Comparative Study on Classification Algorithms for Plant Leaves Disease Detection

S.Dhivya[1], Dr.R.Shanmugavadivu[2]
Research Scholar, Department of Computer Science, PSG College of Arts and Science, Coimbatore, India[1]
Assistant Professor, Department of Computer Science, PSG College of Arts and Science, Coimbatore, India[2]

Abstract

Image Segmentation is one of the important research areas in plant leaf disease detection. Many researchers proposed several methods in the field of agricultural area for segmenting the leaf diseases. The plant leaves are affected mostly by fungi, bacteria and viruses. When the one part of the leaf is attacked and the other part of the leaf is also infected. Here, the main task is the detection of various Leaf diseases in plant leaves. This paper proposed an algorithm for image segmentation which is used for detection and classification of plant leaf diseases. It also surveys various leaf diseases and classification techniques that are used for leaf disease detection. Image Segmentation is one of the important feature for detecting the diseases in plant leaf diseases is done by using K nearest neighbor, Support Vector Machine (SVM) and Probabilistic Neural Network (PNN).

Keyword: *Image segmentation, Plant Leaf Diseases, K nearest neighbor and Support Vector Machine (SVM) and Probabilistic Neural Network (PNN).*

I. INTRODUCTUION

Today agriculture plays an important role in the process of economic growth of a country. In agriculture, plant leaf segmentation plays the main concept for many researchers. Plants are vitally important for environmental protection, without that there will not be the survival of the ecology of the earth. Plant leaf classification and disease detection is the most important problem for many farmers, it also decrease the life time of the plant leaf. In most of the cases the pests or diseases are affected a large amount of the leaves. The identification of plant leaves and finding out the pests or diseases, and the percentages of pests affected in leaves and also the symptoms of the diseases play the key role in successful cultivation of crops.

Leaf disease recognition plays an important role in plant classification. The leaf disease can reduce the quality and the quantity of the crops and also reduce the future growth. Finding detection of disease in leaf manually takes too much time. So, there is the need for automatic detection of the leaf diseases. Here, computer plays the major role for detecting the diseases in leaf. There are various classification algorithms for finding the leaf diseases, use K nearest neighbor (KNN), Support Vector Machine (SVM) and Probabilistic Neural Network (PNN).

Several other approaches used their pre-defined features. Miao et al. proposed an evidence-theory-based rose classification [1] based on many features of roses. Many approaches above employ k-nearest neighbor (k-NN) classifier [2] [3] [4] while some papers adopted Artificial Neural Network (ANN). 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 [5].

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 [9], Meeta Kumar, Mrunali Kamble, Shubhada Pawar, Prajakta Patil, Neha Bonde survey various classification techniques which can be used for plant leaf classification and it deals with associating a given input pattern with one of the distinct classes.

II. LEAF DISEASES & SYMPTOMS

Leaves are mainly affected by bacteria, fungal or viral. Here, we are discussing about these various leaf diseases and the symptoms of the leaves.

Damaged Leaf	*Pests or Diseases*
Deformed leaves, Sucking damage	Aphids
Discolored leaves, sucking damage	Thrips and mites
Chewed or skeletonized leaves	Beetles, caterpillars and sawflies
Leaf galls (abnormal plant growths)	Cynipid wasps, certain aphids, psyllids and mites
Leaf mines (white patterns on leaves)	Beetle, fly, or moth larvae
Folded leaves	Caterpillars, tree cricket and spiders
Rolled leaves	Certain mites or some caterpillars
Chewed leaves, slime trails	Slugs and snails

Table 1: List of Various Damaged leaf and pests

III. PROPOSED METHOD

The different types of leaf images are taken and then those are used to recognize the affected area in leafs. Then various types of classifications techniques are applied on them, to classify those affected and unaffected leaf images. There are various steps to process the images. The following steps used for plant leaf disease detection.

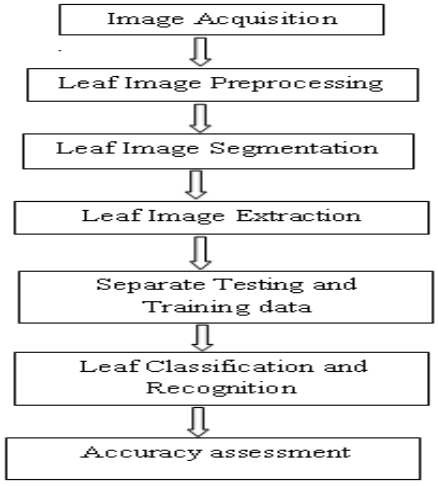


Fig 1: Block Diagram for Plant Leaf Classification

1. Image acquisition is the first step that requires capturing the leaf images with the help of a digital camera.
2. Leaf image preprocessing of input images to improve the quality of the images and remove the unwanted distortions from the image. Image clipping i.e. cropping of the leaf image to get the concerned image region. Image smoothing is done using the smoothing filter. For increasing the contrast image enhancement is also done.
3. Leaf segmentation of the input images is the process of portioning the digital image in to multiple segments i.e. pixels, boundaries.
4. After the segmentation, disease portion from the images is extracted. Mainly feature extraction involves the color, shape and pattern of images.
5. Finally, classifying images are used for training and testing of the datasets. The classifications may be k-nearest neighbor, support vector machine and probabilistic neural network based. These methods are used to classify and detect the diseased and health leaves.

IV. CLASSIFICATION ALGORITHMS

There are many classification algorithms for identifying the diseases on leaves. Here, we only focused by k-nearest neighbor, support vector machine and probabilistic neural network.

A. K nearest neighbor algorithm

K nearest neighbor algorithm (KNN) is a method for classifying objects based on closest training examples in the feature space. K-nearest neighbor algorithm is among the simplest of all machine learning algorithm. K-nearest neighbor algorithm is a method for classifying objects. Training process for this algorithm only consists of storing feature vectors and labels of the training images.

Here, the KNN uses the Euclidian distance for counting the number of infected and uninfected leaf images and also finds its accuracy. In KNN the classification i. e. to which class the given point is belongs 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. 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. The steps in KNN classification are as follows

Xi : Training data,
Yi : Class label,
xi : unknown sample
Step 1 : for i= 1 to n do
Step 2 : compute the distance d(xi, x)
Step 3 : End for
Step 4 : Compute the step 1 containing indices for the K smallest distance d (Xi, X)
Step 5 : Return majority label.
For (Yi where I belong to I)
Thus the proposed leaf disease identification is used for classifying the leaf disease based on specific type of leaves effectively and the classified dataset is utilized during the testing phase.

B. Support Vector Machine

Support Vector machine (SVM) is a non-linear Classifier and is used to identify the classes which are closely connected to the known classes. The idea behind the method is to non-linearly map the data to some high dimensional space, where the data can be linearly separated, thus providing great classification performances. The main concept of SVM first to transfer the input data into high dimensions space by means of a kernel function and then constructs an OSH (Optimal Separating Hyper Plane) between the two classes in the transformed space [17]. For plant leaf classification it will perform feature vector extracted from leaf's curve. SVM finds the OSH by maximizing the margin between the classes. Data vectors nearest to the constructed line in the transformed space are called the support vectors. The SVM estimates the function for classifying data into two classes. Using a nonlinear transformation that depends on a regularization parameter, the input vectors are placed into high dimensional feature space, where a linear separation is employed. To construct a nonlinear support vector classifier, the inner product (x,y) is replaced by kernel function K (x,y), as in (1).

$$f(x) = sgn\left(\sum_{i=1}^{l}(a_i y_i K(x_i x) + b)\right) \quad (1)$$

Where, f(x) = determines the membership of x. We assume normal subjects were labeled as -1 and other subject as +1. The SVM has two layers [4]. During the learning process, the first layer selects the basis K(x_i, x), i=1, 2…..N from the given set of kernels, while the second layer constructs a linear function in the space. This is equivalent to finding the optimal hyper plane in the corresponding feature space. The SVM algorithm can construct a variety of learning machines using different kernel functions. Fig (2) shows the linear separating hyper plane where support vector are encircled.

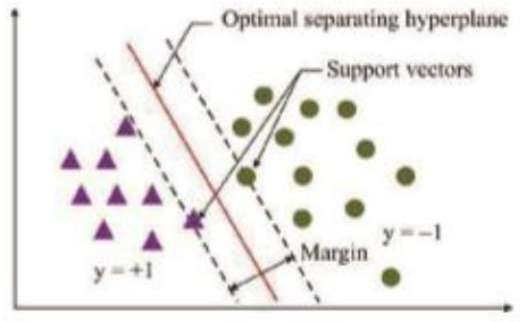


Fig 2: Linear Support Vector Machine

C. Probabilistic neural network

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a "magical" black box trained to achieve expected intelligent process, against the input and output information stream. Thus, there is no need for a specified algorithm on how to identify different plants. PNN is derived from Radial Basis Function (RBF) Network which is an ANN using RBF. RBF is a bell shape function that scales the variable nonlinearly.

PNN has many advantages. The training speed is many times faster than other classification algorithm. Here weights are not trained but it will be assigned and also existing weights are not alternated but only new vectors are inserted into weight metrics when training[7][8]. So it can be used in real-time. Since the training and running procedure can implemented by matrix manipulation, the speed of PNN is very fast. PNN has three layers: the input layer, Radial Basis Layer and the Competitive Layer. Radial Basis Layer evaluates vector distances between input vector and row weight vectors in weight matrix. These distances are scaled by Radial Basis Function nonlinearly. Then the Competitive Layer finds the shortest distance among them, and thus finds the training pattern closest to the input pattern based on their distance. In this paper, 12 leaf features are extracted and orthogonalized into 5 principal variables which consists the input vector of the PNN. The PNN is trained by 1500 leaves to classify 25 kinds of plants with accuracy greater than 80%.

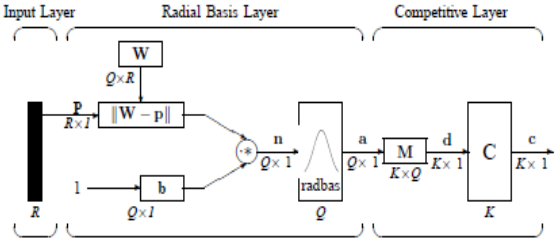


Fig 3. Network Structure R= 5, Q=1500, K=25

3.1 Input Layer

The input vector p is denoted as black vertical bar in fig 3. Its dimension is R × 1, here R=5.

3.2 Radial Basis Layer:

In this, the vector distance between input vector p and the weight vector made for each row of weight matrix W are calculated. Here, the vector distance is defined as dot product between two vectors [18]. Assume the dimension of W is Q × R. The dot product between p and the i-th row of W produces the i-th element of the distance vector ||W−p||, whose dimension is Q×1, as shown in Fig. 3. The minus symbol, "–", indicates that it is the distance between vectors.

Then, the bias vector b is combined with ||W− p|| by an element-by-element multiplication, represented as "·*" in Fig.3. The result is denoted as n = ||W− p|| · *p.

The transfer function in PNN has built into a distance criterion with respect to a center. In this paper, we define it as

$$\text{radbas}(n) = e^{-n^2} \qquad (2)$$

Each element of n is substituted into Eq. 2 and produces corresponding element of a, the output vector of Radial Basis Layer. We can represent the *i*-th element of a as

$$a_i = \text{radbas}(\| W_i - p \| .* b_i) \qquad (3)$$

Where, Wi = vector made of the i-th row of W and

bi = i-th element of bias vector b.

3.3 Competitive Layer

There is no bias in competitive layer and the vector a is firstly multiplied with layer weight matrix M, producing an output vector d. The competitive function, denoted as C in Fig. 3, produces a 1 corresponding to the largest element of d, and 0's elsewhere. The output vector of competitive function is denoted as c. The index of 1 in c is the number of plant leaf that our system can classify. In this paper, the output vector of dimension, K, is 25.

V. IMPLEMENTATION AND RESULT

In this work we have collected our own databases. We collected various plant leaf images from World Wide Web in addition to taking up some photographs of plant leaf images that can be found in and around our places. Here, we have used MATLAB as our computer language. The datasets contains 1500 images from various infected and uninfected leaf images are collected from the various plant leaves like pumpkin, pomegranate, tomato, potato, bean leaf, rose leaf and etc. In this paper we resize the images to 256 ×256 pixels. The colored images are converted into gray scale images. For Calculating Accuracy the given formula is used. This formula is used to calculate accuracy of prediction of plant leaf diseases as given as input.

$$\text{Accuracy} = \frac{Number\ of\ Right\ Prediction}{Total\ number\ of\ sample\ present} \times 100 \qquad (4)$$

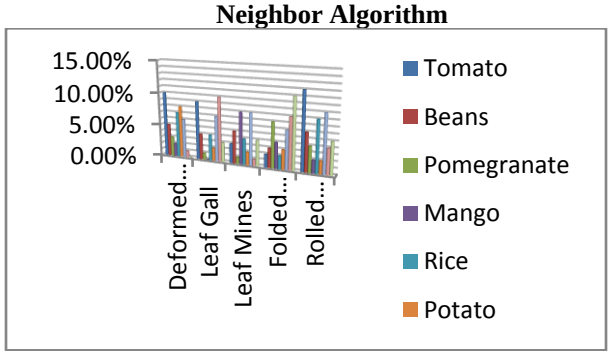
I Sample Output for Infected Plant Leaf Images

Fig 4: Sample Output of Infected plant leaf image

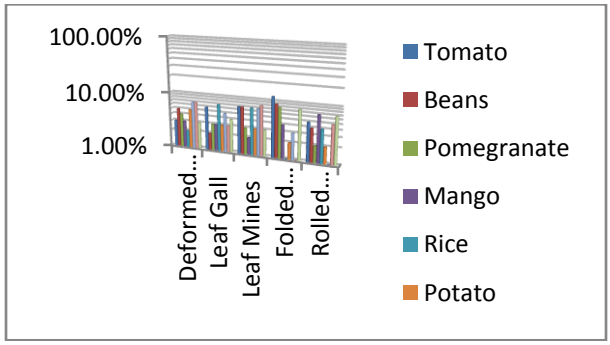
Comparison of results by KNN, SVM and PNN

Name of the plant	*Total no. of images used for training*	*Total no. of images used for testing*	*Accuracy Detection %*		
			KNN	*SVM*	*PNN*
Tomato	50	45	65.3	84	89.5
Rose	42	35	81	85	91.3
Beans	55	52	89	91.8	95.6
Pomegranate	35	35	91	89.9	95.2
Mango	30	30	93	95	98
Rice	42	42	78.3	79.5	81.25
Potato	40	40	82.4	86.4	88.9
Guava	36	36	94.5	96.7	98.2
Ladies Figurer	50	50	85.9	88.3	92.4
Sapotta	56	48	81.4	85.6	86.2
Overall Accuracy %			87.54	90.35	94

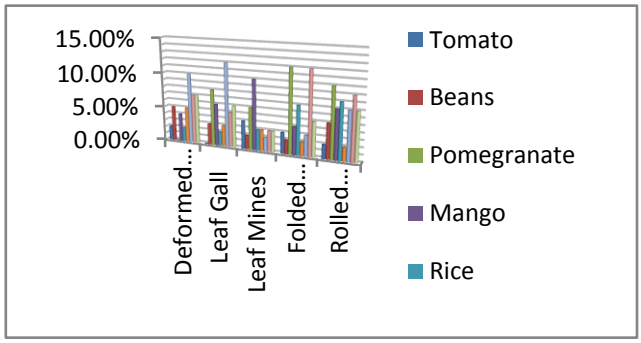
a. Sample Percentage Accuracy of K –Nearest Neighbor Algorithm



b. Sample Percentage Accuracy of Support Vector Machine



c. Sample Percentage Accuracy of Probabilistic Neural Network



VI. CONCLUSION

The accurate Disease detection and classification of the plant leaf image is very significant for the successful development of cropping and this can be done using image processing. In this paper, the methodology is used for classification of various kinds of classification algorithm can be used to classify various plant diseases working upon the resultant images. To improve the accuracy by using three classification algorithms like K-Nearest Neighbor Algorithm (KNN), Support Vector Machine (SVM) and Probabilistic Neural Network (PNN). Compare to other the Probabilistic Neural Network (PNN) classifier is able to classifying the disease in better accuracy.

REFERENCES

[1] Z. Miao, M.-H. Gandelin, and B. Yuan, "An oopr-based rose variety recognition system," Engineering Applications of Artificial Intelligence, vol. 19, 2006.

[2] J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition," Applied Mathematics and Computation, vol. 185, 2007.

[3] X. Gu, J.-X. Du, and X.-F. Wang, "Leaf recognition based on the combination of wavelet transform and gaussian interpolation," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS 3644. Springer, 2005.

[4] X.-F. Wang, J.-X. Du, and G.-J. Zhang, "Recognition of leaf images based on shape features using a hypersphere classifier," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS 3644. Springer, 2005.

[5] 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.

[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] D. F. Specht, "Probabilistic neural networks," Neural Networks, vol. 3, 1990.

[8] T. Master, Practical Neural Network Recipes. New York: John Wiley, 1993.

[9] Meeta Kumar, Mrunali Kamble, Shubhada Pawar, Prajakta Patil, Neha Bonde, "Survey on Techniques for plant Leaf Classification", International Journal of Modern Engineering Research, Volume 1, Issue 2.

[10] J.Jeya Priyankha and K.Suresh kumar, "Crop Disease Identification Using a Feature Extraction HOG Algorithm", Asian Journal of Applied Science and Technology (AJAST) Volume 1, Issue 3, Pages 35-39, April 2017

[11] S. W. Zhang* Y. J. Shang and L. Wang, " Plant Disease Recognition Based On Plant Leaf Image", The Journal of Animal & Plant Sciences, 25 (3Suppl. 1) 2015 Special Issue Page:42-45 ISSN: 1018-7081.

[12] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, Paulus Insap Santosa, "Leaf Classification Using Shape, Color, and Texture Features", International Journal of Computer Trends and Technology- July to Aug Issue 2011.

[13] Lakhvir Kaur, Dr. Vijay Laxmi, "A Review on Plant Leaf Classification and Segmentation", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 5 Issue 8 August 2016 Page No. 17658-17661.

[14] ManishaAmlekar, Dr. Ramesh R Manza, Dr. Pravin Yannawar and Dr. Ashok T Gaikwad, "Leaf Features Based Plant Classification Using Artificial Neural Network", IBMRD's Journal of Management and Research, Print ISSN: 2277-7830, Online ISSN: 2348-5922.

[15] Jayamala K. Patil , Raj Kumar, "Advances In Image Processing For Detection Of Plant Diseases", Journal of Advanced Bioinformatics Applications and Research ISSN 0976-2604 Vol 2, Issue 2, June-2011, pp 135-141.

[16] Sharada P. Mohanty, David P. Hughes and Marcel Salathe, "Using Deep Learning for Image-Based Plant Disease Detection", Frontiers in Plant Science, Volume 7, September 2016.

[17] Krishna singh, Dr.Indira Gupta and Dr. Sangeetha Gupta, Retrivel and classification of leaf shape by support vector machine using binary decision tree, probabilistic neural network and genetic fourier moment technique: a comparative study, IADIS International Conferences Computer Grapics, Visualization, Computer Vision and Image Processing 2010.