Cotton Crop Disease Detection using Image Processing and Machine Learning

Vishwajeet Gawali[1], Rohit Suryawanshi[2], Nitin Sonawane[3], Prof.Pranjali Kuche[4], Prof. Pranjali Kuche[5]
[1234]*BE Student*, [5]*Professor,*
[12345] *Department of Information Technology Marathwada Mitra Mandal Collage of Engineering , Pune Maharashtra, India*

Abstract- India is an agriculture country and above seventy percent of our population depends on the agriculture. One-third of our national income comes from agriculture. Agriculturalist are facing loss due to various crop disease and it becomes tedious to cultivators to monitor the crop regularly when the cultivated area is huge(acres).So the plant disease detection plays an very important role in agriculture field. Timely and accurate disease detection is important for the loss caused due to crop diseases which affects adversely on crop quality and yield. Early diagnosis and intervention can reduce the loss of plant disease and reduce the unnecessary drug usage. Earlier, automatic detection of plant disease was performed by image processing. For disease detection and classification we are proposing machine learning mechanisms and image processing tools. Crop disease will be detected through various stages of image processing such as image acquisition, image pre-processing, image feature extraction & feature classification. For image feature extraction we will be use image global feature extraction technique.

Keywords- Image Processing, Machine Learning, Feature Extraction, Image Global Features, Classification.

I. INTRODUCTION

Farmer's economic growth relies on the quality of the product that they grow, which is directly dependent on the plants growth and yield they get. Plants are attacked by the different disease which target different parts of plant body such as leaf, stem, seed, and fruit and so on. To solve this problem machine learning seems to be a better option various machine learning technique are recently proposed for identification and classification of plant disease from plant images. Cotton is one of the most important cash crops of India and plays a dominant role in the industrial and Agriculture Economy of the country. It provides basis raw material (cotton fiber) to cotton textiles industry. Cotton in India provides direct livelihood to 6 million farmers and about 40-50 million people.

Various image processing concepts such as image filtering,segmentation, image feature extraction have emerged to detectthe leaf diseases. There are various image segmentationmethods available such as k-means clustering, Canny andSobel segmentation, and Otsu thresholding. Techniques such asSupport Vector Machine (SVM), Neural Network (NN), andHomogeneous Pixel Counting technique for Cotton DiseasesDetection (HPCCDD) can be used for classification. Featuresplay an important role in the classification process. Previousproposed works for detecting disease has some limitations such as low resulting accuracy and less number of images used to detect disease.Themain source for the disease is the leaves of the cotton plant. About 80 to 90 % of disease onthe cotton plant is on its leaf. So f our study of interest is the leaf of the cotton tree rather thanwhole cotton plant the cotton leaves is mainly suffered from diseases like insecticide(tudtude,mawa) fungus, Foliar leaf on leaf of cotton, Alternaria leaf spot of cotton. The machine vision system now a day is normallyconsists of computer, digital camera and application software. Various types of algorithms are integrated in the application.Image processing is one important method that helps segment image into objects and background image. One of the key steps in image analysis is feature detection. Image recognition has attracted many researchers in the area of pattern recognition, similar flow of concept are applied tothe field of pattern recognition of plant leaf, that is used in diagnosing the cotton leaves diseases. There are numerous methodshave been proposed in the last two decades which are not fully solved. However this is challenging problems. The critical issueis how to extract the discriminative and stable feature for classification.

II. RELATED WORK

The primary focus of this work is to detect diseaseand estimate its stage for a cotton plant using images. Mostdisease symptoms are reflected on the cotton leaf. The proposed work uses two cascaded classifiers, so using local statistical features, first classifier segments leaf from the background. Then using hue and luminance from HSV color space another classifier is trainedto detect disease and find its stage. The developed algorithm isa generalized as it can be applied for any disease [1].

This work presents asurvey on detection and classification of cotton leaf diseases. It isdifficult for human eyes to identify the exact type of leaf diseasewhich occurs on the leaf of plant. Thus, in order to identify thecotton leaf diseases accurately, the use of image processes and, machine learning techniques can be helpful. The images used forthis work were acquired from the cotton field using digitalcamera. In pre-processing step,

background removal technique isapplied on the image in order to remove background from theimage. Then, the background removed images are furtherprocessed for image segmentation using otsu thresholdingtechnique [2].

Leaf diseases on cotton plant must be identified early and accurately as it can prove detrimental to the yield. The presented work presents a pattern recognition system for identification and classification of three cotton leaf diseases i.e. Bacterial Blight, Myrothecium and Alternaria. The images required for this work are captured from the fields at Central Institute of Cotton Research Nagpur, and the cotton fields in Buldana and Wardha district. Active contour model is used for image segmentation and Hu's moments are extracted as features for the training of adaptive neuro-fuzzy inference system [3].

Farming is important sector in India for human being, as near about 55-60% people are depends directly andindirectly on it. Among all crops, Cotton is main cash crop in India gives more income to the farmer. Due to diseases oncotton there may be chances of decrease in production and drastic change is occurred on crop. The fungal diseases likeVerticilium Wilt, Bacterial blight, Red spot, Alternaria, Downy Mildew are responsible for production loss. So, this work presents various types of diseases and control on it using image processing technique. The comparative study of artificialneural network, Support vector machine is discussed [4].

In this work, author present survey on the various types of leaf diseases in plants and their identification process. An identification problem deals with associating a given input pattern with one of the distinct classes. Plant leaf disease identification is a technique where leaf spot disease is identified based on its different morphological features. There are various successful identification techniques like Probabilistic Neural Network, Genetic Algorithm, Back Propagation Neural Network and Principal Component Analysis (PCA). Deciding on the method for identification is often a difficult task because the quality of the results can be varying for different input data. Plant leaf disease identification has wide applications in the field of Agriculture to increase the productivity [5].

In this work, author aim is to solve the cotton disease detection problem using the image processing techniques automatically from the input image. The disease classification will primarily based upon the visibility of the disease on the cotton leaves, which further can be used for the identification using the classifier. The proposed model implementation would be done using the MATLAB simulator and the proposed model results would be obtained in the form of the accuracy, precision, recall, elapsed time and many other similar parameters [6].

In this workauthor express Technological Strategies uses mobile captured symptoms of Cotton Leaf Spot images and categorize the diseases using support vector machine. The classifier is being trained to achieve intelligent farming, including early detection of disease in the groves, selective fungicide application, etc. This proposed work is based on Segmentation techniques in which, the captured images are processed for enrichment first. Then texture and color Feature extraction techniques are used to extract features such as boundary, shape, color and texture for the disease spots to recognize diseases [7].

This work addresses the problem of diagnosis ofdiseases on cotton leaf using Principle Component Analysis(PCA), Nearest Neighborhood Classifier (KNN). Cotton leafdata analysis aims to study the diseases pattern which are definedas any deterioration of normal physiological functions of plants,producing characteristic symptoms in terms of undesirable colorchanges mainly occurs upon leaves; caused by a pathogen,which may be any agent or deficiencies. The predictions ofdiseases on cotton leaves by human assistance may be wrongin some cases. Using machine vision techniques, it is possible toincrease scope for detection of various diseases within visible aswell invisible wavelength regions [8].

III. PROPOSED METHODOLOGY

The methodology for diagnosing cotton leaf diseases involves several tasks, such as Image acquisition, image preprocessing, image feature extraction and cotton leaf diseases classification based on image feature that is color features, shape features and texture features.The first phase is the image acquisition phase. In this step, image is uploaded from the images of the variouscotton leaves dataset. In the second phase image preprocessing is completed. In the third phase, image feature extraction for the infected part of the leaf is completed based on specific properties among pixels in the image or their texture. After this step, certain statistical analysis tasks are completed to classify the features that represent the given image using machine learning to compare image features. Finally, classification result shows the identified cotton leaf disease.

Proposed System Advantages:

1. It consist two algorithms for classification and feature extraction which effectively able to extract disease from image and gives the actual final result.

2. This proposed system effectively able to extract all the spatial characteristics of an image.

Proposed System Architecture:

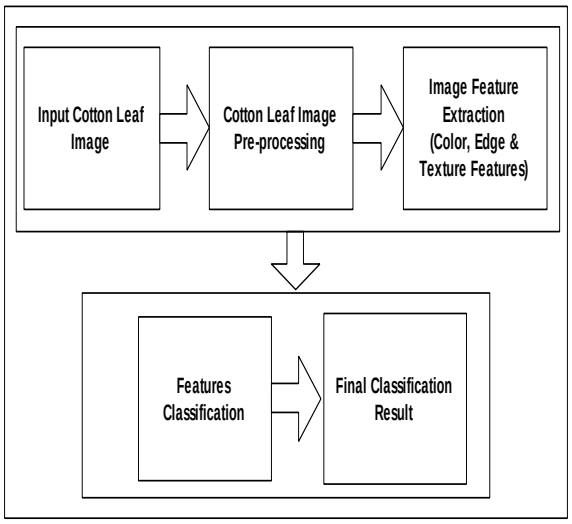

Fig.1: Proposed System Architecture

IV. METHODOLOGY

1. Feature Extraction for Disease Identification:

Color feature: As most of the color distribution information can be captured by the low-order moments, using only the first three moments: mean, variance and skewness, it is found that these moments give a good approximation and have been proven to be efficient and effective in representing the color distribution of images (Stricker and Orengo 1995).

Edge Detection: Most of the shape information of an image is enclosed in edges. So first we detect these edges in an image and by using these filters and then by enhancing those areas of image which contains edges, sharpness of the image will increase and image will become clearer.

Texture feature: Describes the structure arrangement of surfaces and their relationship to the environment, such as fruit skin, clouds, trees, and fabric. The texture feature in our method is described by hierarchical wavelet packet descriptor (HWVP). A 170- D HWVP descriptor is utilized by setting the decomposition level to be 3 and the wavelet packet basis to be DB2.

2. Support Vector Machine

- Support Vector Machine (SVM) is used to classify the fruit quality. SVM Support vector machines are mainly two class classifiers, linear or non-linear class boundaries.
- The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.
- The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class.

V. MATHEMATICAL MODEL

Cotton Leaf Image Feature Extraction using Global Features

1. Mathematical Equations of Color Feature Extraction Method

The color distribution information can be captured by the low-order moments, using only the first three moments: mean, variance and skewness, it is found that these moments give a good approximation and have been proven to be efficient and effective in representing the color distribution of. These first three moments are defined as:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{ij} - \mu_i)^2}$$

$$S_i = \left[\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - \mu_i\right)^3\right]^{\frac{1}{3}}$$

Where, Pij is the value of the i^{th} color channel of the j^{th} image pixel. Only 3 x 3(three moments for each color component) matrices to represent the colorcontent of each image are needed which is a compact representation compared to other color features.

2. Mathematical Equations of Canny Edge Detector Method

Step1: Smooth the image with a Gaussian filter to reduce noise and unwanted details and textures.

$$g(m,n) = \square_{\square}(m, n) * f(m, n)$$

Where

$$\square_{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{\square^2 + \square^2}{2\sigma^2}\right)$$

Step2: Compute gradient of g (m, n) using any of the gradient operations (Roberts, Sobel, Prewitt, etc) to get:

$$M(n, n) = \sqrt{\square_{\square}^2(\square, \square) + \square_{\square}^2(\square, \square)}$$

And

$$\square(m, n) = \square\square\square^{-1}[\square_{\square}(m, n) / \square_{\square}(m, n)/]$$

Step3: Threshold M:

$$\square_{\square}(m, n) = \begin{cases} \square(\square, \square) \ \square\square \ \square(\square, \square) > \square \\ 0 \qquad\qquad \square\square\square\square\square\square\square\square\square \end{cases}$$

3. Mathematical Equations of Texture Feature Extraction Method

According to co-occurrence matrix, there are several textural features measured from the probability matrix to extract the characteristics of texture statistics of remote sensing images. Correlation measures the linear dependency of grey levels of neighboring pixels.

$$\text{Correlation} = \frac{\sum_{\square=0}^{\square-1}\sum_{\square=0}^{\square-1}(\square,\square)\square\,(\square,\square)-\square_{\square}\,\square_{\square}}{\square_{\square}\square_{\square}}$$

4. The Mathematics of the Support Vector Machine

We have k sub-spaces so that there are k classification results of sub-space, called CL_SS1, CL_SS2,...,CL_SSk. Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$CL = \frac{1}{\square}\sum_{\square=1}^{\square}\square\square\_\square\square_{\square}$$

Or weighted mean value:

$$CL = \frac{1}{\square}\sum_{\square=1}^{\square}\square_{\square}\square\square\_\square\square_{\square}$$

Where w_i is the weight of classification result of subspace SSi, and satisfies:

$$\sum_{\square=1}^{\square}\square_{\square} = 1$$

The centroid of a leaf is calculated as follows:

$$\bar{X} = \frac{\sum_{i=0}^{k} xi}{k}, \bar{Y} = \frac{\sum_{i=0}^{k} yi}{k}$$

Where $(\bar{X}, \bar{Y})$ represents the centroid of the leaf, xi and yi are x and y coordinates of the i^{th}pixel in the infected region and k denotes the number of pixels that represent only the leaf portion.

In the next step, the distance between the centroid and the pixel value was calculated. For distance, the following Euclidean distance was used:

$$\text{Distance} = \sqrt{(\square 2-\square 1)^2-(\square 2-\square 1)^2}$$

Where (x1, x2) and (y1, y2) represent the two co-ordinate values.

VI. RESULT

We compared the proposed disease recognition accuracy on number of samples and show the result graphically. Let see the following graph andtable shows the disease recognition accuracy result of Alternaria leaf disease, Cucumber disease and foliar disease respectively.

The overall accuracy of Alternaria leaf disease, Cucumber disease and foliar disease are 95%, 92% and 91.3% resp. So our proposedsystem recognition accuracy is better than existing system.

So this works gives better recognition result compare toexisting method.

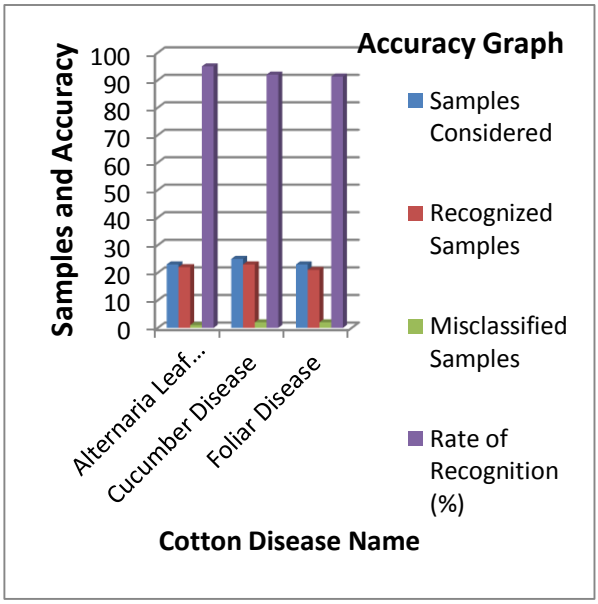

Fig.2: Accuracy Graph

Table: Recognition Accuracy

Cotton Disease	No. of Samples	Recog. Samples	Miss classified Samples	Rate of Recog. (%)
Alternaria Leaf Blight	23	22	1	95
Cucumber Disease	25	23	2	92
Foliar Disease	23	21	2	91.3

VII. CONCLUSION

In this work, addressed how the disease analysis is possible for the cotton leaf diseases detection, the analysis of the various diseases present on the cotton leaves can be effectively detected in the early stage before it will damage the whole plant. Here the technique presented can able to detect the disease more accurately. Mainly focuses on the plant disease detection and through the application of various methodologies.Usage of various features extraction technique and a stable, sufficient data set have facilitated in obtaining satisfactory experimental

result.The usage of classification and feature extraction processes has enhanced the performance of the system which provides better results.

VIII. REFERENCES

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

[2]. Bhumika S.Prajapati Vipul K.Dabhi Harshadkumar B.Prajapati, "A Survey on Detection and Classification of Cotton Leaf Diseases", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2016.

[3]. P. R. Rothe, R. V. Kshirsagar, "Cotton Leaf Disease Identification using Pattern Recognition Techniques". International Conference on Pervasive Computing (ICPC)

[4]. Naik Durgesh Manikrao, Dr. Prof. A.J.Vyavahare, "Disease Detection of Cotton crop using Image Processing Technique: A Survey". Volume 3 Issue VI, June 2015

[5]. Bed Prakash, Amit Yerpude, "A Survey on Plant Leaf Disease Identification". Volume 5, Issue 3, March 2015

[6]. Kapil Prashar, Rajneesh Talwar, Chander Kant, "A Review on Efficient Identification of American Cotton Leaf Diseases through Training Set". International Journal of Computer Applications (0975 – 8887) Volume 132 – No.7, December2015

[7]. Prof. Sonal P. Patil, Ms. Rupali S.Zambre, "Classification of Cotton Leaf Spot Disease Using Support Vector Machine". Ms. Rupali S.Zambre et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 4, Issue 5(Version 1), May 2014, pp.92-97

[8]. Viraj A. Gulhane, Maheshkumar H. Kolekar, "Diagnosis of Diseases on Cotton Leaves Using
Principal Component Analysis Classifier", 2014 Annual IEEE India Conference (INDICON)