Survey on Classification Techniques for Plant Leaf Disease Classification

K.Muthukannan*, P.Latha#, R.Pon Selvi®, P.Nisha©

*Associate Professor, Einstein college of Engineering,Tirunelveli

#Associate Professor, Government college of Engineering, Tirunelveli

©,®PG Student ,Einstein college of Engineering , Tirunelveli

***Abstract*– This paper present survey on various Classification techniques which can be used for plant leaf disease classification. Plant leaf disease classification is a technique where disease classified based on its texture features. There are various effective classification techniques like k-Nearest Neighbour Classifier, Probabilistic Neural Network, Genetic algorithm, Learning Vector Quantization etc. But Choosing on method for classification is a difficult task, because the accuracy of the result may be varied for different input data. The goal of this survey is to classify the plant leaf diseases to increase the quality and quantity of the agricultural product. And also help the farmer to take superior decision about many aspects of crop development process.**

Keywords – Leaf classification, Classifier, PNN, Genetic algorithm, KNN, LVQ

I. INTRODUCTION

Agriculture is the mother of all nations. It played a vital role in the development of human civilization. Research in agriculture domain is aimed towards increase the quality and quantity of the product at less expenditure with more profit. The quality of the agricultural product may be degraded due to plant diseases. Disease in plants that interrupts its vital functions such as photosynthesis, fertilization etc. These diseases are caused by pathogens viz.., fungi, bacteria, viruses.

Therefore, detect and classify the plant disease in early stage is an important task. Farmers require continuous monitoring of experts which might be prohibitively expensive and time consuming. Therefore looking for fast, less expensive and accurate method to automatically detect and classify the diseases from the symptoms that appear on the plant leaf. Depending on the applications, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing and some automatic classification tools.

II. REVIEW OF LITERATURE

Suhaili Beeran kutty, Noor ezan Abdullah et al., in paper titled Classification of Watermelon Leaf Diseases Using Neural Network Analysis [1] proposed the colour feature extraction to extract the RGB pixels. The proposed automated classification model involved the process of disease classification using Statistical Package for the Social Sciences (SPSS) and Neural Network Pattern recognition toolbox.

Al-Hiary, H., S. Bani-Ahmad, M. Reyalat, M. Braik, and Z.AlRahamneh. proposed Fast and accurate detection and classification of plant diseases [2] ,neural network classifier was used as a classifier. The future scope of this paper is to automatically estimate the severity of the detected disease.

P. Revathi M. Hemalatha detected Cotton leaf spot diseases in [3] by using Homogenous Segmentation based Edge Detection Techniques. This system is analysed 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 [4], Al-Bashish, D., M. Braik, and S. Bani-Ahmad developed Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification developed color image segmentation algorithms to get disease spots. The edge of the diseased spot identified by homogenize techniques like sobel filter, canny filter etc. These extracted edges are given as the input to the classification.

S.Arivazhagan, R.Newlin Shebia et al., developed Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features [5], color transformation structure for the input RGB is created. Green pixels are masked and removed using specific threshold value. For useful segments, texture statistics can be computed. Extracted features are passed through the classifier.

In [6], S.Phadikar, et al developed Classification Of Rice Leaf Diseases Based On Morphological Changes, proposed an automated system has been developed to classify the leaf brown spot and the leaf blast diseases of rice plant based on the morphological changes of the plants caused by the diseases. Radial distribution of the hue from the center to the boundary of the spot images has been used as features to classify the diseases by Bayes and SVM Classifier.

Sanjeev S Sannakki., et al., in paper titled Diagnosis and Classification of Grape Leaf Diseases Using Neural Networks [7], proposed thresholding process to mask the green pixels. Segmentation is done by k-means clustering. Feed forward back propagation neural network was trained for classification.

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 textural features like homogeneity, energy, entropy etc. Some of the classification techniques used as Neural Network, Genetic Algorithm, k-Nearest Neighbour Classifier and learning vector quantization. Plant leaf disease classification has wide application in Agriculture.

1. K – Nearest Neighbour

K-Nearest Neighbour is an example of lazy learner. The main function of the lazy learner is simply stores the training tuple, and process the test tuple based on training tuple similarity.

Nearest neighbour classifiers are based on learning by analogy, that is by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n-dimensional pattern space. When given an unknown tuple, a k-nearest neighbour classifier searches the pattern space for the k training tuple that are closest to the unknown tuple. These k training tuples are the k "nearest neighbours" of the unknown tuple.

"Closeness" is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples,

$X1 = (x_{11}, x_{12}, x_{13}, \ldots\ldots x_{1n})$

$X2 = (x_{21}, x_{22}, x_{23}, \ldots\ldots x_{2n})$

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}$$

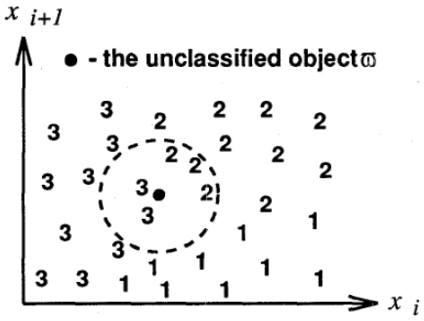

Fig 1: Example for classification using k-NN rule

Advantages

1. Simpler classifier as exclusion of any training process.
2. Applicable in case of a small dataset which is not trained.

Drawbacks

1. Speed of computing distance increases according to numbers available in training samples.
2. Expensive testing of each instance and sensitive to irrelevant inputs.

2. Probabilistic Neural Network (PNN)

A probabilistic neural network (PNN) is a feed forward neural network, which was derived from the Bayesian networkand a statistical algorithm called Kernel Fisher discriminant analysisin which the operations are organized into a multi-layered feed forward network having four layers viz. input layer, pattern layer, summation layer, and output layer.

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.

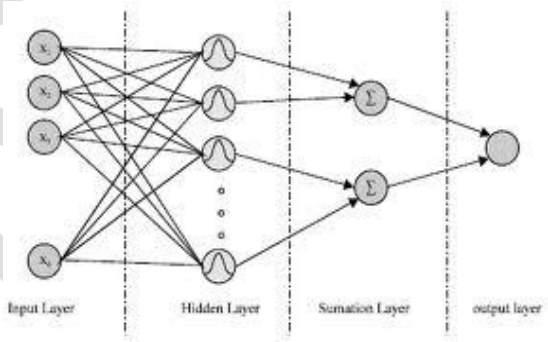

Fig 2: Multilayer Feed Forward Network

Advantages

1. PNNs are much faster than multilayer perceptron networks.
2. PNNs can be more accurate than multilayer perceptron networks.
3. PNN networks are relatively insensitive to outliers.
4. PNN networks generate accurate predicted target probability scores.
5. PNNs approach Bayes optimal classification.

Disadvantages

1. PNN are slower than multilayer perceptron networks at classifying new cases.
2. PNN require more memory space to store the model.

3.Genetic algorithm

Genetic Algorithms are mainly used for feature classification and feature selection. The basic purpose of genetic algorithms (GAs) is optimization .GAs is adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. They represent

intelligent exploitation of a random search used to solve optimization problems.

Based on natural selection, this algorithm evolves three operators

1. Selection which equates to survival of the fittest.
2. Crossover which represents mating between individuals.
3. Mutation which introduces random modification.

ALGORITHM

Begin
t = 0
Initialize population P (t)
Compute fitness P (t)
Repeat
t=t+1
SelectP (t) from P (t-1)
CrossoverP (t)
MutateP (t)
Compute fitness P (t)
Until termination criterion is achieved
End

Advantages

1. GA algorithm is less sensitive to noise.
2. Robust search technique.

4. Learning vector quantization

Learning Vector Quantization (LVQ) is a supervised version of vector quantization that can be used when we have labelled input data. This learning technique uses the class information to reposition the Coronoid vectors slightly, so as to improve the quality of the classifier decision regions. It is a two stage process – a SOM followed by LVQ: This is particularly useful for pattern classification problems.

1.The first step is feature selection – the unsupervised identification of a reasonably small set of features in which the essential information content of the input data is concentrated.

2.The second step is the classification where the feature domains are assigned to individual classes.

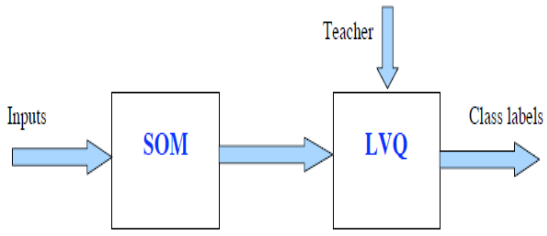

Fig 3: Block diagram for LVQ algorithm

The LVQ Algorithm

The basic LVQ algorithm is a straightforward method for shifting the Voronoi cell boundaries to result in better classification. It starts from the trained SOM with input vectors $\{\mathbf{x}\}$ and weights/Voronoi vectors $\{W_j\}$, and uses the classification labels of the inputs to find the best classification label for eachW_j. The LVQ algorithm then checks the input classes against the Voronoi cell classes and moves the W_j appropriately

1. If the input **x** and the associated Voronoi vector/weight$W_{I(X)}$(i.e. the weight of the winning output node $I(\mathbf{x})$) have the same class label, then move them closer together by $\Delta W_{I(X)}(t) = \beta(t)(\mathbf{x} - W_{I(X)}(t))$as in the SOM algorithm.

2. If the input **x** and associated Voronoi vector/weight $W_{I(X)}$have the different class labels, then move them apart by $\Delta W_{I(X)}(t) = -\beta(t)(\mathbf{x} - W_{I(X)}(t))$.

3.Voronoi vectors/weights W_j corresponding to other input regions are left unchanged with $\Delta W_j(t) = 0$.

Where $\beta(t)$ is a learning rate that decreases with the number of iterations/epochs of training. In this way we get better classification than by the SOM alone.

Advantages

1. Easy to interpret in various application domain.
2. LVQ can be applied for multi class classification problems.

Disadvantages

1. The classification generalization performance will not be as good as possible.

IV. COMPARATIVE STUDY

Table 1: Comparative Study of classification techniques for plant leaf classification

Techniques	Advantages	Disadvantages
K-Nearest Neighbour (KNN)	1.Simple classifier. 2.Applicable for small dataset which is not trained.	1.Sensitive to irrelevant inputs. 2.Based on the number of training samples, speed is increased.
Probabilistic Neural Networks (PNN)	1.Tolerant of noisy inputs	1. Long training time. 2. Need more memory for training tuple.
Genetic algorithm	1. GA algorithm is less sensitive to noise. 2. Robust search technique.	1. Computation speed is slow.

Learning Vector Quantization(LVQ)	1. LVQ can be applied for multi class classification problems. 2. Can handle data with missing values.	1. Need to be able to generate useful distance measures for all attributes. 2. Difficult to fine vectors for a given problem.

V. RESEARCH POSSIBILITIES

1. High resolution camera will be used.
2. Increase the number of training and testing samples.
3. Use other color models like CMY,HSV,HSL also.
4. Increase the number of training and testing samples.
5. Use hybrid algorithms for classification such as Genetic algorithms and NNs.
6. Automatically estimate the severity of the detected disease.
7. Shape feature, color feature can be taken as the input for disease classification.

VI. CONCLUSIONS

From study of above classification techniques we come up with following conclusion. The nearest-neighbour method is perhaps the simplest of all algorithms for predicting the class of a test example. The main disadvantage of the k-NN method is the time complexity of making predictions. Considerable amount of work has been done for recognizing plant species using k Nearest Neighbour technique.

The present paper reviews and summarizes various classification techniques for several plant species that have been used for classifying plant leafdiseases. The major techniques for classification of plant diseases are: K-NN, PNN, Genetic algorithm, Learning Vector Quantization (LVQ) techniques are used to analyses the affected plant leaves.

The k nearest-neighbour method is the simplest algorithm for predicting the class of a test example and also adaptable for small training set which is not trained. The main disadvantage of the k-NN method is the time complexity of making predictions. The most important advantage of PNN is that training is easy and instantaneous. Additionally, neural networks are tolerant to noisy inputs. But in neural network it's difficult to understand structure of algorithm and also it takes longer time for training. It needs lot of memory for training tuple. Another technique we studying is genetic algorithm. Genetic algorithms are good at refining irrelevant and noisy features selected for classification. But representation of training/output data in genetic programming is complicated. Genetic algorithms provide a comprehensive search methodology for machine learning and optimization. The learning vector quantization algorithm is a prototype based supervised learning algorithm. Classes are predefined and we have a set of labelled data. The goal is to determine a set of prototypes the best represent each class.

This review suggests that this disease classification technique shows a good potential with an ability to classify plant leaf diseases and some limitations. This application can be used in agricultural field to classify various plant leaf diseases and helps the farmer to take superior decision about many aspects of crop development process.

ACKNOWLEDGEMENT

I would like to convey my special thanks to my guide Dr.K.Ramar for his constant support for this work.

REFERENCES

[1]. Suhaili Beeran kutty, Noor ezan Abdullah et al., "Classification of Watermelon Leaf Diseases Using Neural Network Analysis" , *IEEE Business Engineering and Industrial Applications Colloquium (BEIAC), 2013.*

[2]. Al-Hiary, H., S. Bani-Ahmad, M. Reyalat, M. Braik, and Z.AlRahamneh. 2011. "Fast and accurate detection and classification of plant diseases". *International Journal of Computer Applications, 17(1): 31-38.*

[3]. *P.Revathi.,M.Hemalatha*, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques", 2012 - *International Conference on Emerging Trends in Science, Engineering and Technology*.

[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". *Information Technology Journal, 10(2): 267-275.*

[5]. S.Arivazhagan,R.Newlin Shebia et al., "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", *Agricultural Engineering Institute : CIGR journal, volume 15, No.1 ,March 2013.*

[6]. S.Phadikar, et al, "Classification of Rice Leaf Diseases Based on Morphological Changes", *International journal of Information and Electronics Engineering*, volume 2, No. 3, May 2012.

[7]. Sanjeev S Sannakki., et al., "Diagnosis and Classification Of Grape Leaf Diseases Using Neural Networks", *4th ICCCNT, July 2013.*

[8]. Robert M Haralick et al., "Textural Features for Image Classification", *IEEE transactions on Systems, Vol.SMC-3, No.6, November 1973, pp 610-621.*

[9]. Bauer, S. D., F. Korc, W. Forstner. 2011. "The potential of automatic methods of classification to identify leaf diseases from multi spectral images". *Precision Agriculture, 12: 361-377.*