Pest and Disease Identification in Cotton Plants

Charmi Kalani: 1401067, Divya Patel: 1401030, Himani Patel: 1401111, Rupam Jogal: 1401003
School of Engineering And Applied Sciences
Ahmedabad University

***Abstract*—Aim of the project is to determine stage of disease Gray Mildew in cotton plants which is widely prevalent fungal disease in Gujarat. We have used some images to train the model. The main objective is image preprocessing. The reason behind that is images can be captured under uncontrolled conditions in field and can be captured by mobile phone or camera by an untrained person. Leaf diseases on cotton plant can easily be identified early and accurately as it can prove detrimental to the yield. Here the problem is of classification of disease stages on cotton leaf images. For that we can apply Nearest Neighbour Classifier. Using machine vision techniques, it is possible to increase scope for detection of various diseases within visible as well invisible wavelength regions. Using local statistical features, rst classier segments leaf from the background. Then using Hue and Luminance from HSV colour space another classier is trained to detect disease and nd its stage.The algorithm is a generalized as it can be applied for any disease.**

Keywords:Nearest Neighbourhood Classifier, Leaf diseases

I. INTRODUCTION

Cotton is a principal cash crop in India and affects Indias economy in many ways. Large number of people depends on Cotton crop either by its cultivation or processing. The cotton productivity become more mission oriented in recent times due to increasing domestic demand. Automatic detection of plant diseases is an essential research topic as it may be advantageous in monitoring huge fields of crops, and detect the symptoms of diseases as soon as they appear on plant leaves. Therefore the need for fast, automatic, less expensive and accurate method to detect plant disease cases is of great significance. So, using machine learning techniques we can detect the diseases in early stages,so we can reach our goal.There are many challenges in identifying diseases.The cotton diseases are classied as: 1) Bacterial, e.g. Grey Mildew, Bacterial Blight, Crown Gall; 2) Fungal e.g. Anthracnose, Leaf spot, Alternaria; 3) Viral, e.g. Leaf curl, Leaf crumple, Leaf roll; 4) Pest, e.g. White ies, Pink ball worm.There is a production loss due to diseases.

So there came the need for automation in detection of diseases in plants.And here we trying to focus on a single disease that is Grey Mildew and we identified the 3 stages of the disease with machine learning approach.

II. LITERATURE REVIEW

Though diseases are mostly seen on the leaves of plant, precise quantification of these visually observed disease traits has not well studied yet because of the complexity of visual patterns.Hence there has been increasing demand for more specific and sophisticated image pattern understanding techniques.

Computer aided detection of plant diseases is an important research topic as it may prove benefits in detecting diseases automatically from the symptoms that appear on the plant leaves in early stages.In a machine vision based solution, image is segmented into its constituents objects, suitable features are extracted from the foreground and disease region is classified. The system proposed intends at processing the images captured in natural conditions from varying distances.

Some of the other approaches dealing with overlapping leaves are as follows:

- Authors in other paper developed an algorithm to extract the leaf from a plant image automatically in a controlled condition.
- They used machine vision to segment leaves from the background, extract leaf curvature and identify the tree species.
- The algorithm runs a binary classifier to first segment leaf and non-leafy part.

There are various research papers for Disease Identification in the plants. A few of the papers have been reviewed.

1) Cotton Leaf Disease Identification using Pattern Recognition Techniques by Viraj A. Gulhane and Maheshkumar H. Kolekar
   It gives idea about problem of diagnosis of diseases on cotton leaf using **Principle Component Analysis (PCA), Nearest Neighbourhood Classier (KNN)**. After implementing PCA/KNN multi-variable techniques, it analyse the statistical data related to the Green (G) channel of RGB image since disease are reected well by green channel. In most of the cases diseases are seen on the leaves of the cotton plant such as Blight, Leaf Nacrosis, Gray Mildew, Alternaria, and Magnesium Deciency. The classication accuracy of PCA/KNN based classier observed is 95 percent.
2) Cotton Leaf Disease Identification using Pattern Recognition Techniques by P. R. Rothe and R. V. Kshirsagar
   It discusses about a pattern recognition system for identification and classification of three cotton leaf diseases i.e. Bacterial Blight, Myrothecium and Alternaria. Active contour model is used for image segmentation and Hus moments are extracted as features for the training of adaptive neuro-fuzzy inference system. The

classification accuracy is found to be 85 percent.

3) Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification by Srdjan Sladojevic, Marko Arsenovic,Andras Anderla, Dubravko Culibrk, and Darko Stefanovic
This paper is concerned with a new approach to the development of plant disease recognition model, based on leaf image classification, by the use of **deep convolutional networks**.It is able to recognize 13 different types of plant diseases out of healthy leaves, with the ability to distinguish plant leaves from their surroundings and deep CNN training was performed. The experimental results on the developed model achieved precision between 91 percent.

III. KNN ALGORITHM

The model representation for KNN is the entire training dataset. It is as simple as that. KNN has no model other than storing the entire dataset, so there is no learning required.

When KNN is used for classification, the output can be calculated as the class with the highest frequency from the K-most similar instances. Each instance in essence votes for their class and the class with the most votes is taken as the prediction.

Class probabilities can be calculated as the normalized frequency of samples that belong to each class in the set of K most similar instances for a new data instance. For example, in a binary classification problem (class is 0 or 1)

- p(class = 0) = count (class = 0) / (count(class = 0)+ count(class = 1))

If you are using K and you have an even number of classes (e.g. 2) it is a good idea to choose a K value with an odd number to avoid a tie. And the inverse, use an even number for K when you have an odd number of classes.

Ties can be broken consistently by expanding K by 1 and looking at the class of the next most similar instance in the training dataset.

IV. ALGORITHM

First We Model the data using two techniques.

The algorithm is two - step process. 1. Part (i) We segment the leaf into two class foreground and background. Foreground is cotton plant leaves, while Background is non leafy part of the cotton plant including soil.

2. This foreground is an input to step two classifier, which segments image into the diseased class and the non-diseased class.

After segmentation in step two, next task is to find the severity of the disease which is defined by a formula.

In the first step, after doing foreground and background separation we apply KNN using fitcknn to classify the model.

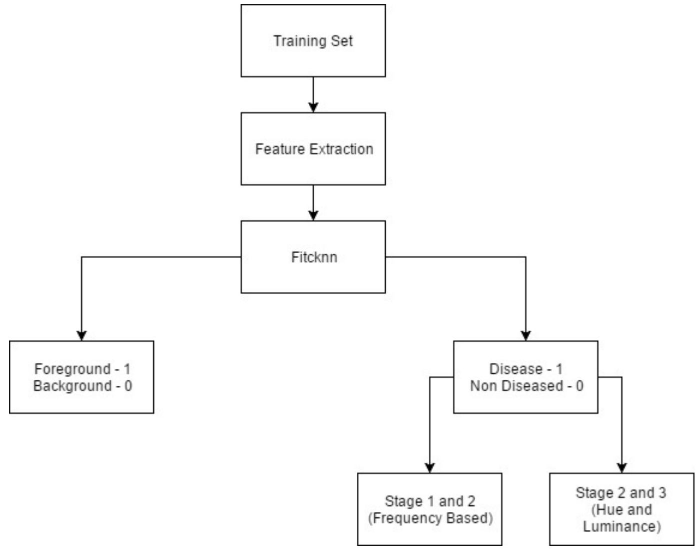

Fig. 1: Flow Chart

We have assign a class 1 to foreground and class 0 to background.

Now the class 1 having foreground images are again classified using Step 2. i.e Binary Classification of Foreground images. Foreground images are classified into Disease and Non-Disease. We are assigning 1 to Disease and 0 to Non-Disease.
Now Disease is classified into Stage 1 and Stage 2. Frequency based on KNN with the 15 features. Images with having More white spots are classified into Stage 1 and Stage 2.
Stages 2 consists of images of leafs having more fungus which are white due to decaying. Stages 3 consists of images of leafs that are brown.

V. MAIN RESULT

We have Tried to segment the affected part of leaf from image using training model computed before. Ideally if there is given a fresh leaf(figure 4 and 5) then there is not any sign of disease so severity is zero and whole region is black.

In Second trial we have taken a leaf with disease at third stage. And algorithm is detecting the affected part of that leaf. Based on comparison of figure 6 and 7 we can say that the estimation of severity is 0.56.

VI. CHALLENGES

1) Color of the background same as the color of the diseased part.

No.	Table 1 Features
1.	Mean of red channel
2.	Mean of green channel
3.	Mean of blue channel
4.	Standard deviation of red channel
5.	Standard deviation of green channel
6.	Standard deviation of blue channel
7.	Contrast = $\sum_{i,j} \|i-j\|^2 \ p(i,j)$
8.	Correlation = $\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)(p(i,j))}{\sigma_i \sigma_j}$
9.	Energy = $\sum_{i,j} p(i,j)^2$
10.	Homogeneity = $\sum_{i,j} \frac{p(i,j)}{1+\|i-j\|}$
11.	Entropy = - $\sum_i P_i \ log_2 P_i$
12.	Gradient of the image
13.	Range of hue component
14.	Mean(Maximum value - other values)
15.	Mean of Luminance

Fig. 2: Table for the features that are considered.

$$Severity = \frac{\sum_{i,j} P(i,j)}{T(i,j)} * 100$$

2) Bright as well as dark pictures were difficult to process.
3) Low quality images won't let detect very small patches of disease.

VII. FUTURE WORK

- To improve the calculation of severity for better stage classification of the disease.
- Improve the results with the implementation of neural network to identify other diseases and its stages.

VIII. REFERENCES

1) Parikh, M. S. Raval, C. Parmar, and S. Chaudhary, Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images, 2016.

2) V. A. Gulhane, Diagnosis of Diseases on Cotton Leaves Using Principal Component Analysis Classifier, 2014.

3) E. Engineering, Cotton Leaf Disease Identification using Pattern Recognition Techniques, vol. 0, no. c, 2015.

Algorithm 1 Decision for stage selection

Find the severity with the stage two and the stage three training set.
if $S > Th$ **then**
  if S calculated from stage three training set **then return** Stage three disease
  else
    if $10\% < S < 30\%$ **then return** Stage one disease
    elseStage two disease
    end if
  end if
else
  No disease
end if

Fig. 3: Decision for stage selection

Fig. 4: Original Image

Fig. 5: Severity : 0

Fig. 6: Original Image

Fig. 7: Severity :0.56