ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Automatic Leaf Disease Detection and Classification using Hybrid Features and Supervised Classifier

S.Raj Kumar[1], S.Sowrirajan[2]

Professor, Dept. of ECE, Panimalar Engineering College, Chennai, Tamilnadu, India[1]

UG Student, Dept. of ECE, Panimalar Engineering College, Chennai, Tamilnadu, India[2]

ABSTRACT: Farmers at present find it difficult to visit the Plant Clinic in regular intervals to check the possible occurrences of disease in plants which is more time consuming and also results in loss of crops. To avoid this problem we propose this method to detect early diseases in the leaf. We collected the leaf samples of groundnut, mangoes, brinjal, tomato and maize. The samples are acquired using high resolution camera. The proposed decision making system utilizes image content characterization and supervised classifier Back Propagation with Feed Forward neural network (BPN-FF). At the preprocessing stage, the resizing of image to 256x256 pixels, color space conversion and region of interest selection is performed on an input image. Color, texture and geometric features of the image are extracted by the HSV conversion, GLCM, Lloyd's clustering respectively. The proposed method incorporates all the hybrid features with the aid of Lloyd's Clustering and BPN-FF classifier will be used for classification based on learning with the training samples and thereby providing the information on the abnormality (Early leaf spot, late leaf spot and alternaria leaf spot) as well as the respective medication to be sought out. Thus making it easy for the farmers to save their crops.

KEYWORDS: BPN-FF, HSV, GLCM, LLOYD'S CLUSTERING, ABNORMALITY

I. INTRODUCTION

Agriculture though being biological faces many a problems in growing of fine crops. This growing of crops is being affected by insects or any other plant malady thereby affecting the entire plantation as the infection spreads. It is estimated that the outbreak of helminthosporiose of rice in north eastern India in 1943 caused a heavy loss of food grains and death of a million people. Since the effects of plant diseases were devastating, some of the crop cultivation has been abandoned. It is estimated that 2007 plant disease losses in Georgia (USA) is approximately $653.06 million (Jean, 2009). In India no estimation has been made but it is more than USA because the preventive steps taken to protect our crops are not even one-tenth of that in USA. 70% of the Indian population depends on agriculture. Farmers have wide range of diversity to select suitable Fruit and Vegetable crops.

But the cultivation of these crops for optimum yield and quality produce is highly technical. It can be improved by the aid of technological support. Plants like tomato, potato, brinjal and many others face distinct threats from insects, bacteria, chemicals and even insecticides and pesticides. These diseases when left unidentified may cause huge loss of crops in the future. Also many of these diseases can be identified only during the fruit or vegetable producing stage of the plant. This project involves the use of Image Processing to detect these defects and also to instruct the owners of what to be done. Every plant disease usually begins with the leaf. So, the images of the plant leaves are collected and processed where hybrid features of the leaves are extracted. The hybrid features involving color, texture and shape features contribute to the test values used for training the neural network. A Back Propagation-Feed Forward (BPN-FF) classifier is used to train the artificial neural network. The system accepts the leaf images as input and charts out the hybrid features, then the classifier compares these features with the target vector matrix (comprising of hybrid features of normal and diseased leaves taken as references with which the network was trained) and hence provides the user with the information about the type of disease and also the cure for that disease. Thus Image processing is used to

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

monitor the traits of the plants for possible infirmities that might occur and provide the necessary data to ameliorate them before infection spreads.

II. LITERATURE SURVEY

Dheeb Al Bashish et. al. [1], proposed that the RGB images be converted into HSI plane and then the color features are extracted (by SGDM generation). The texture features are extracted by obtaining GLCM (Grey Level Co-occurrence Matrix). The input images are segmented using K-means clustering technique and then the segmented images are analysed by a pre-trained BPN network for detection and classification of plant leaf and stem diseases like early scorch, cottony mold, ashen mold, late scorch and tiny whiteness. The author also compares between various models incorporating various components such as HS,H,S,I,HSI and found that model HS provides the best efficient output amongst all other models with efficiency of 92.7%.

Niket, A.,et. al. [2], stated that the RGB images upon acquisition undergo color space transformation into HSI plane and upon segmentation using K-means clustering the green pixels are masked from the appropriate cluster and the masked green pixels are removed. Then the useful segments are obtained. Then the texture features are extracted using Color-Co-Occurrence Matrix (CCM). The classifier used in BPN-FF. But addition color or shape features or both might have improved the efficiency of classification.

Arivazhagan, S.,et. al. [3], proposed a new method in which the RGB images are converted to HSI plane and only the Hue component is used for further analysis. Then the green pixels are masked and the masked green pixels are removed. The useful components are obtained upon segmentation and only the texture feature is extracted using Co-occurrence matrix. Then the neural network employing SVM classifier is used to detect and classify early scorch, yellow spots, brown spots, late scorch with efficiency of 94.74%. Here only a single feature extraction is employed and the classifier which is not that efficient in classifying the disease but effectively detects whether the leaf is diseased or not.

Gurpreet Kaur and Himanshu, M.,et. al. [4], provide the study of various classifiers namely K-Nearest Neighbors(KNN) classifier, Support Vector Machines(SVM) classifier, Back Propagation Neural Network-Feed Forward(BPN-FF) classifier, Probabilistic Neural Network(PNN) classifier, General Regression Neural Network(GRNN) classifier.

Sanjeev, S.S.,et. al. [5], provided a new way in which the k-means segmentation is done, followed by feature extraction using GLCM and the classification is done via BPN. Here only hue component is used and instead of k-means segmentation other algorithms could have employed to extract lesion more accurately.

III. PROPOSED METHOD

The proposed approach starts first by creating device independent color space transformation structure. This converts the color values in the leaf image to color space specified in the color transformation structure. The color transformation structure specifies various parameters of transformation. A device independent color space is the one where the resultant color depends on the equipment used to produce it. To improve the precision of the disease detection and classification process, a device independent color space is required. In device independent color space, the coordinates used to specify the color will produce the same color regardless of the device used to take the pictures.
Thus the RGB to HSV and RGB to gray-scale conversions are made. Using the H (hue) and S (saturation) values color features are extracted. These features include Mean, Standard deviation, Skewness and Kurtosis each from H and S planes of the image. The texture features are extracted from the gray-scale image. It involves applying DWT (Discrete Wavelet Transform) to the grayscale image so as to get the high frequency components of the image. The DWT represents an image as a sum of wavelet functions, known as wavelets, with different location and scale. In DWT the data is represented in a set of high pass (detailed) and low pass (approximation) coefficients .In order to extract the texture features we make use of a 2D-DWT.
The next feature extracted is the Geometrical feature (Shape). This is done by the application of Lloyd’s Clustering. Initially four empty clusters are assigned. The next step involves determining the centroid of each cluster. Once the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

centroid is determined, the distance between each pixel and the centroid is calculated and the pixel is moved to the cluster whose centroid gives the minimum distance. This segmentation process is repeated until there are no more pixels left to repartition. Finally four clusters with repartitioned pixels from the image is obtained. The color, texture and shape features extracted are used to train the Back Propagation Feed Forward Neural Network to produce the desired output about the leaf image input.

Methodologies:

i. Pre-processing
ii. Color Space Conversion
iii. Color Feature Extraction
iv. Multi wavelet Transformation
v. GLCM (Texture Feature Extraction)
vi. Lloyd's clustering
vii. Geometrical Feature Extraction

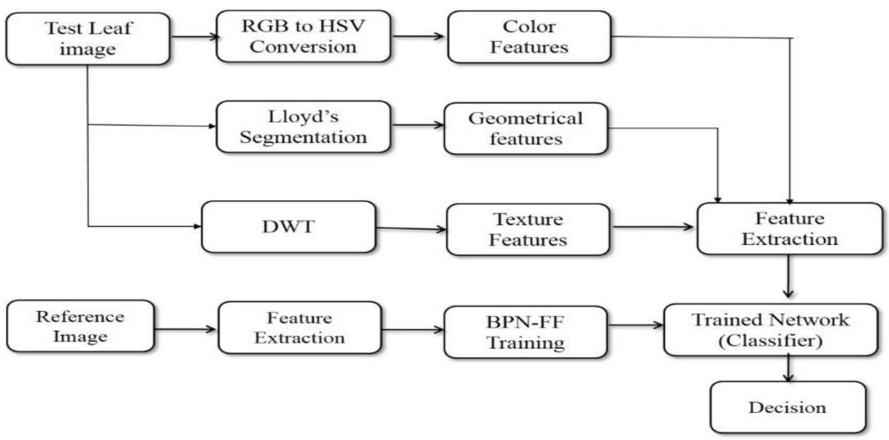

Fig.1 Block diagram of proposed model

IV. HYBRID FEATURES EXTRACTION

In our paper we have incorporated all the hybrid features of a leaf color, texture shape (geometric feature) by the respective methodology.

a. Colour Feature Extraction

The first step is to extract color features. In order to do that we need to convert a RGB image to HSV plane from which the features such as mean, standard deviation, skewness and kurtosis are extracted. The disadvantages of RGB model are that the RGB color coordinates are device dependent. This implies that the RGB model will not in general reproduce the same color from one display to another and also it is difficult to relate this model to color appearance because its basis is to device signals and not display luminance values .HSI stands for Hue, Saturation and Intensity. Hue represents dominant color as perceived by an observer. Saturation refers to the purity or the amount of white light mixed with a hue. Intensity reflects the brightness.

Theoretically Hue component is obtained by

$$H = \begin{cases} \theta & \text{if } B < G \\ 360 - \theta & \text{if } B \geq G \end{cases} \qquad (1)$$

$$\theta = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{\frac{1}{2}}}\right\} \qquad (2)$$

Saturation component is obtained by

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

$$S = 1 - \left[\frac{3}{(R+G+B)}[\min(R,G,B)]\right] \quad (3)$$

Intensity component is given by

$$I = \frac{1}{3}[R + G + B] \quad (4)$$

b. *Texture Feature Extraction*

After color feature extraction, we move on to extract the texture feature. For this purpose we make use of the Discrete Wavelet Transform. Discrete Wavelet Transform is the most popular transform used in image compression. The DWT represents an image as a sum of wavelet functions, known as wavelets, with different location and scale. In DWT the data is represented in a set of high pass (detailed) and low pass (approximation) coefficients .In order to extract the texture features we make use of a 2D-DWT. A 2-D DWT can be seen as a 1 -D wavelet scheme which transform along the rows and then a 1–D wavelet transform along the columns.

In the first stage which is the horizontal decomposition the input data is passed through set of low pass and high pass filters. The output of high pass and low pass filters are down sampled by 2.The output from low pass filter is an approximate coefficient and the output from the high pass filter is a detail coefficient. The second stage comprises of decomposing the outputs of the first stage further along the columns which is the vertical decomposition.

Consider the input to be of size MxN (256×256), at the end of the first stage the image has a size of (M/2) x (N) and at the end of the second stage the image has a size of (M/2) x (N/2). The end result consists of four sub bands LL, LH, HL and HH(128×128 each) where LL represents the smoothing details and LH, HL and HH represent the sharpening details. The alphabet L means low pass signal and H means high pass signal. LH signal is a low pass signal in row and a high pass in column. Hence, LH signal contain horizontal elements. Similarly, HL and HH contains vertical and diagonal elements, respectively. But for more accuracy ,we apply 2D-DWT to LL element which is down-sampled by 2 in both column and vertical decomposition into LL1,LH1,HL1 and HH1(64×64 each).Then the GLCM (Grey Level Co-occurrence Matrix) for a particular sub-band is obtained by the matlab command “graycomatrix( )” , then the textures features are extracted from high frequency components LH1,HL1 and HH1(thus 3×5=15 texture features in total for a single image) using the matlab command “graycoprops( )”(except entropy) while the entropy is calculated by normalizing the COM(Color-Occurrence Matrix) and hence by the respective formula of entropy.

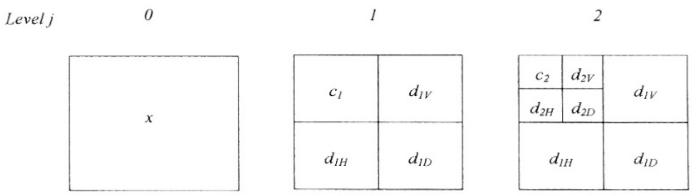

Fig.2 2D-DWT in two successive levels

The texture features extracted by the above process are Energy, Entropy, Contrast, Correlation and Homogeneity.

Energy is a measure the homogeneousness of the image and can be calculated from the normalized COM. It is a suitable measure for detection of disorder in texture image.

$$\text{Energy} = \sum_{i,j=0}^{N-1} p(i,j)^2 \quad (5)$$

Entropy gives a measure of complexity of the image. Complex textures tend to have higher entropy.

$$\text{Entropy} = -\sum_{i,j=0}^{N-1} p(i,j) \ln p(i,j) \quad (6)$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Contrast measures the local variations and texture of shadow depth in the gray level co-occurrence matrix.

$$\text{Contrast} = \sum_{i,j=0}^{N-1}(i,j)^2\, p(i,j) \qquad (7)$$

Correlation-Coefficient measures the joint probability occurrence of the specified pixel pairs.

$$\text{Correlation Coefficient} = \sum_{i,j=0}^{N-1}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j} \qquad (8)$$

Where μ_i, μ_j are the mean and σ_i, σ_j are the standard deviations

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1}\frac{p(i,j)}{1+(i-j)} \qquad (9)$$

c. *Shape Feature Extraction*

Lloyds Clustering: The segmentation refers to the process of partitioning a digital image into multiple segments. The goal is to simplify and change the representation of an image into something that is more meaningful and easier to analyse. The segmentation is performed by using Lloyd's clustering algorithm and iterative process of K-means clustering. It is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other.

The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. Initially for a 256x256 grayscale image four clusters are chosen. The next step involves finding the centroid of each cluster. This is determined theoretically by calculating the ratio of difference between the Maximum intensity and the Minimum intensity of the image to the total number of clusters considered initially.

$$\text{Centroid} = \frac{(I_{max} - I_{min})}{\text{No. of clusters}}$$

In such way the centroid of all four clusters are calculated. The final step involves the repartitioning of the image into the clusters. The distance between an individual pixel of the image and the centroids of the four clusters are calculated. The minimum of the four distances is determined and the pixel is moved to the corresponding cluster. The centroid of the cluster is updated every time a pixel is moved into the cluster. This occurs iteratively until there are no more pixels left in the image to repartition. The four clusters are then obtained and are subjected to get their respective Histogram to determine the two clusters having more relevant information. This is done by executing the number of occurrences of each pixel value of the particular cluster. Then Probability Density Function of each pixel in the cluster and thereby using the extracted values to train the BPN-FF classifier.

V. SUPERVISED CLASSIFIER - BPN-FF

Once all the hybrid features are extracted, the next step is to train the ANN (Artificial Neural Network) and then test the ANN with test samples. Consider a network with a single real input x and network function F. The derivative F'(x) is computed in two phases:

Feed-forward: the input x is fed into the network. The primitive functions at the nodes and their derivatives are evaluated at each node. The derivatives are stored.

Back propagation: The constant (say 1) is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is multiplied by the value stored in the left part of the unit. The result is transmitted to the left of the unit. The result collected at the input unit is the derivative of the network function with respect to x.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

The back propagation algorithm can be decomposed in the following four steps: (i) Feed-forward computation, (ii) Back propagation to the output layer, (iii) Back propagation to the hidden layer, (iv)Weight updates. The algorithm is stopped when the value of the error function has become sufficiently small.

A. Classification

The classification of the leaf disease can be achieved by two different phases (i) Training the network with feature data and (ii) Testing the network with feature data. The extracted features are used to train the classifier BPN-FF and thereby creating a target vector matrix which the network uses to classify the test leaf samples. For BPN networks, the decision layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category.

VI. SIMULATION RESULTS

From the input images collected, the corresponding hybrid features are extracted so as to form the target vector matrix which will be utilized by the BPN-FF classifier to classify the test leaf images into any one of the four abnormalities and hence provide the corresponding cure for the disease.

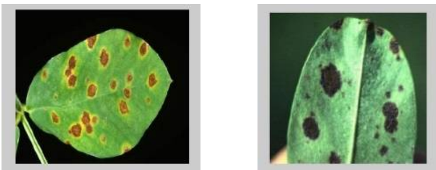

Fig.3 Input Images

The above results showcase the input images of two leaves Fig.3 from the available database

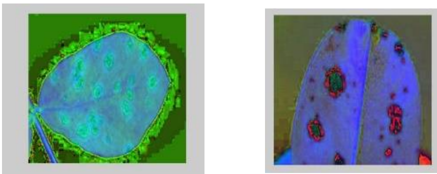

Fig.4 HSV color conversion of images

The first step is the color space conversion of RGB image to HSV plane as shown in Fig.4.

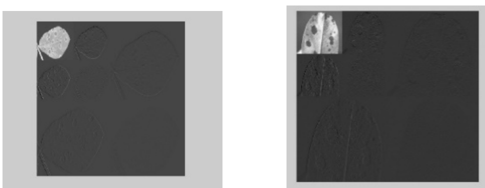

Fig.5 2D-DWT of grayscale images

Upon Gray-scale conversion and applying 2-level 2-D DWT the simulation result is shown in Fig.5 and hence the texture features are extracted from LH1,HL1,HH1 frequency components which results in 3×5=15 texture features.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

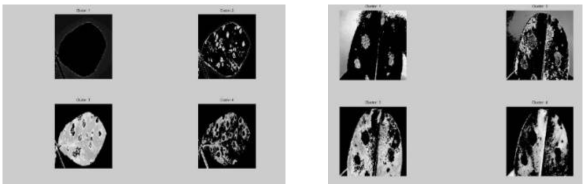

Fig.6 Output of Lloyd’s clustering

On applying the Lloyd’s segmentation, the input image is divided into four clusters as shown in Fig.6, from which the cluster having maximum abnormality is chosen for extracting geometric features.

Fig.7 Final Output of Classifier

Then BPN-FF classifier makes use of these extracted features and hence classifies the type of disease or as normal leaves with respect to the data availed from target vector matrix Fig.7.

The color and texture features for about eight input images are tabulated in Table.1.

TABLE I

COLOR AND TEXTURE VALUES OF EARLY & LATE LEAF SPOT

Color and Texture Features	**Leaf 1** (1.0e+003*)	**Leaf2** (1.0e+003*)
Mean of Hue	0.0003	0.0002
Mean of Saturation	0.0003	0.0005
Standard Deviation of Hue	0	0
Standard Deviation of Saturation	0	0
Kurtosis of Hue	2.7769	1.9343
Kurtosis of Saturation	1.5787	1.2446
Skewness of Hue	0.0585	-0.1007
Skewness of Saturation	-0.1684	0.2754
Energy LH1	0.0002	0.0002
Contrast LH1	0.0195	0.0137
Correlation LH1	0.0002	0.0004
Homogeneity LH1	0.0006	0.0007

Color and Texture Features	**Leaf 1** (1.0e+003*)	**Leaf2** (1.0e+003*)
Entropy LH1	0.0026	0.0026
Energy HL1	0.0002	0.0002
Contrast HL1	0.0227	0.0181
Correlation HL1	0	0.0001
Homogeneity HL1	0.0006	0.0006
Entropy HL1	0.0026	0.0029
Energy HH1	0.0002	0.0002
Contrast HH1	0.0237	0.0189
Correlation HH1	0	0
Homogeneity HH1	0.0005	0.0006
Entropy HH1	0.0028	0.0031

VII. CONCLUSION AND FUTURE WORKS

In this paper we have proposed an image-processing based approach to automatically classify the normal or diseased leaves (Early leaf spot, Late leaf spot, Alternaria leaf spot) and also provide the cure for the same which would be beneficial to beginners in farming or gardening (as these diseases are common in flowering plants like rose as well) . In our approach we have incorporated all the hybrid features of a leaf to train the ANN (BPN-FF) and have made use of Lloyd's clustering which is more efficient than the traditional K-means clustering to segment the test images.

As of now we have worked on the images pertaining to groundnut plant and we have planned to expand it to other plants like mango, potato etc., and we are working on to include more diseases and the number of samples in our database to increase the accuracy rate of disease identification.

REFERENCES

[1] Dheeb Al Bashish, MalikBraik, and Sulieman, B., "A Framework for Detection and Classification of Plant Leaf and Stem Diseases",IEEE International Conference on Signal and Image Processing2010.
[2] Niket, A., Bharat Jadhav, Smeeta, N., "Detection and classification of plant diseases by image processing",International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 2, April 2014.
[3] Arivazhagan, S., NewlinShebiah, R., Ananthi, S., Vishnu Varthini,S., "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", Vol.15, CIGR Journal, March 2013.
[4] GurpreetKaur, Himanshu, M., "Classification of biological species based on leaf architecture", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No.2, April 2012.
[5] Sanjeev, S., Vijay, S., Nargund, V. B., Pallavi "Diagnosis and classification of grape leaf diseases using neural network", IEEE-2013 4th ICCCNT 2013, July 4-6, 2013.
[6] EvyKamilah, R., MustikaMentari, RatihKartikaDewi, HariGinardi, R. V., "Sugarcane leaf Disease detection and severity estimation based on segmented spots image"- International Conference on Information, Communication Technology and System, 2014.
[7] Murali Krishnan,Dr.Sumithra, M.G.," A Novel Algorithm for Detecting Bacterial Leaf Scorch (BLS) of Shade Trees Using Image Processing", IEEE 11th Malaysia International Conference on Communications 26th - 28th November 2013
[8] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li ," Application of Neural Networks to Image Recognition of Plant Diseases ",International Conference on Systems and Informatics (ICSAI 2012).
[9] John William, O., Jennifer, D., Leobelle, A., Paul Jensen, S., Ira, V., " Identification of Diseases in Rice Plant (Oryza Sativa) using Back Propagation Artificial Neural Network ",7th IEEE International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM) The Institute of Electrical and Electronics Engineers Inc. (IEEE) – Philippine Section 12-16 November 2013.
[10] Prasad Babu, M. S.,SrinivasaRao, B., "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. Downloaded from www.indiakisan.net on May 2010