leaf diseases such as Alternaria Macrospora, Bacterial Blight, Grey Mildew, Rust, Leaf Curl virus disease etc. Every year 20% to 25% of the cotton production is reduced due to diseases. Here we are detecting bacterial Blight, grey mildew and alternaria macrospora from images.These diseases are very much prevalent here in Gujarat. Grey mildew symptoms include Initial infection appears as triangular, square or irregularly circular whitish spots of 3 to 4 mm size on leaves. For Bacterial Blight the symptoms can appear on stems, leaves as well as bolls. But we will be considering only the leaves since only leaves belong in the domain of the problem. And as for the symptoms of Alternaria macrospora, it is primarily a leaf disease.Infection on cotyledons, leaves and bracts first appears as small, circular brown, grey-brown to tan spots which vary in size from 1-10 mm. A common trait in all the disease symptoms is occurrence of spots of various size and shapes and that is the primary focus for detection of disease in our project by extraction of the diseased area we will further on decide the disease through processing.

Dealing with real image there are overlapping image and dry leaf image which have some kind of similar colour with reference to background.When we are dealing with shape, Fourier descriptors are used to extract the best leaf properties. There are many ways to extract the diseased part, like extracting the color, shape, texture, edge detection,taking histogram, eigen feature extraction etc. But we will be focussing on only the diseased area by extracting only the diseased area.

Then, there are different classification techniques such as Support Vector Machine(SVM), Artificial Neural Network(ANN), Deep Neural Network (deep learning), Principal Component Analysis(PCA), K-means clustering,Otsu Segmentation etc. With these techniques and training data, we train the machine to identify the diseases from given image of cotton leaf.

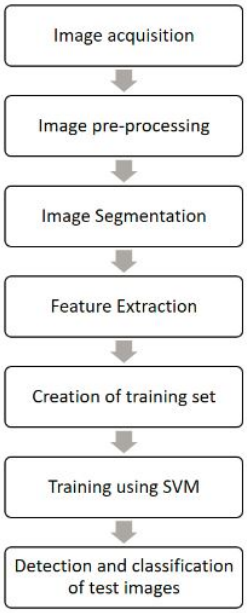
II. System Description

The basic approach towards disease identification that we have adopted is , first we have done pre-processing on images.Then image segmentation waas done.After that we extracted the features and created training set with that then gave it as an input to SVM classifier.As an output we got labels for each image.These labels are leaf diseases which we have considered.

1) Image acquisition - Image acquisition is basically creation of database to feed into the system for training as well as for testing and classification of images. We acquired our images online from various sources. The images are of bacterial blight, Alternaria macrospora , Grey mildew and healthy leaves. Healthy images are added into the dataset to identify both healthy and diseased images.

2) Image pre-processing - For preprocessing the images are converted into 256x256 pixels and which is later on transformed into CIElab color space.The Lab color space describes mathematically all perceivable colors in the three dimensions L for lightness and a and b for the color opponent's green–red and blue–yellow.

3) Image Segmentation - After the image preprocessing, the image is segmented into three parts using knn where k is



selected to be three. The separation is done on the basis of colors of the image. There are three primary color groups in an image which is diseased. These groups are :

- Soil part which is generally brownish in color
- The leaf part - which is the green area
- The diseased area comprising of sports or blobs which can range from light yellow/brown to grey in color depending up on the disease.

4) Feature Extraction - After feature extraction, only the diseased area is extracted and processing is done on it. There is a combination of 13 features used to classify the images into four classes - Alternaria macrospora, Bacterial Blight, Grey mildew and healthy leaves. These features are extracted only for the diseased part to identify the classes.

5) Creation of training set - With the help of the features that are extracted from the diseased are of the image a training set in the form of a .mat file is prepared which is used to later used as a training set to make the algorithm learn.

6) SVM Training : For disease detection using multi class SVMs, the feature vector is used as an input to a multiclass SVM system. Four classes were considered for the experiments,each one representing one of the diseases (Alternaria Macrospora,Bacterial Blight,Grey Mildew,Healthy Leaf).More specifically, the feature vectors are used as an input and the output of the SVM system is a label that classifies the grid deformation to one of the four diseases. For training we are giving two types of vectors (g1,l1),...,(gn,ln). Where gi are statistical feature vectors of each image and li are leaf disease labels of the feature vector. It constructs four disease rules, where it separates training vectors of one class from the rest of the vectors, by minimizing the objective function.

7) Detection and classification of test images - For detection of the disease some of the images from the data set itself are used for testing because of availability of a limited data set. Similar kind of preprocessing and image segmentation is performed here in this step too before feeding into the system for detection. A region is selected which consists the most of the diseased area and features from that area are used to input to the system and detection. In the end affected area percentage is also calculated to get an average idea about the stage of the disease. This percentage is calculated with reference to the affected area to the entire image.

III. Results

Images for the database for the processing were collected online from google and an image data set of our own was created. The images were belonging to either of the diseases i.e. Alternaria macrospora, Bacterial Blight , Grey mildew or healthy leaves so that the system was enabled to differentiate between the diseased and undiseased images. Later on completely rotten images which were withered off from the plants were also added to the data set but since all of the disease look pretty much the same after withering we removed it later on.

Initially we started off with 43 images with lesser noise in the background consisting of evenly distributed images from each class. These images belonged to the four basic classes as discussed above.The result accuracy we got with these images was almost 80%.

Later on addition of images with higher level of noise were added totalling to about 108 images. As a result the accuracy drastically dropped to about 55%. This was happening due to poor segmentation of the images which lead to confusing areas after seperation which did not clearly define the three areas that were meant to be separated. As a result an additional step in the preprocessing was added. In this step, the entire leaf was separated from the image first and then rest of the step were carried out. This seperation was done with the help of activecontour and creating a mask for the actual image to separate the leaf.


```
valyrian =

  Columns 1 through 15

     3     1     1     1     1     1     1     1     1     1     1     1     3     1     1

  Columns 16 through 30

     2     2     2     2     2     2     2     1     2     2     2     2     2     2     2

  Columns 31 through 43

     3     2     3     3     3     2     1     1     3     3     3     2     3
```

```
  Columns 1 through 15

     1     3     3     1     1     1     4     3     1     4     1     1     1     1     4

  Columns 16 through 30

     3     1     1     4     1     1     1     1     2     1     1     3     1     1     3

  Columns 31 through 45

     1     1     3     4     1     1     3     1     1     1     1     1     1     1     3

  Columns 46 through 60

     1     1     1     1     2     4     3     1     3     1     1     1     1     1     1

  Columns 61 through 75

     1     1     1     1     1     1     1     3     1     1     3     1     3     1     1

  Columns 76 through 90

     4     1     4     1     1     3     3     3     3     1     3     3     3     3     3

  Columns 91 through 105

     3     3     3     3     4     3     4     4     1     4     4     1     4     1     4

  Columns 106 through 108
```

Future work - Completion of implementation using neural network to identify the pests. Inclusion of more diseases for identification. Secondary classification of stages for the diseases, not just a vague idea using the area.

IV. Literature Review

Aditya Parikh, Mehul S. Raval, et al.[1] describes the approach to detect disease and identify its stage of cotton plant.In cotton, diseases in leaf are critical issue because it reduces the production of cotton. The region of interest is leaf because in cotton plant, most of diseases occur in leaf only.The identification is done from unconstrained images of cotton leaf.The detection is done after some pre-processing is done on images.The algorithm implemented for detection can be used for any specific disease.

Shayan Hati,Sajeevan G[2] describes how Artificial Neural Network is used to identify plant by inputting leaf image. Compared to earlier approaches, new input features and image processing approach that matter in efficient classification in Artificial Neural Network have been introduced. Image processing techniques are used to extract leaf shape features .These extracted features are used as inputs to neural network for classifying the plants.

V. R. Patil,R. R. Manza[3] describes how features are extracted from leaf images by various image processing techniques.Then digital morphometric feature extraction is discussed.The there are some experimental results of some techniques implemented.

Xiaodong Zheng,Xiaojie Wang[4] describes how leaf features play an important role in plant species and plant identification like leaf vein is an important morphological feature of leaf.The gray-scale morphology processing is applied to the image to eliminate the color overlap in the whole leaf vein and the whole background. Then linear intensity adjustment is adopted to enlarge the gray value difference between the leaf vein and its background. Calculate a threshold with OSTU method to segment the leaf vein from its background. Finally, the leaf vein can be got after some processing on details.

Sushma S. Patil,Mr. Suhas K. C[5] describes the symptoms of some cotton leaf diseases and then an algorithm is proposed to detect different diseases using Support Vector Machine(SVM). This algorithm contains some of image processing steps such as resizing, converting RGB to grayscale,detecting edges of spots,segmentation,features extraction and then SVM is applied to trained dataset.

V. Discussion

Basically, first we are doing clustering for each and every image, to identify the region of interest(ROI). After manually selecting one cluster among three options, we are extracting feature values (for now, 13 training features are considered), and storing it in an excel sheet. Thus, making a 2-dimensional array of size 48x13. Then, a new diseased/healthy leaf image is given in algorithm, and then it gives the output as label, depicting the type of disease. But, we are facing many problems because of unconstrained images.

Firstly, distinguishing between two similar diseases like alternaria macrospora and bacterial blight, which shows similar symptoms of brown spots. The difference is their spread pattern, one has diseased brown area across veins, while other has multiple small spots. Third disease, namely grey mildew is quite different because it has white colour spots for diseased section. So, to have distinguished classification between these two diseases is a challenge.

Secondly, there are some problems in the database itself like - background colour similar to diseased part colour, human hand in the image, contrast or brightness of images, resolution, shape etc. To assure our algorithm to work despite these difficulties, is another challenge.

REFERENCES

[1] Aditya Parikh, Mehul S. Raval,Chandrasinh Parmar and Sanjay Chaudhary Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images In IEEE International Conference on Data Science and Advanced Analytics (DSAA),IEEE, 2016.

[2] Shayan Hati,Sajeevan G Plant Recognition from Leaf Image through Artificial Neural Network In International Journal of Computer Applications (0975 – 8887) Volume 62– No.17, January 2013.

[3] V. R. Patil,R. R. Manza A Method of Feature Extraction from Leaf Architecture In International Journal of Advanced Research in Computer Science and Software Engineering 2015

[4] Xiaodong Zheng,Xiaojie Wang. Leaf Vein Extraction Based on Gray-scale Morphology In MECS, December 2010.

[5] Sushma S. Patil,Mr. Suhas K. CIdentification and Classification of Cotton Leaf Spot Diseases using SVM Classifier In International Journal of Engineering Research & Technology (IJERT) April 2014.