Tomato Plant Disease Classification in Digital Images using Classification Tree

H. Sabrol and K. Satish

***Abstract*--The applications based on image processing for plant disease recognition and classification is the wide area of research these days. These applications are useful for timely recognition of plant disease. The disease like fungal, bacterial and virus are the destructive disease for any plant. In the study, five types of tomato diseases i.e. tomato late blight, Septoria spot, bacterial spot, bacterial canker, tomato leaf curl and healthy tomato plant leaf and stem images are classified. The classification conducted by extracting color, shape and texture features from healthy and unhealthy tomato plant image. The feature extraction process is done after the segmentation process. Extracted features from segmented images fed to classification tree. Finally, the disease classification was based on these six different types of classes. The classification of six types of tomato images yielded overall 97.3% of classification accuracy.**

***Index Terms*--Plant disease classification, image processing, feature extraction, classification tree**.

I INTRODUCTION

Plant disease is an abnormal state of the plant that disturbs the normal growth of plant [1]. There are several types of plant disease could cause several losses to the production of crops. The presence of the pathogen depends on the favorable environment conditions and varieties of crops grown, which is the reason for occurrence and prevalence of plant diseases. There are various plant disease management programs that will help to reduce losses in yields and grain quality. From the three or four decades, research in the field of plant disease recognition and classification contributed greatly. The correct identification of diseases on early will may help in taking action to prevent losses produce a high-quality yield of great good grain. Image recognition of plant diseases is a widely concern now these days, which develops the visual applications and also the reason for the popularization of digital technologies.

H.Sabrol, Research Scholar is with the Department of Computer Science and Applications, Panjab University, Chandigarh, India. (e-mail:hiteshwarisabrol@gmail.com)

K. Satish is with the Department of Computer Applications, P.U. SSG Regional Centre (Panjab University, Chandigarh), Hoshiarpur, India. (e-mail:satishnotra@yahoo.com)

An automatic recognition and quantification of plant diseases based on image processing techniques and methods proposed and studied [2][3][4][5][6][7][8]. The automatic diagnosis system based on plant disease features reduces the dependency on experts in the area concerned. The characteristic feature of plant disease extracted from the diseased region of the disease affected images using image processing techniques. The recognition of the disease-infected plant done by pattern recognition techniques such as "neural networks"[9][10], "support vector machine" [11][12], etc. The features extracted from the digital images of plant diseases include "color" [11][13][14], "shape" [15], "texture"[16] features and so on.

To achieve automatic plant disease identification and diagnosis based on image processing, image recognition of various types of diseases that found on most of the horticulture plants but in this study we were taking the plant leaf images of three different types of disease like fungal, bacterial and virus of tomato plants, are caused by the different types of fungus, bacteria and virus. The fungal disease which taken for the study are tomato septoria leaf spot and tomato late blight, bacterial diseases are bacterial leaf spot and bacterial canker and virus included tomato leaf curl. The tomato plant disease classification includes seven steps as mentioned in the Fig 2. Step 1. Total types of images are Six included fungal: 2, bacterial: 2, viral:1 and healthy (Normal):1. Step 2. Created training and testing dataset. Step 3. To perform segmentation using Otsu's method on training and testing dataset. Step 4 and Step5. Extracted total 24 features, i.e., color, shape, and texture from segmented images of tomato plant healthy and unhealthy. Step 6. To give input and target data for classification. Step 7. Final classification using classification tree. The classification tree used for classifying the different types of tomato plant healthy and unhealthy images. The "classification tree" [17], there is training samples of n observations on I

978-1-5090-0396-9/16/$31.00 ©2016 IEEE

variable classes that take the values 1,2,3…j and prediction values X1…Xj. The main aim of the classification tree is to find the best prediction values for Y on from the new predicted values X. Paper is organized as follows: materials and methods are explained in section II. Result and discusion are presented in section III. Finally section IV concludes the paper.

A. B. C. D.

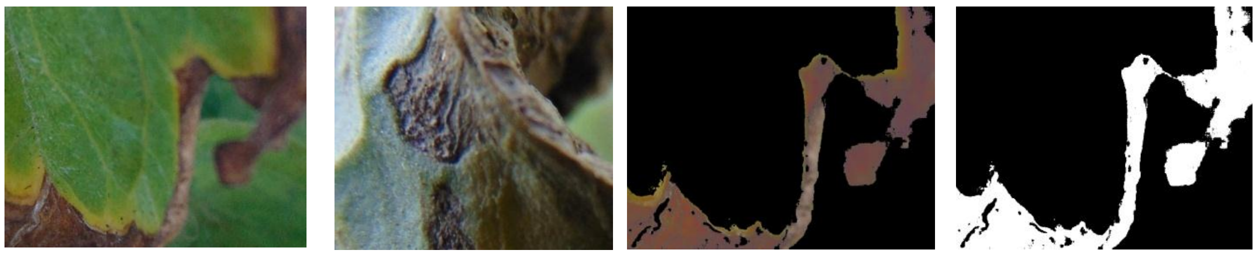


Fig. 1.1. A) & B) Late Blight Affected Tomato Leaves C) Color Segmented Image D) Binary segmented images

II. MATERIAL AND METHODS

In the current study, a dataset of total 383 digital images of tomato plant disease was obtained by using the common digital camera. The Otsu's segmentation applied on the images dataset. Total ten times the images are randomly used for training and testing purpose. For training, randomly three times 345 and seven times 349 images used. For testing, randomly the set of 39 images, three times and 38 images, seven times for testing are used. These sets of images included bacterial leaf spot 80 (Cat 1), fungal septoria leaf spot 26 (Cat 2), fungal late blight 95 (Cat 3), healthy 58 (Cat 4), bacterial canker 46 (Cat 5) and tomato leaf curl 78 (Cat 6). In MATLAB R2012b total 24 features including color 9, Shape 11 and texture 4 extracted from the training and testing images.

To color features extraction, we computed total nine features from RGB healthy and unhealthy images of the tomato plant. The R, G, B color components extracted from the RGB images. Then computed the maximum of each R, G, B components. i.e. max(R), Max(G), Max(B). Two feature vectors computed from each R, G, B, i.e., mean and standard deviation. Then computed mean(R), mean(G), mean (B) and std2(R), std2 (G) and std2 (B) [18].

Mean:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \qquad (1)$$

Standard Deviation:

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^2} \qquad (2)$$

To compute the shape features, we extracted eleven shape features i.e. area, Euler number, orientation, extent, perimeter, convex area, filled area, eccentricity, majoraxis length, equidiameter, and minaxislength. For extracting these features, we developed the function using *regionprops* function from Matlab Image Processing Toolbox, which applied on segmented binary images.

From the each region of interest, the gray-level co-occurrence matrix determined. The co-occurrence matrix formed by considering neighboring pixels a distance of one unit and at four angles (0°, 45°, 90°, and 135°) from each pixel of interest. It will result in 8*8

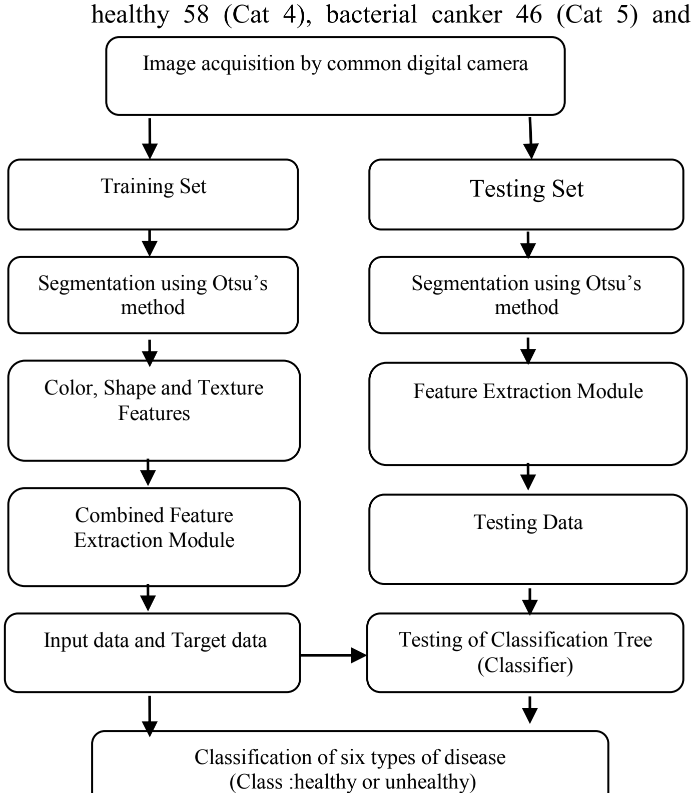


Fig. 2. Steps for Tomato Plant Disease Classification using Classification

symmetric matrix, in which element (i, j) is a frequency count of the number of times a pixel of intensity j neighbors by a predefined distance (two pixels in the current study) at any one of the four defined angles, a pixel of intensity i. For this study, the neighbor distance was set to two. Thus, each pixel is compared against two of its neighbor pixels: the pixel two columns to the right and the pixel two rows down. The pairs formed by both of these relationships are used to increment the appropriate pixels in the output image. The texture defined by the matrix of relative frequencies, P(i, j). Further, this matrix is subsequently normalized by dividing each element into the normalized array by the total number of co-occurrence pairs. For a four angle, one neighbor operation on an image of n, the total number of pixel pairs

$$p(i,j) = \frac{P(i,j)}{n} \quad (3)$$

Four common texture properties were calculated using the normalized matrix, as described following

T1: Correlation:

$$\sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(i-m_r)(j-m_c)pij}{\sigma_r \sigma_c} \quad (4)$$

where m_r or m_c and $\sigma_r or \sigma_c$ refer to the mean and standard deviation, respectively, of the rows or columns.

T2: Contrast: $\sum_{i=1}^{K} \sum_{j=1}^{K} (i-j)^2 pij$ (5)

T3: Energy: $\sum_{i=1}^{K} \sum_{j=1}^{K} [p(i,j)]^2$ (6)

T4: Homogeneity: $\sum_{i=1}^{K} \sum_{j=1}^{K} \frac{pij}{1+|i-j|}$ (7)

The classification tree is non-parametric supervised learning technique. The classification tree described the creation of "binary decision tree" [17]. The method follows the greedy technique in which classification tree are created in a top-down recursive divide and conquer manner. The classification tree adopts the top-down approach, which begins with the tuples of training set and their associated class labels. For the generation of classification tree, the tree partitioned in smaller subsets recursively. In the study, five types of diseases and healthy is assumed as six types of different classes with its category labels i.e. Cat 1, Cat 2, Cat3, Cat 4, Cat 5 and Cat 6 as mentioned in Section II. The splitting criterion for generating classification tree is Gini index.

Gini Index: $1 - \sum_{i=1}^{m} p_i^2$ (8)

The Gini index generated for the root node based on 24 features is n=98.703 as shown in the Fig 3.b. Total nodes created for classification is 83. Following is an example of if-then rule used in the proposed method:

If vertex n<98.703 then
    Create node 2
else if vertex n>=98.7037 then
    Create node 3
else
    Image belongs to class =Cat3
end

The if-then rules as mentioned above generated for each 383 images by using features extracted from images.

The tomato plant disease classification using classification tree is following:

Input: Random Data set partition Data_Set (345 and 346) images (tuples) for training with their associated categories.

Testing: Data set of 38 and 39 images

Attribute list (attributes): Total 24 features (Color 9, Texture 4 and Shape 11)

Attribute Selection method: split point (continuous-valued) splitting_criteria: Gini index.

Output: A classification tree based on six categories of tomato plant images.

Method: plant_disease_classification_tree(training data partition, Data_Set)

1. Create a vertex n:
2. If all the images are same in Data_Set with plant_image_category, Cat then
    Return n as a leaf vertex with the named Category;
3. If attributes is null then
    Return n as a leaf vertex with the majority class in Data_Set;
4. Employ selection method(Data_Set, attributes) to apply the splitting_criteria;
5. Named vertex n with splitting_criteria;
6. If splitting_point is continuous-valued then
    Restricted to binary trees
    attributes=attributes- splitting_point;
7. For each output i of splitting_criteria
    Data_Set i the set of data tuples in Data_Set satisfying output I;
   If Data_Set i is null then
    Connect a leaf node named with greater number class in Data_Set to node n;
   Else
    Connect the node returned by plant_disease_classification_tree(Data_Set i, attributes) to vertex n;
8. End For
9. Return n.

See the output in Fig 3.a and 3.b.

III. RESULTS AND DISCUSSION

The classification accuracy of the six different kinds of tomato plant images is yielding classification 97.3% (Fig 3.a). In the plant disease classification tree (Fig 3.b), total 83 nodes are created for classification. The splitting criteria based on Gini index used for dividing the sub nodes further till the leaf nodes not generated for various categories. The tomato plant images categorized into three types of diseases fungal, bacterial and virus. Further, subdivided into five types of disease and healthy. These images segmented by using otsu's method into segmented binary and color images. Then total 24 combined feature extracted and fed into classification tree. The Classification tree is resulting 97.3% classification accuracy as a classifier (Fig 3.a).

IV. CONCLUSION AND FUTURE SCOPE

In this study, supervised learning technique is used for learning purpose to classify the tomato plant leaves into six classes: healthy and unhealthy, i.e., five types of diseases due to fungal, bacterial and virus. The combined features included color, shape and texture features extracted from both type of plant images (diseased affected or normal). The classification results showed that classification tree is resulting good accuracy. The method proposed in the paper could also use for plant disease image recognition and classification. There are more sophisticated techniques are available for classification like Adaptive neuro-fuzzy, Neural Networks, Genetic algorithm. Support vector machines etc. for image classification. These techniques can also use for plant image recognition and classification.

Fig. 3.a. Tomato Plant Disease classification accuracy: 97.3%

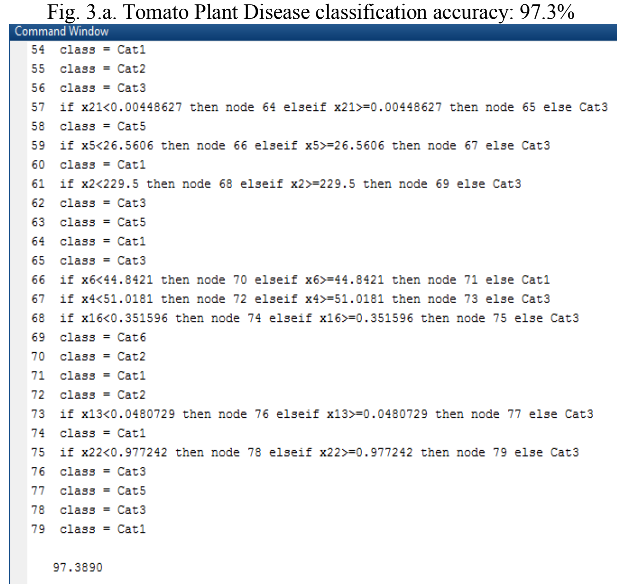
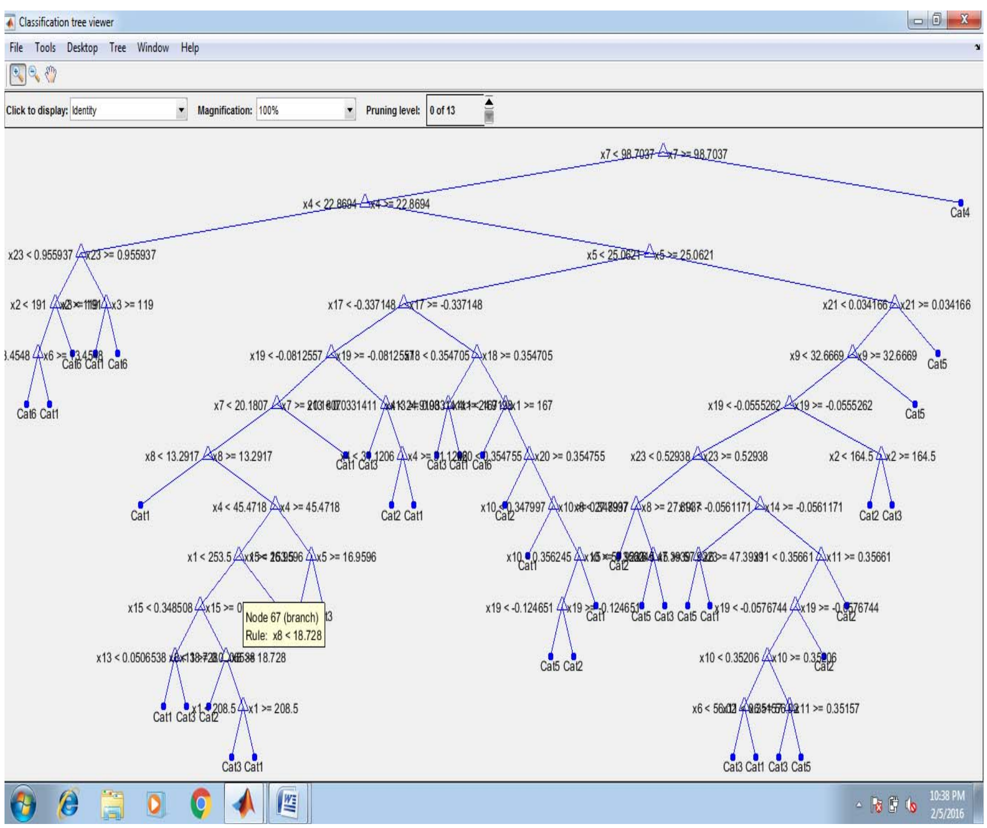


Fig. 3.b. Tomato Plant Disease Classification Tree. Root node created based on Gini index=98.703



REFERENCES

[1] Arthur Kelmam, "Plant Diseases Encyclopedia. Britannica", 2012. Available: http://www.britannica.com/science/plant-disease.

[2] C. C. Tucker and S. Chakraborty. "Quantitative assessment of lesion characteristics and disease severity using digital image processing", J. Phytopathology, 1997, vol.145, pp.273–278.

[3] S. Ahmad, J. F. Reid, M. R. Paulsen, and J. B. Sinclair, "Color classifier for symptomatic soybean seeds using image processing PlantDisease",1999, vol. 83, pp.320–327.

[4] J. W. Olmstead, A. Gregory, and G. A. Lang., "Assessment of severity of powdery mildew infection of sweet cherry leaves by digital imageanalysi", Hortscience, 2001,vol. 36, pp.107–111.

[5] J. C. Lai, S. K. Li, B. Ming, N. Wang, K. R. Wang, R. Z. Xie, and S. J. Gao., "Advances in research on computer-vision diagnosis of cropdiseases. Scientia Agricultura Sinica",2009, vol. 42, pp.1215–1221.

[6] C. H. Bock, A. Z. Cook, P. E. Parker, and T. R. Gottwald., "Automated image analysis of the severity of foliar citrus canker symptoms. PlantDisease",2009, vol.93, pp. 660–665.

[7] C. H. Bock, G. H. Poole, P. E. Parker, and T. R. Gottwald., "Plant disease severity estimated visually, by digital photography and imageanalysis, and by hyperspectral imaging", Critical Reviews in Plant Sciences, Mrach 2010, vol. 29, pp. 59–107.

[8] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z.ALRahamneh, "Fast and accurate detection and classification of plantdiseases", International Journal of Computer Applications, 2001, vol. 17, pp.31–38.

[9] D. W. Zhang and J. Wang, "Design on image features recognition system of cucumber downy mildew based on BP algorithm", Journal of Shenyang Jianzhu University (Natural Science), May 2009, vol. 25, pp. 574–578.

[10] D. T. Zhao, Y. H. Chai, and C. L. Zhang, "Inspection of soybean frogeye spot based on image procession", Journal of Northeast Agricultural University, vol. 41, pp. 119–124, April 2010

[11] Y. W. Tian, and Y. Niu, "Applied research of support vector machine on recognition of cucumber disease", Journal of Agricultural Mechanization Research, vol. 31, pp. 36–39, March 2009

[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] Y. L. Cui, P. F. Cheng, X. Z. Dong, Z. H. Liu, and S. X. Wang, "Image processing and extracting color features of greenhouse diseased leaf", Transactions of the CSAE. vol. 21(Supp.), pp. 32–35, December 2005.

[14] Z. X. Cen, B. J. Li, Y. X. Shi, H. Y. Huang, J. Liu, N. F. Liao, J. Feng, "Discrimination of cucumber anthracnose and cucumber brown speck base on color image statistical characteristics", Acta Horticulturae Sinica, vol. 34, pp. 1425–1430, June 2007.

[15] Y. X. Zhao, K. R. Wang, Z. Y. Bai, S. K. Li, R. Z. Xie, and S. J. Gao, "Research of maize leaf disease identifying system based image recognition", Scientia Agricultura Sinica, vol. 40, pp. 698–703, April 2007.

[16] R. Pydipati, T. F. Burks, and W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis", Computers and Electronics in Agriculture,vol. 52, pp. 49–59, June 2006

[17] L. Breiman, JH. Friedman, RA. Olshen, CJ. Stone.1984 "Classification and Regression Trees". CRC Press.

[18] M. Stricker and M. Orengo. "*Similarity of color images",* In SPIE Conference on Storage and Retrieval for Image and Video Databases III, March 1995,volume 2420.