Classification Methods Used In Plant Leaf Disease Identification: A Review

Sushil R.Kamlapurkar
Karmaveer Kakasaheb Wagh Institute of Engineering Education and Research,Nashik

***Abstract*—**The paper presents survey of different classification methods in plant disease. The symptoms of disease can be observed on the parts of the plants such as leaf, stems, lesions and fruits. The leaf shows the symptoms by changing colour, showing the spots on it. A classification is a technique that deals with classification of pattern of its own class of disease. A leaf is classified according to its morphological features. There are different classification techniques used such as k-Nearest neighbour, Principle Component Analysis, Probabilistic Neural Network, Artificial Neural Network, Support Vector Machine, Fuzzy Logic. Selection of the technique is a very difficult task as result quality varies with data.

Keywords— *Principal Component Analysis, Support Vector machine, Artificial Neural Network, Fuzzy Logic*

I. INTRODUCTION

Now days, self operating systems are employed to monitor and to control the field conditions in smart farming. Such self operating systems recognize the disease based on the identification of the symptoms of disease. Therefore, the occurrence of the disease can be quickly and accurately identified and this information is provided to the farmers, experts and researchers. This reduces the need of human being to monitor large fields. The detection of disease from image the key is to extract the characteristic feature of the diseased region. With the disease its features may vary. The features extracted from the image are color, shape, texture etc.Sometimes more features are to be extracted for accurate identification and detection of the disease. This may result in increasing cost of hardware as well as software tools. This ultimately causes complexity and computation time. Hence only required features must be extracted.

The automated systems are used to minimize the problems by using image processing, pattern recognition and automated tools. The next section of paper tries to represent those systems in meaningful manner.

II. LITRATURE REVIEW

The paper titled New Optimize Spectral Indices for Identifying and monitoring Winter Wheat Diseases the authors have proposed an effective method for indirect monitoring of plant diseases by developing vegetation indices based on the hyper spectral data. Three different diseases considered are powdery mildew, yellow rust and aphids on winter wheat in China. First using spectrometer reflectance and transmittance of the upper faces of leaves is measured. Then using RELIF-F algorithm most and ye least relevant wavelength for different wavelengths are extracted. To obtain the indices the reflectance of the a single band extracted from the most relevant wavelengths and the normalized wavelength difference from all possible combinations of the most and least relevant wavelengths were used. For these indices the classification accuracy obtained is: Yellow rust index 91.6%, Powdery mildew index 85.2% and Aphid index93.5% [1].

Sanjeev S Sannakki, Vijay S Rajpurohit, V B Nargun, Pallavi Kulkarni used k-means clustering and back propagation network to recognise the diseases on grape leaf. First background is removed using masking. In second stage, image is pre-processed using antistrophic diffusion to prevent an affected region. Segmentation is done by K-means clustering. To extract features of diseased parts Gray Level Co-occurrence Matrix is used. Classification of disease is done by using three layer back-propagation networks [2].

Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li used Principal Component Analysis for detection of plant leaf disease from images. The images are of two wheat and two grape disease.PCA is used for feature data reduction. For wheat diseases detection is based on PCA and BP network. With PCA and BP network the accuracy of detection is 100%.Whereas for grape diseases the accuracy is 94.29%[3].

Jun Pang and Zohng ying-Bai identified leaf spot on images using Integrating Local Threshold and Seed growing region. As leaf colour and spot is uneven, authors used improved segmentation method for detection spot. The method is called LTSRG (Local Thresholding and Seeded Region Growing). The results are compared with threshold based Ostu and threshold based EM. On comparing its found that LTSRG is easy to realize and improve the spot image segmentation [4].

Dheeb Al Bashish,Malik Braik, and Suleiman Bani - Ahmead used K-means clustering and Neural network for detection and classification of leaf disease. Leaf images are taken from area in Jordan for study. The system has four phases. In the first phase colour transformation structure is used for leaf image and then device independent colour space transformation is applied. Second phase is segmentation using k-means clustering .In the next phase, feature extraction is done by using colour co-occurrence methods. In the last phase, the selected features are passed through trained Neural network. Using Neural network the diseases are detected with precision around 93% [5].

The paper titled SVM based Multiple Classifier System for Recognition of Wheat leaf Diseases use different classifiers to obtain higher accuracy. These classifiers can complement the errors created by using them individually .The design of MCS consists of designing of classifier ensemble and design of

decision function. This decision function produces output using input and of MCS. The systems provide accuracy of MCS 95.16% compared with SVM, ANN, KNN [6].

A paper titled Grape leaf Disease Detection from Colour Imagery using Hybrid Intelligent System develops a system that eliminates need of expert. The system contains three parts leaf colour segmentation, leaf disease segmentation and analysis and classification. In leaf colour segmentation, irrelevant background in image is removed. Leaf disease segmentation is done by self organizing map with genetic algorithm .The obtained image is filtered by Gabor wavelet and finally disease is identified using SVM. The system classifies image as Scab disease, Rust disease and No disease [7].

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 proposed the fuzzy feature selection - fuzzy curves (FC) and surfaces (FS) for cotton leaves disease image feature selection. Using fuzzy curves isolate a small set of significant features from a set of original features according to their significance and eliminate spurious features. Apply fuzzy surfaces to isolate the features dependent on the significant features. This is useful to reduce the dimensionality of the feature space. It has faster execution speed and higher classification rate [8].

III. CLASSIFICATION TECHNIQUES

This section will discuss some of the popular classification techniques used for plant leaf classification. The classification is based on different morphological features of leaf. Some of the classification techniques used are K-nearest neighbour, Artificial Neural Network, Support Vector Machine, Fuzzy Logic, Principal Component Analysis

(1)*K-Nearest Neighbour:* It is a simple type of classifier in machine learning.The classification is achieved by identifying the nearest neighbours to a query examples and then using those neighbours for determination of the class. In KNN the classification the minimum distance between the given point and other points. It does not include any training process. For the classification, the Euclidean distance between the test samples and training samples is calculated. Thus similar measures and accordingly the class for test samples is determined. Based upon the highest number of votes from the k neighbours, the sample is assigned to the most common class amongst its k nearest neighbours where k is a positive integer, typically small. If k = 1, then the sample gets assigned to the class of its nearest neighbour. In two class, classification problems, selecting k to be an odd number avoids tied votes. The results are good enough if selection if features is done carefully. But it is a slow learner. Another drawback is its not robust to noisy data if used for large number of training examples.

(2)*Support Vector Machine:* Support Vector machine (SVM) is a non-linear Classifier. Basically it is developed for two classes. SVM can analyse the linear separable instance directly. For non-linear separable instance it uses non –linear mapping algorithm. This algorithm maps low dimensional linear sample of input space to high dimensional feature space. Based on Structural Risk Minimization a segmentation hyper plane is developed in feature space. In SVM, different SVM algorithms adopt different kernels. The hyper plane divides two classes. This is done by maximizing the margin from the hyper plane to the two classes. The samples that are close to the margin are known as support vectors. Multiclass classification is also applicable.

Advantages of SVM are:

- Simple geometric interpretation.
- Computational complexity does not depend on the dimensionality of the input space.
- Robust working even if training examples contain errors.
- It's simple geometric interpretation and a sparse solution.

Drawbacks of SVM are:

- Time consuming.
- Difficult to understand the learned function (weights).
- For classification large no of support vectors required.

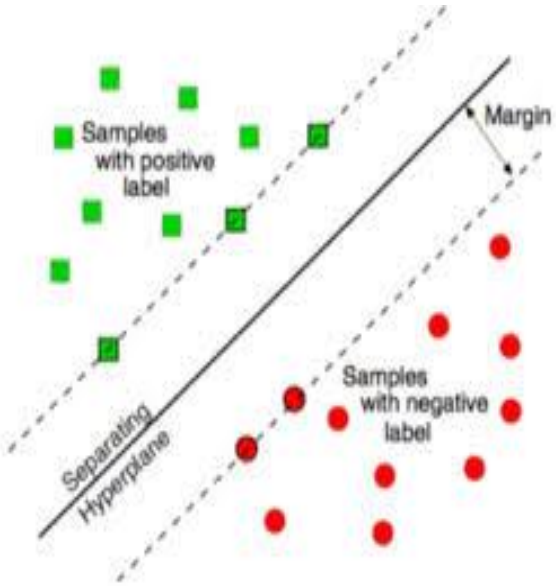

Fig.1.Supprot Vector Machine

(3) *Fuzzy Logic:* It is a system that uses fuzzy sets or fuzzy logic to convert data values into membership degrees through membership function. These rules are used for the classification. Therefore; category is defined for each one of the attributes. A Fuzzy logic classifier has very high speed and used where limited precision is accepted in the data values or real time. Fuzzy image processing processes 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 image defuzzification. 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 performance is poor when there

is a limited amount of knowledge that the designer can incorporate in the system.

(4)*Principal Component Analysis:* Sometimes, the dimension of the input vector is large with redundancy in its component values. Hence input vector size must be reduced in dimensions. The effective way to perform this operation is Principal Component Analysis. This result in: Firstly it orthogonalizes input vector components. Secondly, it orders the components such that a component with large variation is first. Finally it eliminates those components producing least variation in the data set.

(5)*Artificial Neural Network:* An artificial neural network (ANN) is an information-processing paradigm that uses the biological nervous system. An artificial neuron is like a device having many inputs and one output.
Originally, ANN started in the form of a single neuron proposed in the McCulloch and Pitts model in the 1940s. 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. The network is made up of only input neurons and output neurons. It can solve only linear problems. The two modes of operation for neuron: the training mode and the using mode. In the training mode, for particular input pattern neuron is fires. In the using mode, when a known input pattern is detected at the input, its corresponding output becomes current output.Fig.2 and Fig.3 explains the network.

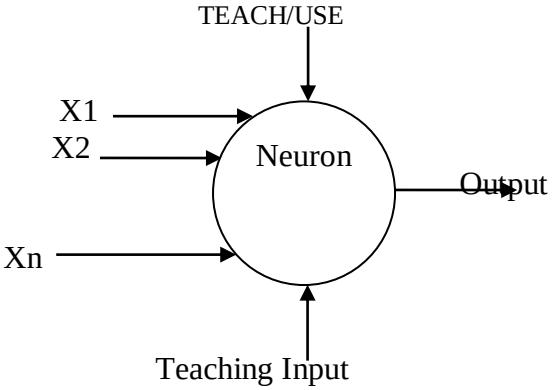

Fig.2.General Framework of Neural Network

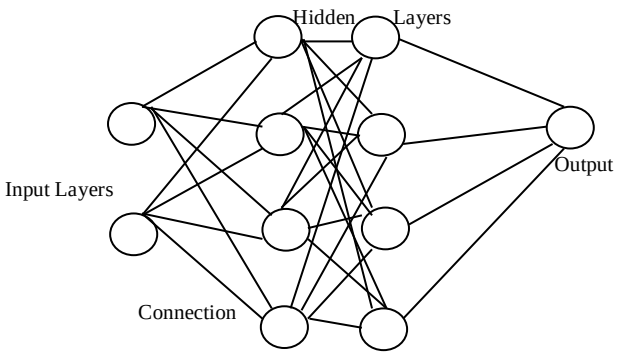

Fig.3. Internal Connections of Network

If the input pattern does not belong to list of the input patterns, the firing rule decides whether neuron is to fire or not. Neural networks are typically organized in layers. The layers are containing a number of interconnected 'nodes' that contain an 'activation function'. The input layer introduces the pattern to network, and communicates to one or more hidden layers. Hidden layers are the layers where the actual processing is done through weighted connections. The hidden layers are linked to an output layer that produces output. This can be observed in fig.3

Depending upon on the connection pattern; ANN's can be grouped into two categories. Feed-forward networks, in which graphs have no loops. Recurrent or feedback network, in which loops occur because of feedback connection. Feed-forward networks are static, i.e. they produce only one set of output values. Recurrent or feed–back networks, on the other hand are dynamic systems.

(6)*Self Organizing Map:* The Self Organizing Map was developed by professor Kohonen in 1980.This model of neural network is based on unsupervised learning. This unsupervised learning produces a low dimensional representation of input space of training samples called map.

To preserve the properties of the input space it uses neighbourhood functions. Therefore, it is useful to visualize low dimensional view of high dimensional data similar to multidimensional scaling. It operates in two modes: training and mapping. Using input training builds maps and mapping classifies new input vector. Usually nodes are 2D spacing in rectangular or hexagonal grid. To lace a vector onto map a node with closest weight vector is determined.

IV. CONCLUSION

From study of above classification techniques we come up with following conclusion. K-nearest neighbor is the simple in all algorithms for predicting the class of a test example. But major disadvantage of K-NN is the time complexity required for predictions. The neural networks are tolerant to noisy inputs. In neural network, algorithm structure is difficult to understand. SVM is competitive with 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. The drawback of SVM is determination of optimal parameters while training data is not linearly separable. Also its more complex to understand and implement.

ACKNOWLEDGEMENT

A research work of this magnitude has been a journey with various ups and downs. I extend my sincere & heartfelt obligation towards all the personages who have helped me in this endeavour. I am ineffably indebted to Prof. Dr. M.P.Satone, Associate Professor, K.K.W.I.E.E.R, for her conscientious guidance and encouragement to accomplish this work. Last, but not the least, I would like to thank Department of Electronics & Telecommunication Engineering of my

institute and all my colleagues who have been my moral support in the completion of my work.

V. REFERENCES

[1] Wenjiang Huang, Qingsong Guan, Juhua Luo, Jingcheng Zhang,Jinling Zhao, Dong Liang, Linsheng Huang, and Dongyan Zhang "New Optimized Spectral Indices for Identifying and monitoring Winter Wheat Diseases" IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 7, NO. 6, JUNE 2014

[2] Sanjeev S Sannakki, Vijay S Rajpurohit, V B Nargund, Pallavi Kulkarni"Diagnosis and Classification of Grape Leaf Diseases using Neural Networks" 4th ICCCNT 2013

[3] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiao Long "Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks" 8th International Conference on Natural Computation (ICNC 2012)

[4] Jun Pang ,Zhong-ying Bai,Jun-chen Lai,Shao -kun Li "Automatic Segmentation of Crop Leaf Spot Disease Images by Integrating Local Threshold and Seeded Region Growing" IEEE 2011

[5] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad "A Framework for Detection and Classification of Plant Leaf and Stem Diseases" IEEE 2010

[6] Yuan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo "SVM based Multuple Classifier System for Recognition of Wheat leaf Diseases"CDC-2010

[7] A.Meunkaewjinda,P.Kumsawat,K.Attakitmongcol,A.Srikaew "Grape leaf Disease Detection from Color Imagery using Hybrid Intelligent System"ECIT-CON 2008

[8] Yan cheng zang, Han- Ping Mao, Bo Hu, Ming-xi"Features Selection of Cotton Disease Leaves Image based Feature Selection Techniques"IEEE-2007

Author: Kamlapurkar Sushil R has his completed Bachelor of Engineering in Electronics and Telecommunication from Savitribai Phule Pune University(formerly University of Pune).He is currently pursuing Master of Engineering from the same.