

SVM-based Classifier System for Recognition of Cotton Leaf Diseases

P.R. Rothe and Dr. R. V. Kshirsagar
Department of Electronics Engineering,
Priyadarshini College of Engineering, Nagpur, India

Abstract: The allocation and recognition of cotton leaf diseases are of the major importance as they have a cogent and momentous impact on quality and production of cotton. This paper presents a *modus operandi* for automatic classification of cotton leaf diseases through feature extraction of leaf symptoms from digital images. Otsu's segmentation method is used for extracting color and shape features. Support vector machines (SVM) had been used to do classification on the extracted features. Three diseases have been diagnosed, namely Bacterial Blight, Myrothecium and Alternaria. The testing samples of the images are gathered from CICR Nagpur, cotton fields in Buldhana & Wardha district.

Keywords: Unsharp filtering, Otsu's segmentation, Support vector machine, Classification.

I. Introduction

Cotton is an important cash crop in India. Cotton plants suffer from multifarious diseases that can prove detrimental to the quality and quantity of the yield. In Andhra Pradesh state government reported that 20.46 lacs farmers suffered from cotton crop failure in 33.73 lacs acres and lost Rs.3071.6 crores. In Maharashtra the estimated loss is about Rs.10000/- crores in recent years.

Mainly the detection and identification of leaf diseases can be done by naked eye observation[1]. This in turn requires continuous monitoring which is not viable in large farms. Currently, the detection of crop diseases is mostly depend on manual recognition, but it can be erringly diagnosed by farmers because they usually judge the symptom by their experiences. The misidentification leads to some erroneous control measurements such as desultory and untimely use of pesticides.

Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring and supervising large crop fields and early detection of the symptoms of diseases as soon as they appear on plant leaves [2;3]. Such systems include image enhancement, image segmentation, feature extraction and training blocks.

Usually the detection and identification of cotton leaf diseases is done by naked eye observation which leads to some erroneous control measurements. This leads to requirement of automatic system that can identify the leaf disease early and accurately. The proposed system will help identification of three types of leaf diseases namely Bacterial Blight, Myrothecium and Alternaria.

II. Cotton Leaf Diseases

In this paper the leaf diseases considered are Bacterial Blight, Myrothecium and Alternaria. The symptoms of these diseases are as:

Bacterial Blight 	The appearance of many minute water soaked angular spots scattered on the ventral surface of leaves, these spots turn brown and then are converted into black dead lesions on both sides of the leaf. Many a time one or more veins in part or whole become water soaked and then turn brown or black, this symptom is called as "Black vein" or "Vein blight". It may appear independently or together with angular leaf spots.
Myrothecium 	The appearance of circular to semicircular light brown to tan colored spots with violet to reddish brown margins. Later on shield shaped small fruiting bodies are produced in the central portion of the spot. Big shot holes appear in leaves as the center dries and drops off.
Alternaria 	In early stage of infection pale green spots with irregular margins develop on leaves. Later on, these spots enlarge and concentric rings and cracked centre develops. Severe infection causes drying and falling of leaves.

III. Literature Review

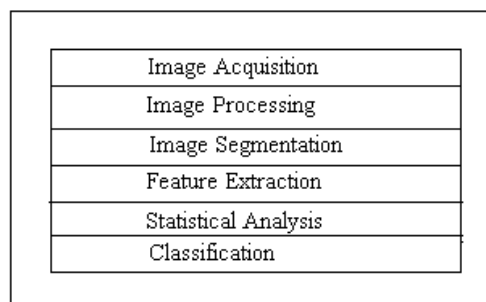
Li, Gao and Shen [18] had proposed the automatic identification system to detect powdery mildew, sharp eye spot and stripe rust diseases in wheat. The parameters such as sphericity, roundness, Hu1, Hu2 and equivalent radius were selected for identification. Song kai & Liu zhikun [15] proposed the maize disease image recognition system for Corn. It uses YCbCr color space technology to segment disease spot and uses the co-occurrence matrix spatial grey level layer to extract disease spot texture feature, and uses BP neural network for classification. The accuracy was as high as 98%.

A.A. Bernardes & J.G. Rogeri suggested automatic classification of cotton diseases namely Ramularia, Bacterial Blight and Ascochyta Blight using Wavelet transform and SVM. The accuracy was between 71.4% to 97.1% [16].

H. Al-Hiary, S. Bani-Ahmad, M. Reylat [17] proposed automatic system for detection and classification of plant diseases. They used K-means clustering technique for segmentation and back propagation algorithm for classification to get the average efficiency of 94.67%.

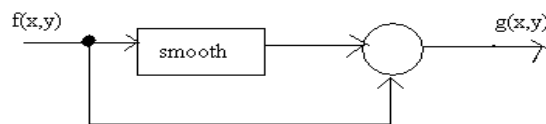
IV. The Proposed Approach

The all-embracing concept for any vision related approach for image classification remains the same. Initially the digital images are procured using a digital camera. Then image processing techniques are applied to these images to extract features. The useful features are used to train the network which performs the classification.



V. Image Enhancement

The images required for the experimentation are acquired by using Cannon A460 digital camera and Leica Wild M3C stereo zoom microscope. The acquired images are then preprocessed i.e. sharpened by using unsharp filter which is a simple sharpening operator. It enhances edges (and other high frequency components in an image) by subtracting an unsharp, or smoothed, version of an image from the original image.



The function 'fspecial' creates the unsharp filter from the negative of the Laplacian filter with parameter alpha. Alpha controls the shape of the Laplacian and must be in the range 0.0 to 1.0. The selected value for alpha is 0.2. These images are segmented using Otsu's thresholding method

VI. Image Segmentation

Segmentation involves separating an image into regions (or their contours) corresponding to objects. Usually by identifying common properties the regions are segmented. The simplest property that pixels in a region can share is intensity. So, a natural way to segment such regions is through *thresholding*, the separation of light and dark regions. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one. If $g(x, y)$ is a thresholded version of $f(x, y)$ at some global threshold T ,

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{Otherwise} \end{cases}$$

Otsu's thresholding method involves iterating through all the possible threshold values [4]. It calculates a measure of spread for the pixel levels on each side of the threshold, i.e. the pixels that either falls in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum. The next step involves calculation of the 'Within-Class Variance'. This is simply the sum of the two variances multiplied by their associated weights.

$$\text{Within Class variance } \sigma_w^2 = W_b \sigma_b^2 + W_f \sigma_f^2$$

The within-class variance for all the possible threshold values is calculated and then the threshold value for which the within class variance is lowest is selected for thresholding. In case of the RGB image a Karhunen-Loeve transform is first performed on the three R, G, B channels. The segmentation is then performed on the image component that contains most of the energy. The segmented images are then complemented to identify

the background (BC) of infected region as shown in Fig. (3). Now to separate the core (CR) and border (BR) regions of the infected images, second level segmentation is performed the results is shown in Fig. (4).

VII. Feature Extraction

A. Color feature extraction

Colors features are obtained by calculating mean (M) and standard deviation (SD) of the intensity of pixels creating spots in three classical planes; red (R), green (G) and blue (B) of the segmented images. The global sampling mean value of an image is the average intensity of all the pixels in the image. Let A be N×M image then its global mean is given by:

$$m = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} A(i, j)$$

The global mean is called simply mean.

Variance is the second moment of intensity about its mean and is given by

$$\sigma^2 = \mu_2(r) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (A(i, j) - m)^2$$

where 'm' is the mean.

Standard deviation is the square root of the variance and is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{k=0}^{l-1} (r_k - m)^2 p(r_k)} = \sqrt{\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (A(i, j) - m)^2}{MN}}$$

All the extracted features are listed in Table 1 using their abbreviated name. For example, BC_M_R, and BC_SD_R represent mean and standard deviation of the spot in the background region by considering red colour plane. Similarly for border and core regions, the feature values are extracted by considering green and blue colour planes.

B. Shape based feature extraction

Area (AR), Sharpness (SH), Perimeter (PR), Eigen values (EV) and Aspect-ratio (ASR) are computed as shape based features to detect the diseases. Area of an object S is defined as the number of pixels comprising the object. So area can be enumerated by counting the object pixels. The perimeter of an object is the length of its boundary and can be determined by counting the pixels on the boundary of the object spot. The length of line segment connecting the two extreme points that compromise the diameter gives major axis length. The minor axis of a boundary is the line perpendicular to major axis and of such length that a box passing through outer four points of intersection of the boundary with the two axes completely encloses the boundary. The ratio of major axes to minor axes length is called aspect ratio. It provides information about the horizontal or vertical stretch of the diseased spot. Sharpness of the shape of the spot is determined by calculating the average distance between two boundary points along the major axis. The eigenvector of a matrix M is some vector x which when multiplied by the matrix M results in the same vector multiplied by some factor. This factor is known as eigen values (ρ).

$$Mx = \rho x$$

where ρ is called eigen value. The eigen value represents valuable information, about the image of the infected region [5-6]. The table(1) gives the sample of features of images.



Original image Fig. (1)



Filtered image Fig. (2)



Otsu segmented image Fig. (3)



Otsu double segmented image Fig. (4).

Table 1: Samples of image features

Sr. No.	Feature	Myrothecium	Bacterial leaf Blight	Altamaria
1.	BcMRed	35.44182	59.51219	39.65875
2.	BcMGm	40.11133	62.73288	43.25169
3.	BcMBlu	21.37432	32.11552	41.71329
4.	CrMRed	32.75282	45.81294	23.70931
5.	CrMGm	37.63609	48.45696	20.52739
6.	CrMBlu	19.63965	25.93067	19.38514
7.	BrMRed	2.68901	13.69926	15.94943
8.	BrMGm	2.47524	14.27592	22.72430
9.	BrMBlu	1.73466	6.18485	22.32815
10.	BcSdRed	63.62281	85.33488	57.07435
11.	BcSdGm	71.49907	90.30237	53.07392
12.	BcSdBlu	39.62155	50.55120	51.79843
13.	CrSdRed	62.30380	83.15030	56.06706
14.	CrSdGm	70.77257	88.10177	47.20886
15.	CrSdBlu	38.64649	50.06459	45.13815
16.	BrSdRed	18.49983	40.28990	29.50034
17.	BrSdGm	17.01990	42.14417	39.00141
18.	BrSdBlu	12.01882	19.22798	38.87541
19.	Area	50381.0000	22568.00000	766.00000
20.	Aspect-Ratio	1.14090	1.23327	1.30979
21.	Sharpness	3031.68708	2230.46849	444.35819
22.	Eigen values	117529.9989	0	8599.8242
		23380.3356	0	-2470.2514

VIII. Classification

Support vector machines are used as a classifier in the work because their performances are superior to other learning algorithms [7-8]. SVM was introduced by Boser, Guyon and Vapnik in 1992. It is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. SVMs provide answer to two-class classification problems by mapping the input vectors into a new high-dimensional feature space. For this purpose it uses some nonlinear mapping and constructs an optimal separating hyper plane by determining the largest margin to separate positive and negative classes. The Gaussian radial basis function kernel is used as network mapping function in this work. The concept is illustrated in Figure (5).

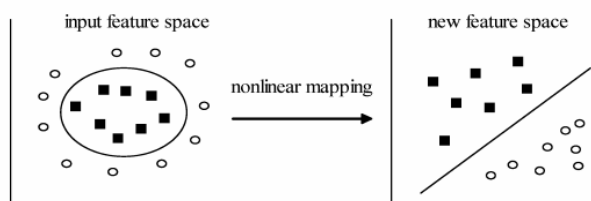


Fig. (5)

The architecture of the SVM is shown below. All the 22 features extracted are used as the input elements. The hidden layer has number of neurons equal to the number of training examples which will make the net convergence easy.

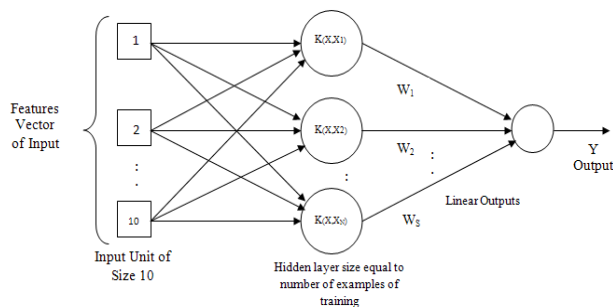


Fig. (6)

IX. Comparison to related work

We found that comparatively less research work has taken with cotton as research subject. The automatic classification system suggested by A.A. Bernardes & J.G. Rogeri for cotton diseases namely Ramularia, Bacterial Blight and Ascochyta Blight has accuracy between 71.4% to 97.1% [16].

H. Al-Hiary, S. Bani-Ahmad, M. Reylat [17] proposed automatic system for detection and classification of plant diseases. to get the average efficiency of 94.67%.

Song kai & Liu zhikun [15] proposed the maize disease image recognition system for Corn. The accuracy was as high as 98%.

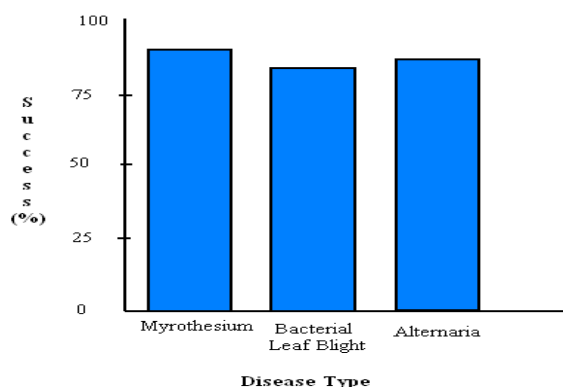
Phadikar and Sil studied rice crops achieving 92% of correct disease identification.

The proposed work for cotton leaf diseases namely Myrothecium, Bacterial Leaf Blight and Alternaria with average accuracy of 90% provide good tool for the farmers in the region.

X. Conclusion

Digital image analysis has proven to be a viable approach to the identification of cotton leaf disease. The results presented have demonstrated that the identification of Bacterial Blight, Myrothecium and Alternaria is achievable using color features such as mean and standard deviation of the intensity of pixels creating spots in three classical planes; red, green and blue of the segmented images and shape features such as area, sharpness, perimeter, eigen values and aspect-ratio. The training and test data sets include the images of three diseases of cotton leaf. Some of the images are gathered from CICR Nagpur, cotton fields in Buldhana & Wardha district. The following graph summarizes the performance of our system.

TABLE II: PERFORMANCE OF THE SYSTEM



The research work in disease identification and classification was mainly done in human medical science. Majority of work in crop diseases identification was done for wheat, rice and maize crop. The proposed system provides the simple and efficient tool for farmers in the local region where cotton happens to be a major cash crop. The proposed system will take care of only Bacterial Blight, Myrothecium and Alternaria leaf diseases. There are other leaf, root and fruit diseases which are not addressed by the system. The efficiency of the system may get affected in case of faulty images. Expansion on this work can target identification of other cotton leaf diseases like Leaf curl, Rhizoctonia, Powdery mildew and root diseases like Root rot. We can also extend this work to develop hybrid algorithms such as genetic algorithm and neural network in order to increase the accuracy of classification underscoring the advantages of hybrid algorithms.

References

- [1]. Weizheng S., Yachun W., Zhanliang C., and Hongda W., "Grading Method of Leaf Spot Disease Based on Image Processing" in Proceedings of the 2008 international conference on Computer Science and Software Engineering, Volume 06, December 12-14, 2008

- [2]. Rampf T., A. K. Mahlein, U. Steiner, E. C. Oerke, H.W.Dehne, L. Plumer, "Early detection and classification of plant diseases with Support Vector Machine based on Hyperspectral reflectance", Computers and Electronics in Agriculture, Volume 74, Issue 1, October 2010, Pages 91-99, ISSN 0168-1699.
- [3]. Hillnhuetter C., A. K. Mahlein, Early detection and localization of sugar beet diseases: new approaches, *Gesunde Pflanzen* 60 (4) 2008, pp. 143-149.
- [4]. Otsu N.(1979). "A threshold selection method from grey level histogram". *IEEE Trans. Sys., Man., Cyber* 9; 62-66. DOI: 10.1109/TSMC.1979.4310076.
- [5]. Polya G., *Studies in Mathematics and Mechanics Presented to Richard von Mises*, Academic Press, New York, (1954)200-207.
- [6]. Courant R. and Hilbert D., *Methods of Mathematical Physics*, 2nd Edition, Interscience Publishers (1965).
- [7]. Cristianini N. and Shawe-Taylor J. 2000. *An introduction to Support Vector Machines and other Kernel-Based Learning methods*. Cambridge University Press, New York.
- [8]. Li S., Kwok J. T., Zhu H. and Wang Y. 2003. Texture classification using support vector machines. *Pattern Recogn.* 36, 2883-2893.
- [9]. V. Vapnik *The Nature of Statistical Learning Theory*. Springer N.Y., 1995. ISBN 0-387-94559-8.
- [10]. Galbieri R. "Behaviour of Cotton Genotypes in the Presence of Pathogens and Nematodes. Master Dissertation (Tropical and Subtropical Agriculture), Campinas, SP-Brazil, 2007.
- [11]. G.L. Li, Z.H. Ma and H.G. Wang, "Image recognition of white stripe rust and wheat leaf rust based on support vector machine", *Journal of China Agriculture University*, vol.17, pp 72-79, April 2012.
- [12]. G.L. Li, Z.H. Ma and H.G. Wang, "Image recognition of grape downy mildew and grape powdery mildew based on support vector machine", *CCTA 2011, part III, IFIP AICT 370*, 2012, pp. 151-162.
- [13]. Yuan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo, "SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases", *Proceedings of 2010 Conference on Dependable Computing CDC'2010*, November 20-22, 2010.
- [14]. Al-Bashish, D.M.Braik and S.Bani-Ahmed, "Detection and classification of leaf diseases using k-means-based segmentation and neural-networks-based classification", *Inform. Technol. J.*, 265-275, 2011.
- [15]. Song Kai, Liu Zhikun, Su hang, "A Research of maize disease image recognition of Corn Based on BP Networks", *Third International conference on measuring technology and Mechatronics Automation* 2011.
- [16]. A.A. Bernardes, J.G.Rogeri, N.Marranghello and A.S. Pereira, "Identification of Foliar Diseases in Cotton Crop" SP, Brazil.
- [17]. H. Al-Hiary, S. Bani-Ahmad, M. Reylat, "Fast and accurate Detection and classification of plant Diseases", *International Journal of Computer Applications*, Volume 17-No.1, March 2011.
- [18]. Jinghui Li, Lingwag Gao, Zuorui Shen, "Extraction and analysis of digital images feature of three kinds of wheat diseases" 3rd International Congress on Image and Signal Processing (CISP2010).