"Diagnosis and Classification of Grape Leaf Diseases using Neural Networks"

Sanjeev S Sannakki[1], Vijay S Rajpurohit[2], V B Nargund[3], Pallavi Kulkarni[4*]
sannakkisanjeev@yahoo.co.in vijaysr2k@yahoo.com nargund56@gmail.com pallavideshpande66@gmail.com
[1,2,4] Dept. of Computer Science, Gogte Institute of Technology Belgaum, Karnataka, India
[3]Department of Plant Pathology, University of Agricultural Sciences, Dharwad, India

***Abstract*—Plant diseases cause significant damage and economic losses in crops. Subsequently, reduction in plant diseases by early diagnosis results in substantial improvement in quality of the product. Erroneous diagnosis of disease and its severity leads to inappropriate use of pesticides. The goal of proposed work is to diagnose the disease using image processing and artificial intelligence techniques on images of grape plant leaf. In the proposed system, grape leaf image with complex background is taken as input. Thresholding is deployed to mask green pixels and image is processed to remove noise using anisotropic diffusion. Then grape leaf disease segmentation is done using K-means clustering. The diseased portion from segmented images is identified. Best results were observed when Feed forward Back Propagation Neural Network was trained for classification.**

***Index Terms*—plant disease identification, Feed forward neural network, image processing, k-means, co-occurrence matrix, feature extraction**

I. Introduction

Grape (*Vitis vinifera*) cultivation-Viticulture is one of the most remunerative farming enterprises in India [1]. Grapes originated in Western Asia and Europe. Fruit is eaten fresh or made into juice, fermented to wines and brandy and dried into raisins [2]. Grapes also have medicinal properties to cure many diseases.

Grapes generally require a hot and dry climate during its growth and fruiting periods. It is successfully grown in areas where the temperature range is from 15^0-40^0C. High temperatures above 40^0C during the fruit growth and development reduce fruit set and consequently the berry size. Low temperatures below 15 C followed by forward pruning impair the budbreak leading to crop failure. Grapes can be cultivated in variety of soils including sandy loams, red sandy soils, sandy clay loams, shallow to medium black soils and red loams [16].

Grape suffers from huge crop losses on account of downy mildew, powdery mildew and anthracnose [1]. In case of downy mildew, the losses are very high when the clusters are attacked before fruit set. Entire clusters decay, dry and drop down [16]. Plant disease is one of the crucial causes of reduction in quantity and degrades quality of the product. The naked eye observation of experts is the main approach adopted in practice for detection and identification of plant diseases. This approach is prohibitively expensive and time consuming in large farms [3]. Further, in some developing countries, farmers may need to go long distances to contact experts. Diseases are managed by adjusting the pruning time and using various fungicides [2]. Observations during research at NRCG, Pune show that precision farming i.e. using information technology for decision making has improved the yield and quality of crops.

II. Literature Survey

Various researchers have proposed image-processing and pattern recognition techniques in agricultural applications for detection of weeds in a field, sorting of fruits and vegetables, detecting diseases etc. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and detect the symptoms of diseases as soon as they appear on plant leaves. Therefore; looking for fast, less expensive and accurate method to detect plant disease cases is of great significance. Excessive use of pesticides for plant disease treatment increases costs and raises the danger of toxic residue levels on agricultural products. This requires that the disease must be identified accurately and also the stage in which the disease is. Hence an efficient disease identification and diagnosis model is required.

Al-Bashish, D., M. Braik and S. Bani-Ahmad have proposed the leaf disease Detection and classification for early scorch, cottony mold, late scorch and tiny whiteness with K-means-based segmentation, CCM to consider texture feature classification and BPNN to classify diseases into one of the six disease classes [4]. A. Camargo, J.S. Smith discuss converting the RGB image of the diseased plant or leaf, into the H, I3a and I3b color transformations. The transformed image is then segmented by analyzing the distribution of intensities in a histogram. The extracted region was post-processed to remove pixel regions not considered part of the target region. Then the neighborhood of each pixel and its gradient is analyzed [5]. S. S. Sannakki, V. S. Rajpurohit, V B Nargund, et.al proposes K-means clustering for segmentation and Disease grading by Fuzzy Logic where the intensity of disease is decided by the area of diseased portion [6]. H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, et.al propose masking of green pixels before extracting feature which results in more accurate classification [7].

All these researchers gather the leaf samples and take images in controlled environment i.e. plane background, good lightening conditions and click from constant distance.

A. Meunkaewjinda, P. Kumsawat et.al propose SOFM and BPNN to recognize grape leaf color, MSOFM for segmentation, GA and SVM for classification[8], process the images with complex background but use many machine learning algorithms which makes the system complex.

III. System Design

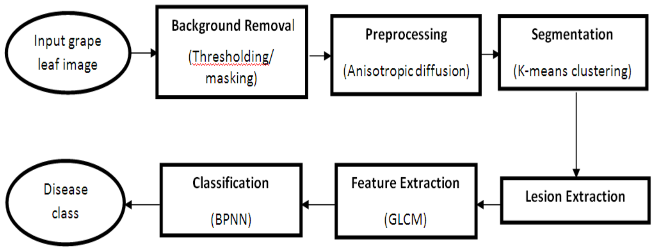


Figure 1: Decision Support System

The proposed system aims at processing the images with complex background, varying lightening conditions and clicked from various distances. This makes the system more dynamic to work under various climatic conditions. Figure 1 shows the corresponding block diagram which is an image processing system for grape leaf in which the following steps are undertaken.

A. **Image acquisition:** The First step is to collect the sample images which are needed to train system. In the working system, this step indicates the input or query image. Leaf images are captured using digital camera, Nikon Coolpix P510, 16.1 Megapixel and are used for training and testing the system. All the images are stored in standard jpg format. In the present study, images are captured from different regions like Pune, Bijapur, Sangali and expert advice is taken for identification. Some images are downloaded from internet to have diverse environment. Images gathered include the leaves affected by two major diseases found in India, downy mildew and powdery mildew.

B. **Background removal:** In this step, the input image is resized to standard size 300x300. Then, mostly green colored pixels are identified. If the green color component is less than threshold i.e. 70 in present work, then all red, green and blue value of that pixel is assigned zero and green channel of that image is assigned to 255. This is called masking green pixels. This fastens the processing in the next step and also improves accuracy [7].

A **mask** is a black and white image of the same dimensions as the original image (or the region of interest you are working on). Therefore, each of the pixels in the mask can have a value either 0 (black) or 1 (white) as shown in Figure 2. When executing operations on the image the mask is used to restrict the result to the pixels that are 1 (selected, active or white) in the mask and the operation restricts to some parts of the image.

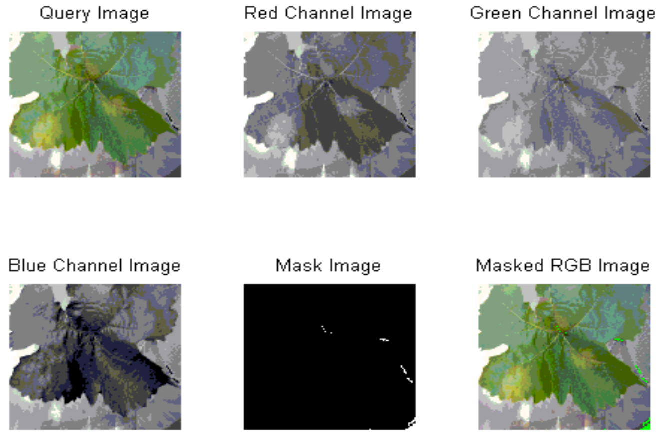


Figure 2: The query image, RGB channels, mask and resulting image after masking.

C. **Preprocessing:** Then it is enhanced by five iterations of Anisotropic Diffusion [10] to preserve the information of affected portion. The diffused image is shown in Figure 3. Anisotropic diffusion is a generalization of this diffusion process; it produces a family of parameterized images, but each resulting image is a combination between the original image and a filter that depends on the local content of the original image. As a consequence, anisotropic diffusion is a *non-linear* and *space-variant* transformation of the original image.

Non-linear, Space variant method:

$$\nabla I\ (x, y, t) = \left(\frac{\partial}{\partial_x}, \frac{\partial}{\partial_y}\right) \tag{1}$$

$$E\ (x, y, t) = \nabla I\ (x, y, t) \tag{2}$$

$$c(x, y, t) = g\left(\left\|\nabla I(x, y, t)\right\|\right) \tag{3}$$

Equation 1 gives gradient of the brightness function. We have an edge/not edge estimation method "E" in equation 2. Equation 3 is used which will not only preserve but also sharpen the edges if g(.) is chosen properly. First equation amongst the two equations proposed by Perona and malik is used. H component from HSV color space is extracted to reduce the illumination effect.

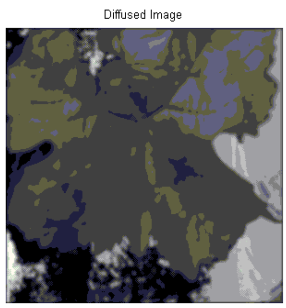


Figure 3: Filtered image after Anisotropic diffusion.

D. **Segmentation:** Clustering of data is a method by which large sets of data are grouped into clusters of smaller sets or segments of similar data. In present work, k-means clustering [10] is used to for segmenting an image into six groups which is found to be optimum as shown in figure 4. One or more (In case more diseases are present at a time) clusters contain the diseased portion of leaf. a and b component from L*a*b space are extracted before clustering.

K-means clustering algorithm:

- $x_1, \ldots, x_N$ are observed data points or vectors.
- Each observation (vector x_i) will be assigned to only one cluster.
- $C(i)$ is cluster number for the i^{th} observation
- For a given cluster assignment C of the data points, compute the cluster means m_k where

$$m_k = \frac{\sum_{i:C(i)=k} x_i}{N_k}, \; k = 1, \ldots, K. \qquad (4)$$

- For a current set of cluster means, each observation is assigned as:

$$C(i) = \arg\min_{1 \le k \le K} \|x_i - m_k\|^2, \; i = 1, \ldots, N \qquad (5)$$

- Iterate last two steps until convergence.

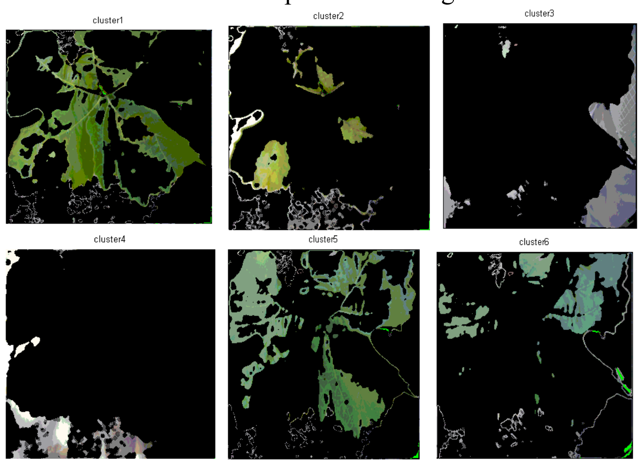
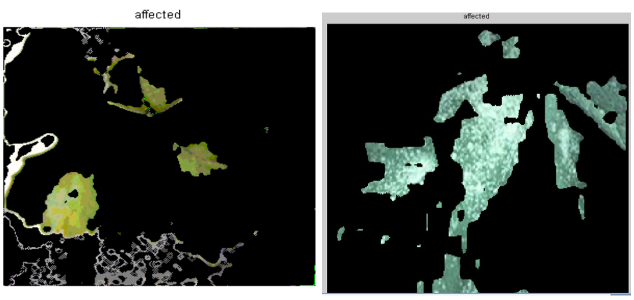


Figure 4: Six clusters formed by K-means clustering.

E. **Extract lesion:** Once the image is divided into six clusters, the mean of each cluster is calculated and means are sorted in ascending order. It is observed that the downy affected lesion is extracted at second (Figure 5) and powdery affected lesion is extracted at sixth in the sorted clusters. It is observed that this is true for the leaves having lesions of both the diseases at same time.



(a) (b)

Figure 5: (a) Downy affected region (b) Powdery affected region

F. **Feature extraction:** The next step is to extract texture features of the extracted diseased portions. This is done by calculating Gray Level Co-occurrence Matrix (GLCM) [12]. The colour co-occurrence texture analysis method was developed through the use of spatial gray-level dependence matrices (SGDM's). Co-occurrence matrices measure the probability that a pixel at one particular gray-level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray-level.

The SGDM's are represented by the function $P(i,j,d,\theta)$ where i represents the gray-level of location (x,y) in the image, and j represents the gray-level of the pixel at a distance d from location (x,y) at an orientation angle of θ., where i is the row indicator and j is the column indicator in the SGDM matrix $P(i,j,d,\theta)$. The nearest neighbour mask, where the reference pixel (x,y) is shown as an asterisk. All eight neighbors shown are one pixel distance from the reference pixel '*' and are numbered in a clockwise direction from one to eight as shown in figure 6. The neighbors at positions 1 and 5 are both considered to be at an orientation angle equal to 0°, while positions eight and four are considered to be at an angle of 45°.

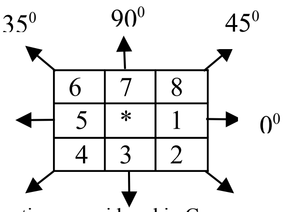


Figure 6: Directions considered in Co-occurrence matrix

Co-occurrence matrix is then normalized. The equation for normalizing co-occurrence matrix is given in Equation 6.

$$p(i,j) = \frac{P(i,j,1,0)}{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P(i,j,1,0)} \qquad (6)$$

In the above equation, $P(i, j, 1, 0)$ is the intensity co-occurrence matrix and N_g represents the total number of intensity levels.

Texture features can be used as useful discriminator when target images do not follow well defined color or shape [16]. Nine texture features listed in table I, are extracted for each image which will be used for training Neural Network in the next step for classification.

Classification: The feed forward Back Propagation Neural Network classifier [13] [15] consisting of three layers namely input layer, a hidden layer, and an output layer is used.

Training Back propagation Neural Network

We need to train the network with available data. Initially the network predicts an output for one input vector for which the true output is known. This combination of known output and input vector is called training sample. The predicted output is compared to the known value. The weights on the arcs are adjusted depending on how the prediction of the actual result. Sigmoid transfer function is used for generating output at each stage. The input layer has 9 nodes, which are related to two 9 texture features—contrast, uniformity, maximum probability, homogeneity, inverse difference, difference variance, diagonal variance, entropy of H bands of lesion area. Output layer contains two neurons. This module assigns an appropriate disease class i.e. Downy or powdery.

Table I. Mathematical formulations of texture features [14]

No.	Features	Formula
1.	Contrast	$\sum_i \sum_j \lvert i-j\rvert^2 p(i,j,d,\theta)$
2.	Uniformity(Energy)	$\sum_i \sum_j p(i,j,d,\theta)^2$
3.	Maximum probability	$\max_{i,j} p(i,j,d,\theta)$
4.	Homogeneity	$\sum_i \sum_j p(i,j,d,\theta)/(1+\lvert i-j\rvert)$
5.	Inverse difference moment of order 2	$\sum_i \sum_j 1/(1+(i-j)^2)\, p(i,j,d,\theta)$
6.	Difference variation	Variance of $\sum_i \sum_j \lvert i-j\rvert p(i,j,d,\theta)$
7.	Diagonal variance	Variance of $p(i,j,d,\theta)$
8.	Entropy	$\sum_i \sum_j p(i,j,d,\theta)\log(p(i,j,d,\theta))$
9.	Correlation	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j}$

IV. Experimental Results

The data consisted of 16 images of powdery mildew (class 1) and 17 images of downy mildew (class 2). MATLAB 7.1 Neural network pattern recognition tool was used for training. In 33, 29 images are used for training and 2 each for testing and validation. Data was given in two files namely input and target. Input file with 9 rows representing texture features and 33 columns representing sample images. Target file consisted two rows with binary values. [0 1] for downy and [1 0] for powdery. It gives good validation results as shown in figure 7. Confusion matrix in figure 8(a) shows images of class 1 and class 2 are classified properly. Figure 8(b) shows sensitivity (True positive) and specificity (False positive) rate of the system.

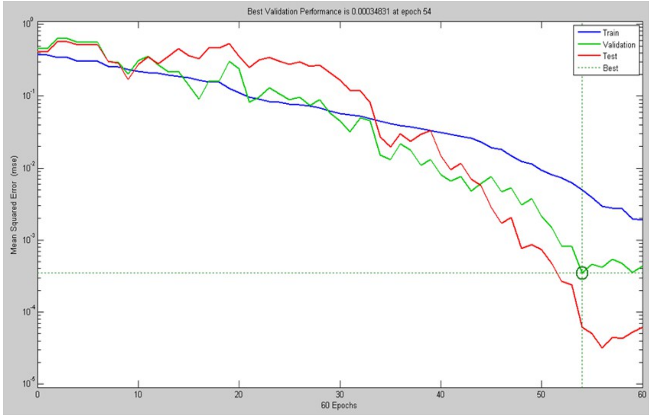
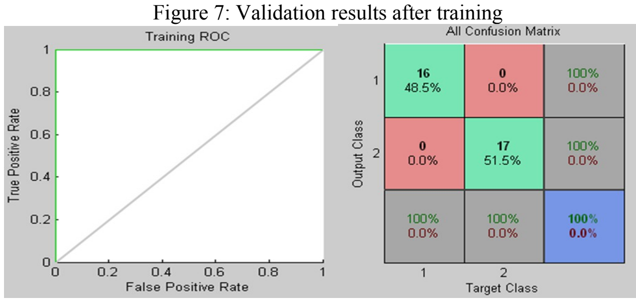


Figure 7: Validation results after training



(a) (b)

Figure 8: Sensitivity and specificity of system and confusion matrix

The green line meeting at upper left corner in Figure 8(a) shows the perfect result of binary classification test. During training it gave 100% correct results which shows the system will work almost accurately. The green box indicates correct classification and red box indicates wrong classification. The model that used hue features gives accurate results reaching a perfect 100%.

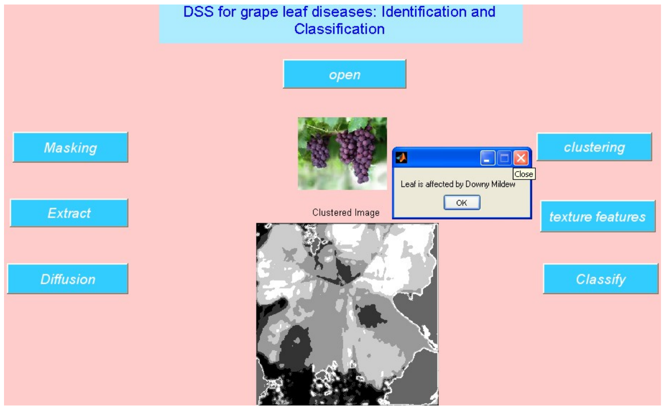


Figure 9: GUI developed using guide toolbox. Downy mildew is recognized by the system

V. Conclusion and Future Work

Study involved collecting leaf samples from different regions. Work was carried out to investigate the use of computer vision for classifying grape leaf diseases. Two classes of grape leaves, i.e., downy mildew and powdery mildew were considered in the experiments. Algorithms based on image-processing techniques, feature extraction and classification, were deployed. The feature extraction process used color co-occurrence methodology, which uses the texture of an image to arrive at unique features, which represent that image. nprtool of MATLAB 7.1 was used to train the neural network for pattern recognition. This training achieved training accuracies of 100% when using hue features alone.

Future work of this study can explore the utility of these techniques to include samples of healthy and other diseases of grapes like anthracnose. Instead of K-means, other segmentation techniques can be used to extract the lesion more accurately. With hue, models can be constructed with the combination of other color components like Saturation and intensity and the results can be compared.

Acknowledgment

- Dr. S. D. Sawant, National Research Centre for Grapes, Manjri Farm, Solapur Road, Pune.
- Dr. A. M. Sheikh, Regional Agriculture Research Station, Hitnalli Farm, Bijapur,

References

[1] A report of the expert consultation on viticulture in Asia and the Pacific. May 2000, Bankok, Thailand. RAP publication:2000/13.

[2] K. Soytong, W. Srinon et.al. Application of antagonistic fungi to control anthracnose disease of grape Journal of Agricultural Technology May 2005.

[3] Weizheng, S., Yachun, W., Zhanliang, C., and Hongda, W. (2008). 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). CSSE. IEEE Computer Society, Washington, DC, 491-494. DOI= http://dx.doi.org/10.1109/CSSE.2008.1649.

[4] Al-Bashish, D., M. Braik and S. Bani-Ahmad, (2011). Detection and classification of leaf diseases using K-means-based segmentation and neural-networks-based classification. Inform. Technol. J., 10: 267-275. DOI: 10.3923/itj.2011.267.275

[5] Camargo, A. and Smith, J. S., (2009). An image processing based algorithm to automatically identify plant disease visual symptoms, Biosystems Engineering, Volume 102, Issue 1, January 2009, Pages 9-21, ISSN 1537-5110, DOI: 10.1016/j.biosystemseng.2008.09.030.

[6] Sanjeev S Sannakki, Vijay S Rajpurohit, V B Nargund, et.al (2011) Leaf Disease Grading by Machine Vision and Fuzzy Logic, Int. J. Comp. Tech. Appl., Vol 2 (5), 1709-1716

[7] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh. Fast and Accurate Detection and Classification of Plant Diseases. International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011

[8] A.Meunkaewjinda, P.Kumsawat, K.Attakitmongcol et.al. Grape leaf disease detection from color imagery using hybrid intelligent system. Proceedings of ECTI-CON 2008.

[9] Peitro Perona ,Jitendra malik, scale space and edge detection using Anisotropic Diffusion, IEEE transactions on pattern analyses and machine intelligence, vol.12, no.7, July 1990.

[10] Macqueen, J.B.,1967. Some methods for classification and analysis of multivariate observations. Proc.Berkely Symp. Math. Statist. Prob., 1:281-297.

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

[12] I.A. Basheer, M. Hajmeer , Artificial neural networks fundamentals, computing, design, and application, Journal of Microbiological Methods 43 (2000) 3–31.

[13] Huang K.Y.Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture feature. Computer and Electronics in agriculture. Pp. 3-11, 2007.

[14] Prasad Babu, M. S. and Srinivasa Rao, B. (2010) Leaves recognition using back-propagation neural network - advice for pest and disease control on crops. Technical report, Department of Computer Science & Systems Engineering, Andhra University, India, www.indiakisan.net on May 2010.

[15] A. Camago,J.S.Smith, Image pattern classification for the identification of disease causing agents in plants, Computers and Elecronics in Agriculture 66(2009) 121-125.

[16] http://nhb.gov.in/fruits/grape/gra012.pdf, gra002.pdf

[17]http://www.nrcgrapes.nic.in/zipfiles/GRAPE%20PROFILE%20-%20NRC%20Grapes.pdf