Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review

[1]Savita N. Ghaiwat, [2]Parul Arora

GHRCEM, Department of Electronics and Telecommunication Engineering, Wagholi, Pune
Email: [1]savita.pusande@gmail.com, [2]parul.arora@raisoni.net

Abstract-- This paper present survey on different classification techniques that can be used for plant leaf disease classification. A classification technique deals with classifying each pattern in one of the distinct classes. A classification is a technique where leaf is classified based on its different morphological features. There are so many classification techniques such as k-Nearest Neighbor Classifier, Probabilistic Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis, Artificial neural network, Fuzzy logic. Selecting a classification method is always a difficult task because the quality of result can vary for different input data. Plant leaf disease classifications have wide applications in various fields such as in biological research, in Agriculture etc. This paper provides an overview of different classification techniques used for plant leaf disease classification.

Index items— k-Nearest Neighbor, Fuzzy logic, ANN, SVM.

I. INTRODUCTION

The purpose of Agriculture is not only to feed ever growing population but it's an important source of energy and a solution to solve the problem of global warming. Plant diseases are extremely significant, as that can adversely affect both quality and quantity of crops in agriculture production. Plant disease diagnosis is very essential in earlier stage in order to cure and control them. Generally the naked eye method is used to identify the diseases. In this method experts are involved who have the ability to detect the changes in leaf color. This method involves lots of efforts, takes long time and also not practical for the large fields. Many times different experts identify the same disease as the different disease. This method is expensive as it requires continuous monitoring of experts.

Depending on the applications, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing, pattern recognition and some automatic classification tools. In the next section this paper tries to present those proposed systems in meaningful way.

II. REVIEW OF LITERATURE

Yan-cheng zang Han-Ping Mao, Bo Hu, Ming-xi Li in paper titled Feature Selection of Cotton Disease Leaves Image Based on Fuzzy Feature Selection Techniques[1] proposed the fuzzy feature selection approach -fuzzy curves (FC) and surfaces (FS) – for cotton leaves disease image feature selection. This research is done in two steps .Firstly to automatically and quickly isolate a small set of significant features from a set of original features according to their significance and to eliminate spurious features they make use of FC. Secondly to isolate the features dependent on the significant features, utilize FS. This approach is useful for practical classification applications which reduce the dimensionality of the feature space.

The feature selection technique has faster execution speed and higher classification success rate because it does not suffer from the local minima problems inherent in the nonlinear modeling techniques typically used in forward selection and backward elimination.

Ajay A. Gurjar, Viraj A. Gulhane describes Eigen feature regularization and extraction technique by this detection of three diseases can be done. This system is having more accuracy, than that of the other feature detection techniques. With this method about 90% of detection of Red spot i.e. fungal disease is detected [2].

Dheeb Al Bashish & et al. proposed image processing based work is consists of the following main steps : In the first step the acquired images are segmented using the K-means techniques and then secondly the segmented images are passed through a pre-trained neural network .The images of leaves taken from Al-Ghor area in Jordan. Five diseases that are prevalent in leaves were selected for this research; they are: Early scorch, Cottony mold, Ashen mold, late scorch, tiny whiteness. The experimental result indicates that the neural network classifier that is based on statistical

classification support accurate and automatic detection of leaf diseases with a precision of around 93% [3].

In [4], diagnosis system for grape leaf diseases is proposed. The proposed system is composed of three main parts: Firstly grape leaf color extraction from complex background, secondly grape leaf disease color extraction and finally grape leaf disease classification.

In this analysis back-propagation neural network with a self-organizing feature map together is utilize to recognize colors of grape leaf. Further MSOFM and GA deployed for grape leaf disease segmentation and SVM for classification. Finally filtration of resulting segmented image is done by Gabor Wavelet and then SVM is again applied to classify the types of grape leaf diseases. This system can classify the grape leaf diseases into three classes: Scab disease, rust disease and no disease. Even though there are some limitations of extracting ambiguous color pixels from the background of the image. The system demonstrates very promising performance for any agricultural product analysis.

In [5] "Extraction of the Rice leaf disease image based on BP neural network " by Libo Liu & et al proposed a system for classifying the healthy and diseased part of rice leaves using BP neural network as classifier. In this study rice brown spot was select as a research object. The images of rice leaves were acquired from the northern part of Ningxia Hui autonomous region. Here the color features of diseases and healthy region were served as input values to BP neural network. The result shows that this method is also suitable to identify the other diseases.

In [6], Tushar H Jaware & et al. developed a Fast and accurate method for detection and classification of plant diseases. The proposed algorithm is tested on main five diseases on the plants; they are: Early Scorch, Cottony mold, Ashen Mold, Late scorch, tiny whiteness. Initially the RGB image is acquired then a color transformation structure for the acquired RGB leaf image is created. After that color values in RGB converted to the space specified in the color transformation structure.

In the next step, the segmentation is done by using K-means clustering technique. After that the mostly green pixels are masked. Further the pixels with zero green, red and blue values and the pixels on the boundaries of the infected object were completely removed. Then the infected cluster was converted into HIS format from RGB format. In the next step, for each pixel map of the image for only HIS images the SGDM matrices were generated. Finally the extracted feature was recognized through a pre-trained neural network. The results show that the proposed system can successfully detect and classify the diseases with a precision between 83% and 94%.

P. Revathi M. Hemalatha detected Cotton leaf spot diseases in [7] by using Homogenous Segmentation based Edge Detection Techniques. This system is analyzed with eight types of cotton leaf diseases they are Fusarium wilt, Verticillium wilt,Root rot,Boll rot,Grey mildew ,Leaf blight ,Bacterial blight,Leaf curl.In these work symptoms of cotton leaf spot images are captured by mobile and classification is done by using neural network.

In this work a homogeneity operator can take the difference of the center pixel and a pixel that is two or three pixels away. The main aim Research work is to use Homogeneity-based edge detector segmentation, which takes the result of any edge detector and divides it by the average value of the area. This work has been implemented in the real time software and produces best results. The software is very fast and time intense, low cost, automatically identify the diseases and pest recommendation to farmers through a mobile phone.

H. Muhammad Asraf and others has proposed, Support Vector Machine (SVM) as classifier with three different kernels namely linear kernel, polynomial kernel with soft margin and polynomial kernel with hard margin [8]. Initial results show that the recognition of oil palm leaves is possible to be performed by SVM classifier. Based on the best performance result, polynomial kernel with soft margin is capable of classifying nutrient diseases accurately in the oil palm leaves with accuracy of 95% of correct classification.

Polynomial kernel with soft margin produces the best performance in average of 95% correct classification as compared to the other types of Kernel function.

In paper titled Pixel-Based Classification Method for Detecting Unhealthy Regions in Leaf Images Satish Madhogaria & et al. implemented an automatic pixel-based classification method for detecting unhealthy regions in leaf images [9]. This proposed system is composed of three main steps .In First step segmentation to divide the image into foreground and background. In the second step, support vector machine (SVM) is applied to predict the class of each pixel belonging to the foreground. And finally, further refinement by neighborhood-check to omit all falsely-classified pixels from second step.

The results presented in this work are based on a model plant (Arabidobsis thaliana), which forms the ideal basis for the usage of the proposed algorithm in biological researches concerning plant disease control mechanisms. The proposed method is compared to the existing method and it is concluded that higher accuracy can be achieved with this method.

Paper [10] presents a method to monitor four main wheat plant diseases: Powdery Mildew, leaf rust Puccinia triticina, leaf blight, Puccinia striiformis. This proposed system by Yuan Tian & et al make use of

stacked structure to combine the results obtained from three support vector machines (SVMs)-based classifiers. The three features obtained are color feature, texture feature, and shape feature which further used as training sets for three corresponding classifiers. This system is mainly divided into three main steps: data acquisition, feature extraction, and classifier design. Multiple Classifier System (MCS) includes number of classifiers which can provide higher classification accuracy.

T. Rumpf & et al.[11] developed a system for the detection and differentiation of sugar beet diseases based on Support Vector Machines and spectral vegetation indices. They used Cercospora leaf spot, leaf rust and powdery mildew diseased leaves as study samples .The main aim was to identify these diseases before their symptoms became visible. In this proposed work nine spectral vegetation were used as features for an automatic classification. The experimental result indicates that the discrimination between healthy sugar beet leaves and diseased leaves classification accuracy up to 97%.

S. Phadikar, J. Sil, and A. K. Das [12] developed an automated classification system based on the morphological changes caused by brown spot and the leaf blast diseases of rice plant. To classify the diseases Radial distribution of the hue from the centre to the boundary of the spot images has been used as feature by using Bayes' and SVM Classifier.

The feature extraction for classification of rice leaf diseases is processed in the following steps: firstly images acquired of diseased rice leaves from fields. Secondly preprocessing the images to remove noise from the damaged leaf and then enhanced the quality of image by using the [mean filtering technique. Thirdly Otsu's segmentation algorithm was applied to extract the infected portion of the image, and then radial hue distribution vectors of the segmented regions computed which are used as feature vectors.

Here classification performed in two different phases .In first phase uninfected and the diseased leaves are classified based on the number of peaks in the Histogram.In the second phase the leaf diseases are classified by Bayes' classifier.This system gives 68.1% and 79.5% accuracies for SVM and Bayes'classifier based system respectively.

The robust technique for fast and accurate detection & classification is proposed in [13].The whole technique is described in the number of steps: Initially RGB images of leaves acquired then a color transformation structure is created for the RGB image. In the next step the color values in RGB converted to the space specified in the color transformation structure. Then images are segmented by applying K-means clustering. Further mostly green pixels are identified and then these pixels are masked. In next step pixels on the boundaries of the infected cluster and the pixels with zeros red, green and blue values were completely removed. Then the infected cluster was converted from RGB format to HSI format. In the next step, the SGDM matrices were then generated for each pixel map of the image for only H and S images. Features are calculated by calling the GLCM functions. After Texture Statistics Computation finally the recognition process was performed through a pre-trained Neural Networks.

Sannakki S.S. & et al. [14] in paper titled "A Hybrid Intelligent System for Automated Pomegranate Disease Detection and Grading " proposed a system not only identifies various diseases of pomegranate plant but also determines the stage in which the disease is. The methodology is divided into four steps: (1) The images acquisition where the images were captured by using digital camera.(2) The image preprocessing creates enhanced image that is more useful for human observer. Image preprocessing uses number of techniques like image resize, filtering,

segmentation, morphological operations etc.(3) Once the image has been enhanced and segmented in image postprocessing noises like stabs, empty holes etc. are removed by applying morphological operations, region filling. Further the features are extracted like color , shape ,texture.(4) Once the features are extracted to which disease class the query image in belongs different machine learning techniques are used like Artificial neural networks, Decision tree learning, genetic algorithms, Clustering, Bayesian networks, Support Vector Machines, Fuzzy Logic etc.

Pranjali Vinayak Keskar & et al.[15] developed a leaf disease detection and diagnosis system for inspection of affected leaves and identifying the type of disease. This system is comprised of four stages: To improve the appearance of acquired images image enhancement techniques are applied. The enhancement is done in three steps: Transformation of HSI to color space in first stage .In the next stage analyzing the histogram of intensity channel to get the threshold. Finally intensity adjustment by applying the threshold. . The second stage is segmentation which includes adaption of fuzzy feature algorithm parameter to fit the application in concern. The feature ex traction stage is comprised of two steps spot isolation and spot extraction. For identification of spot identification algorithm is used is called component labeling. In feature extraction phase three features are extracted namely color, size and shape of the spots. In fourth stage classification is performed by Artificial Neural Network.

III. CLASSIFICATION TECHNIQUES

This section will discuss some of the popular classification techniques that are used for plant leaf classification. In plant leaf classification leaf is classified based on its different morphological features. Some of the classification techniques used are Neural

Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis, k-Nearest Neighbor Classifier. Plant leaf disease classification has wide application in Agriculture.

1) k-Nearest Neighbor:

k-Nearest Neighbor is a simple classifier in the machine learning techniques where the classification is achieved by identifying the nearest neighbors to a query examples and then make use of those neighbors for determination of the class of the query. In KNN the classification i. e. to which class the given point is belongs is based on the calculation of the minimum distance between the given point and other points. As a classifier the nearest neighbor does not include any training process .It is not applicable in case of large number of training examples as it is not robust to noisy data. For the plant leaf classification the Euclidean distance between the test samples and training samples is calculated. In this way it finds out similar measures and accordingly the class for test samples. A sample is classified based on the highest number of votes from the k neighbors, with the sample being assigned to the class most common amongst its k nearest neighbors.k is a positive integer, typically small. If k = 1, then the sample is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes [19][20].

Nearest neighbor method is easy to implement also quite good results if the features are chosen carefully. The K-Nearest Neighbor (KNN) Classifier is works well on basic recognition problems.

The main disadvantage of the KNN algorithm is that it is a slow learner, i.e. it does not learn anything from the training data and simply make use the training data itself for classification. Another disadvantage is this method is also rather slow if there are a large number of training examples as the algorithm must have to compute the distance and sort all the training data at each prediction. Also it is not robust to noisy data in case of large number of training examples. The most serious disadvantage of nearest neighbor methods is that they are very sensitive to the presence of irrelevant parameters.

2) Support Vector Machine:

Support Vector machine (SVM) is a non-linear Classifier. This is a new trend in machine learning algorithm which is used in many pattern recognition problems, including texture classification. In SVM, the input data is non-linearly mapped to linearly separated data in some high dimensional space providing good classification performance. SVM maximizes the marginal distance between different classes. The division of classes is carried out with different kernels.SVM is designed to work with only two classes by determining the hyper plane to divide

two classes. This is done by maximizing the margin from the hyper plane to the two classes. The samples closest to the margin that were selected to determine the hyper plane is known as support vectors .Fig below shows the support vector machines concept. Multiclass classification is also applicable and is basically built up by various two class SVMs to solve the problem, either by using one-versus-all or one versus-one. The winning class is then determined by the highest output function or the maximum votes respectively [18] [19] [20].

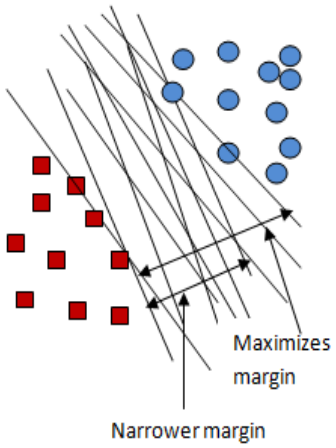

Fig.1 Support vector machine

Main advantages of SVM are:

- Its prediction accuracy is high.
- Its working is robust when training examples contain errors.
- Its simple geometric interpretation and a sparse solution.
- Like neural networks the computational complexity of SVMs does not depend on the dimensionality of the input space.

Drawbacks of SVM are:

- This classifier involves long training time.
- In SVM it is difficult to understand the learned function (weights).
- The large number of support vectors used from the training set to perform classification task.

3) Artificial Neural Network (ANN):

An Artificial Neuron is basically an engineering approach of biological neuron.ANN consists of a number of nodes, called neurons. Neural networks are typically organized in layers. In neural network each

neuron in hidden layer receives signals from all the neurons in the input layer. The strength of each signal and the biases are represented by weights and constants, which are calculated through the training phase. After the inputs are weighted and added, the result is then transformed by a transfer function into the output. The transfer functions used are Sigmoid, hyperbolic tangent functions or a step. Backpropagation is a neural network learning algorithm (Rumelhart and McClelland, 1986) is used in layered feed-forward Artificial Neural Networks. Backpropagation is a form of supervised training [21][17].

Originally, ANNs started in the form of a single neuron, proposed in the McCulloch and Pitts model in the 1940s (McCulloch & Pitts, 1943).

In 1958, frank Rosenblatt proposed Perceptron, is the simplest single layer networks whose weights and biases could be trained to produce a correct target vector when presented with the corresponding input vector. This network made up of only input neurons and output neurons. It can solve only linear problems.

Multi Layer perceptron (MLP) is one of the feedforward neural network with one or more layers between input and output layer. Feedforward means that data flows in one direction from input to output layer (forward). Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

Multilayer perceptrons (MLPs), which can be trained using a backpropagation algorithm (Rumelhart and McClelland, 1986), is a very popular choice for many researchers.

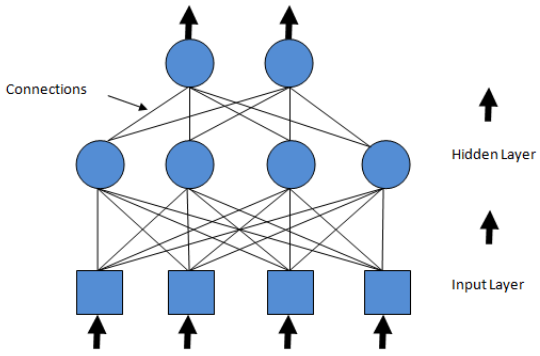

Fig.2 Multilayered Artificial Neural Network

4) Self-Organizing Map

The Self-Organizing Map is one of the neural network models, based on unsupervised learning (human intervention is needed during the learning). The Self-Organizing Map was developed by professor Kohonen in 2001.Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space.

A self-organizing map consists of components called nodes or neurons. Each node has a weight vector of the same dimension as the input data vectors and a position in the map space. The nodes are usually arranged in a two-dimensional regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the smallest distance weight vector to the data space vector [17].

5) Probabilistic Neural Networks

Probabilistic Neural Networks (PNNs) is a feedforward neural network , based on Parzen windows. In a PNN, the operations are organized into a multilayered feedforward network with four layers.PNN is mainly used in classification problems. The first layer is input layer which calculates the distance from the input vector to the training input vectors. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Third Pattern layer contains one neuron for each case in the training data set. It stores the values of the predictor variables for the case along with the target value. The pattern neurons add the values for the class they represent. The output layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category. As PNNs are much faster than multilayer perceptron networks their training phase requires only one pass through the training patterns. PNN can be accurate than multilayer perceptron networks also relatively insensitive to outliers. To improve the overall performance PNNs output can be later processed by another classification system and as this happens very fast, PNNs are used in on-line applications where a real-time classifier is required [17][18] [19] [20].

The main disadvantage of PNN is it requires large storage space.

6) Fuzzy Logic

Fuzzy Logic classifiers are classification systems that make use of fuzzy sets or fuzzy logic (Kuncheva, 2000) which convert real-world data values into membership degrees through the use of the membership functions (Zadeh, 1965) so that these rules then can be used for the classification process. This is done by defining “categories” for each one of the attributes [17] [18].

As Fuzzy logic classifier’s has very high speed they are preferable in cases where there is limited precision in the data values or when classification is required in real time. Fuzzy image processing is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. The

representation and processing depend on the selected fuzzy technique and on the problem to be solved.

Fuzzy image processing is divided into three main steps: image fuzzification, modification of membership values, and, if necessary, image defuzzification .

Because of the uncertainties that exist in many aspects of image processing like additive and non-additive noise in lowlevel image processing, imprecision in the assumptions underlying the algorithms, and ambiguities in interpretation during high level image processing , fuzzy processing is desirable.

The main drawback of Fuzzy logic as classifier is dimensionality because of this classifier is inadequate for problems having a large number of features. Also it gives poor performance while there is a limited amount of knowledge that the designer can incorporate in the system.

IV. CONCLUSION

From study of above classification techniques we come up with following conclusion. The k-nearest-neighbor method is perhaps the simplest of all algorithms for predicting the class of a test example. An obvious disadvantage of the k-NN method is the time complexity of making predictions. Additionally, neural networks are tolerant to noisy inputs. But in neural network it's difficult to understand structure of algorithm. SVM was found competitive with the best available machine learning algorithms in classifying high-dimensional data sets. In SVM computational complexity is reduced to quadratic optimization problem and it's easy to control complexity of decision rule and frequency of error. Drawback of SVM is it's difficult to determine optimal parameters when training data is not linearly separable. Also SVM is more complex to understand and implement.

V. REFERENCES

[1] Yan Cheng Zhang, Han Ping Mao, Bo Hu, Ming Xili ,"features selection of Cotton disease leaves image based on fuzzy feature selection techniques", IEEE Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007.

[2] Ajay A. Gurjar, Viraj A. Gulhane," Disease Detection On Cotton Leaves by Eigenfeature Regularization and Extraction Technique", International Journal of Electronics, Communication & Soft Computing Science and Engineering(IJECSCSE)Volume 1, Issue 1

[3] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad ,"A Framework for Detection and Classification of Plant Leaf and Stem Diseases", 2010 International Conference on Signal and Image Processing.

[4] A.Meunkaewjinda, P. Kumsawat, K. Attakitmongcol and Sri kaew,"Grape leaf disease detection from color imagery using hybrid intelligent system",proceedings of ECTI-CON 2008.

[5] Libo Liu,Guomin Zhou,"Extraction of the Rice Leaf Disease Image Based on BP Neural Network",2009 IEEE.

[6] Tushar H Jaware, Ravindra D Badgujar and Prashant G Patil," Crop disease detection using image segmentation", Proceedings of "Conference on Advances in Communication and Computing (NCACC'12)", April 21, 2012.

[7] P. Revathi M. Hemalatha," Homogenous Segmentation based Edge Detection Techniques for Proficient Identification of the Cotton Leaf Spot Diseases", International Journal of Computer Applications (0975 – 888)Volume 47– No.2, June 2012.

[8] H. Muhammad Asraf, M. T. Nooritawati, M.S.B. Shah Rizam," A Comparative Study in Kernel-Based Support Vector Machine of Oil", Procedia Engineering 41 (2012) 1353 – 1359

[9] Satish Madhogaria, Marek Schikora, Wolfgang Koch, Daniel Cremers," Pixel-Based Classification Method for Detecting Unhealthy Regions in Leaf Images", Informatik 2011 - Informatik schafft Communities 41. Jahrestagung der Gesellschaft für Informatik , 4.-7.10.2011, Berlin

[10] 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, Yichang, China.

[11] T. Rumpf,A.K. Mahlein,U.Steiner,E.C.Oerke, H.W. Dehne, L.Plumer,"Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance",2010 Elsvier B.V.

[12] S. Phadikar, J. Sil, and A. K. Das," Classification of Rice Leaf Diseases Based on Morphological Changes", International Journal of Information and Electronics Engineering, Vol. 2, No. 3, May 2012

[13] Sannakki S.S., Rajpurohit V.S., Nargund V.B., Arun Kumar R.and Yallur P.S.," A Hybrid Intelligent System for Automated Pomgranate Disease detection and Grading", International

Journal of Machine Intelligence ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 2, 2011, pp-36-44

[14] Pranjali Vinayak Keskar, Shubhangi Nimba Masare, Manjusha Suresh Kadam and Prof. Mrs. Seema U.Deoghare," Leaf Disease Detection and Diagnosis", International Journal of Emerging Trends in Electrical and Electronics (IJETEE) Vol. 2, Issue. 2, April-2013.

[15] Zacharias N. Voulgaris,"Discernibility Concept in Classification Problems", University of London may 2009.

[16] Abdallah A. Alshennawy, and Ayman A. Aly," Edge Detection in Digital Images Using Fuzzy Logic Technique", World Academy of Science, Engineering and Technology 27 2009

[17] Vishakha Metre,Jayshree Ghorpade," An Overview of the Research on Texture Based Plant Leaf Classification", IJCSN International Journal of Computer Science and Network, Vol 2, Issue 3, 2013ISSN (Online): 2277-5420

[18] Prof. Meeta Kumar, Mrunali Kamble, Shubhada Pawar, Prajakta Patil,Neha Bonde," Survey on Techniques for Plant Leaf Classification", International Journal of Modern Engineering Research (IJMER) Vol.1, Issue.2, pp-538-544 ISSN: 2249-6645

[19] Noor Ezan Abdullah, Athirah A. Rahim, Hadzli Hashim and Mahanijah Md Kamal," Classification of Rubber Tree Leaf Diseases Using Multilayer Perceptron Neural Network", The 5tn Student Conference on Research and Development -SCOReD 200711-12 December 2007, Malaysia

◈◈◈