PLANT DISEASE DETECTION AND CLASSIFICATION USING IMAGE PROCESSING AND ARTIFICIAL NEURAL NETWORKS

Mr. Sanjay Mirchandani[1], Mihir Pendse[2], Prathamesh Rane[3], Ashwini Vedula[4]

[1]*Assistant Professor, Dept. of Electronics and Telecommunication Engineering, Vivekanand Education Society's Institute of Technology (VESIT), Maharashtra, India*

[2,3,4,]*Student, Dept. of Electronics and Telecommunication Engineering, Vivekanand Education Society's Institute of Technology (VESIT), Maharashtra, India*

---***---

Abstract - *Agricultural productivity is something on which Indian economy highly depends. This is one of the reasons that disease detection in plants plays an important role in agricultural field, as having disease in plants is quite natural. If proper care is not taken in this area then it causes serious effects on plants and due to which respective product quality, quantity or productivity is deteriorated. Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops, and at very early stage itself it detects the symptoms of diseases i.e. when they appear on plant leaves .Visually identifying plant diseases is inefficient, difficult, time consuming, requires expertise in plant diseases and continuous monitoring which might be expensive in large farms. Therefore; a fast, automatic and accurate method to detect plant disease is of great importance. Hence, image processing technique is employed for the detection of plant diseases. The implementation of these technologies will lead to improved productivity.*

Key Words*: *Image Processing, K-means Clustering, Artificial Neural Networks, Feed Forward Neural Networks, Cascaded Feed Neural Networks

1. INTRODUCTION

In this paper a software solution for fast, accurate and automatic detection and classification of plant diseases through Image Processing is presented [1]. Identification of the plant diseases is the key to preventing losses in the quality and quantity of the agricultural product. Health monitoring and disease detection of plant is critical for sustainable agriculture. The typical method of studying plant disease is to rely on visually observable patterns on the plant leaves. Visually identifying plant diseases is inefficient, difficult, time consuming, requires expertise in plant diseases and continuous monitoring which might be expensive in large farms. Therefore; a fast, automatic and accurate method to detect plant disease is of great importance. Hence, image processing technique is employed for the detection of plant diseases. The implementation of these technologies will lead to improved productivity India has a diverse agricultural sector. Agriculture plays a vital role in India's economy and over 58 per cent of the rural households depend on agriculture as their principal means of livelihood. Research in agriculture is aimed towards increase of productivity and quality of food. Three are two main characteristics of plant-disease detection machine-learning methods that must be achieved, they are: speed and accuracy. In this study an automatic detection and classification of leaf diseases has been introduced, this method is based on K-means as a clustering procedure and ANNs(Artificial Neural Networks) as a classifier tool using some texture feature set .The aim of this work is threefold:

1) Identifying the infected object(s) based upon K-means clustering.
2) Extracting the feature set of the infected Leaf images.
3) Detecting and classifying the type of disease using ANNs(Artificial Neural Networks).

2. STEP-BY-STEP APPROACH

2.1 FLOWCHART

The Flowchart gives a brief idea of the various stages; which include Image Acquisition, RGB to Gray conversion ,Median filtering, K-means Clustering and disease detection using Neural Networks. Feed forward back propogation and Cascaded forward back propogation are the two types of neural training networks used.We have collected about 38 images of various plant diseases which we are going to detect as our dataset. The diseases are black spot disease, yellow sigatoka disease, frog eye disease , powdery mildew disease, tobacco ringspot disease, tomato plant disease and valedensia leafspot disease.Pre-processing is done to suppress unwanted image data and to enhance some important image features. It includes RGB to Gray conversion, image resizing and median filtering. The infected part of the plant is then highlighted and made to differ from the healthy part using K-means clustering algorithm. After this various features of the diseased plant images are extracted which include : Energy, Entropy, Co-variance, Standard deviation and Background intensity average. These various feature values serve as an input to the neural network which detects the appropriate disease.

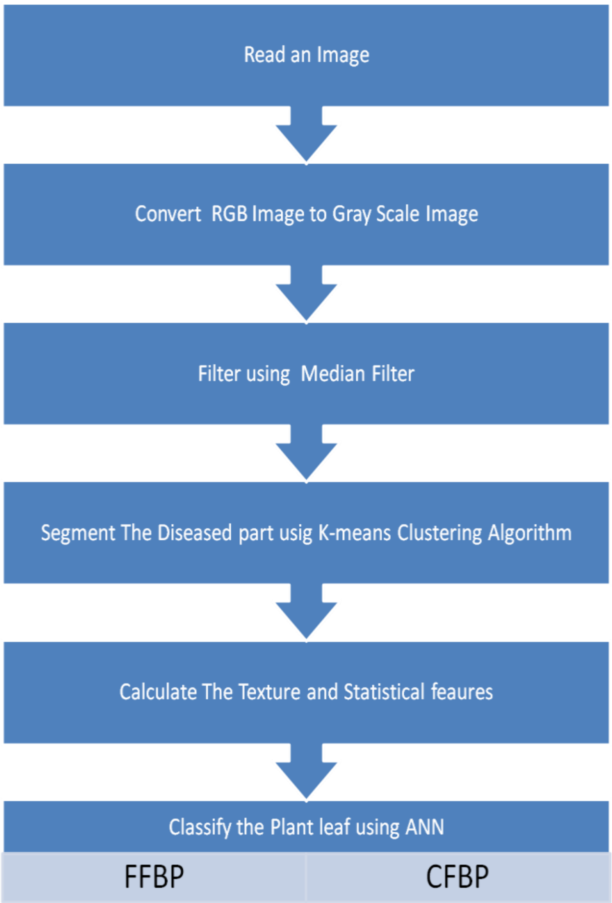

Figure -1: Process Flowchart

2.2 PREPROCESSING

The main aim of Pre-processing is to suppress unwanted image data and to enhance some important image features. It includes RGB to Gray conversion, image resizing and median filtering. Here colour image is converted to gray scale image to make the image device independent. The image is then resized to a size of 256*256. Then median filtering is performed on the image to remove the noise.

2.3 K-MEANS CLUSTERING

K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data (i.e., data without defined categories or groups)[10].The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

Algorithm :

The *K*-means clustering algorithm uses iterative refinement to produce a final result. The algorithm inputs are the number of clusters *K* and the data set. The data set is a collection of features for each data point. The algorithms starts with initial estimates for the *K* centroids, which can either be randomly generated or randomly selected from the data set. The algorithm then iterates between two steps:

1. Data assignment step: Each centroids defines one of the clusters. In this step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More formally, if c_i is the collection of centroids in set C, then each data point x is assigned to a cluster based on

$$\underset{c_i \in C}{\operatorname{argmin}}\ dist(c_i, x)^2$$

Where $dist(\cdot)$ is the standard (L_2) Euclidean distance.

2. Centroid update step: In this step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

$$c_i = \frac{1}{|S_i|}\sum_{x_i \in S_i} x_i$$

The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome.

2.4 ARTIFICIAL NEURAL NETWORKS

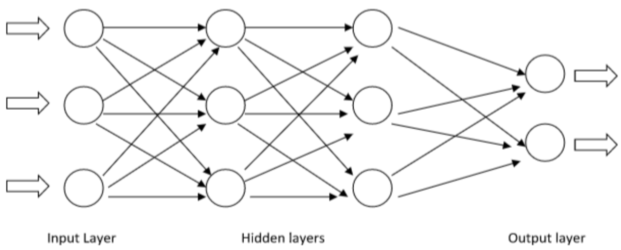

Figure-2: Artificial Neural Networks

Artificial Neural Network is a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs. ANNs(Artificial Neural Networks) are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value. Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values. Here we have used a multiple hidden layer network. This reduces load on neural network and improves its performance. The number of hidden layers taken is 2.

The two types of Neural Networks used are:

1. Feed Forward Back Propagating Neural Network:

A feed forward neural network is an artificial neural network wherein connections between the units do not form a cycle. As such, it is different from recurrent neural networks[12]. The feed forward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. Multi-layer networks use a variety of learning techniques, the most popular being back-propagation. Here, the output values are compared with the correct answer to compute the value of some predefined error-function. By various techniques, the error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. After repeating this process for a sufficiently large number of training cycles, the network will usually converge to some state where the error of the calculations is small. In this case, one would say that the network has learned a certain target function. Back Propagation network is considered to be quintessential Neural Network. Back Propagation is the training or learning algorithm rather than the network itself. To train the network we need to give the output called the Target for a particular input. The input and its corresponding target are called a Training Pair. Once the network is trained, it will provide the desired output for any of the input patterns. The network is first initialized by setting up all its weights to be small random numbers – say between –1 and +1. Next, the input pattern is applied and the output is calculated this is called the forward pass. The calculation gives an output which is completely different to what is expected (the Target), since all the weights are random. We then calculate the Error of each neuron, which is essentially: Target - Actual Output. This error is then used mathematically to change the weights in such a way that the error will get smaller. In other words, the Output of each neuron will get closer to its Target (this part is called the reverse pass). The process is repeated again and again until the error is minimal. The neurons have a sigmoid activation function. The network keeps training all the patterns repeatedly until the total error falls to some pre-determined low target value and then it stops. When the network has fully trained, the Validation Set error reaches a minimum. When the network is overtraining (becoming too accurate) the validation set error starts rising.

2. Cascaded Feed Back Propagating Neural Network:

Cascaded Forward models are similar to feed-forward networks, but include a weight connection from the input to each layer and from each layer to the successive layers. While two-layer feed forward networks can potentially learn virtually any input-output relationship, feed-forward networks with more layers might learn complex relationships more quickly. The function : newcf creates cascade-forward networks. For example, a three layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The three-layer network also has connections from the input to all three layers. The additional connections might improve the speed at which the network learns the desired relationship. CF artificial intelligence model is similar to feed forward back propagation neural network in using the back propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons .Tan-sigmoid transfer function, log - sigmoid transfer function and pure linear threshold functions were used to reach the optimized status. The performance of cascade forward back propagation and feed forward back propagation were evaluated using Root Mean Square Error (RMSE).

3. RESULTS AND DISCUSSION

3.1 RESULTS OF K-MEANS CLUSTERING

The figure below shows the original image, the gray scale representation of same. Also the median filtered image can be seen. The K-means clustered image is present which clearly separates the infected part of leaf from the healthy part.The clustering is performed on the basis of the pixel intensity values.

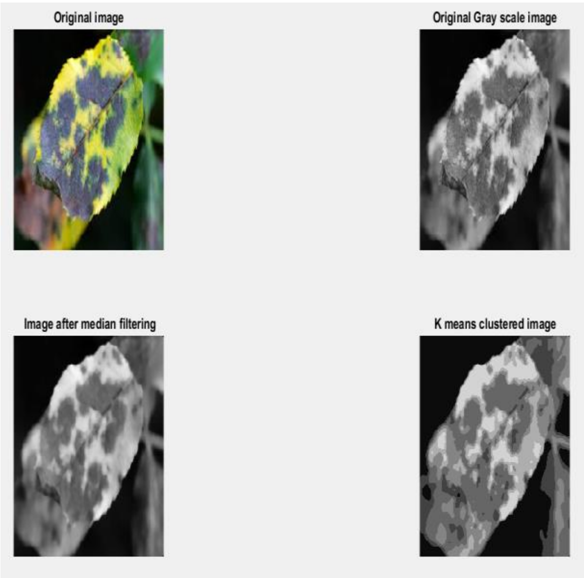

Figure-3 : Output upto K-means Clustered Image

3.2 RESULTS OF INITIAL SIX CENTRES FROM K-MEANS CLUSTERING

The initial centers obtained from the program are shown as seen in the command window. These centers form the basis of the Kmeans algorithm which is used to segment image into different regions i.e healthy an infected. The following image is a tomato leaf diseased image and below are its initial centers calculated:

Figure- 4 : Tomato leaf disease

Ini_Centre = 5 80 103 124 185 208

3.3 RESULT OF FEATURE EXTRACTION

A set of features is extracted from the segmented image which are given as a input while training of the neural network.These features include covariance,standard deviation,entropy and energy. The following is the image of Yellow Sigatoka plant disease and the values of the features are listed below.

Figure-5: Yellow Sigatoka Plant Disease

The values of the features are as follows :

- Average = 171.12
- Energy = 147
- Entropy = 0
- Standard Deviation = 36.2912
- Covariance = 1317.43

3.4 RESULT OF FEED FORWARD NEURAL NETWORK

The results of Feed Forward Neural Network for Black Spot Disease are as shown below

Figure-6: Black Spot Disease

The Feed Forward Neural Network Output is as follows

- Average = 185.32
- Energy = 128
- Entropy = 0
- Standard Deviation = 29.4282
- Covariance = 277.5350

Data of Feature Set =
1.8528
1.2800
0
0.2943
2.7754
y (Output of Feed Forward Neural Network)=

Black Spot disease

3.5 RESULT OF CASCADED FEED NEURAL NETWORK

The results of Feed Forward Neural Network for Tobacco Ring Spot Disease are as shown below:

Figure-7 : Tobacco Ring Spot Disease

- Average = 326.20
- Energy = 147
- Entropy = 0
- Standard Deviation = 29.0432
- Covariance = 119.6732

Data of Feature Set =

3.2620
1.4700
0
0.2904
1.1967
y (Output of Cascaded Feed Neural Network)=

Tobacco Ring Spot Disease

3.6 COMPARISON OF FFBP AND CFBP

FOR FFBP (FEED FORWARD BACK PROPOGATION)

	Number of iterations	Performance	Gradient
Blackspot Disease	120	0.000009	0.0084
Frog Eye Disease	100	0.00131	0.0454
Tomato Leaf Disease	200	0.00000637	0.0194
Ring Spot Disease	150	0.000098	0.0024
Valedinsia Leaf Spot	93	0.0093	0.216
Powdery Mildew	123	0.00989	0.00121
Yellow Sigatoka	110	0.09953	0.00764

FOR CFBP (CASCADED FEED BACK PROPOGATION)

	Number of iterations	Performance	Gradient
Blackspot Disease	59	0.00912	0.370
Frog Eye Disease	51	0.0941	0.492
Tomato Leaf Disease	42	0.986	1.44
Ring Spot Disease	71	0.00882	0.476
Valedinsia Leaf Spot	73	0.0992	0.512
Powdery Mildew	99	0.000897	0.0587
Yellow Sigatoka	60	0.0985	0.477

4. CONCLUSION AND FUTURE SCOPE

4.1 CONCLUSION

The applications of K-means clustering and Neural Networks (NNs) have been formulated for clustering and classification of diseases that affect plant leaves. Recognizing the disease is mainly the purpose of the introduced approach. Thus, Feed Forward and Cascaded Feed Neural Networks Algorithm was tested on seven diseases which influence the plants; they are: Black Spot, Powdery Mildew, Yellow Sigatoka, Tobacco Ringspot, Tomato Leaf Disease, Frog Eye and Valedinsia Leaf Spot. The experimental results indicate that both the approaches can significantly support an accurate detection of Leaf Diseases with little computational effort. Cascaded Feed approach is better as compared to Feed Forward approach as it requires lesser number of iterations because a smaller value of gradient facilitates quick adaption and also reduces the Mean Square Error.

4.2 FUTURE SCOPE

The Feed Forward and Cascaded Feed Neural Network algorithms can be used to design an expert system for the farmers for early detection of Plant Diseases. Presently seven Diseases as mentioned earlier can be detected by this process. The Feed Forward and Cascaded Feed algorithms can be expanded for detection of multiple diseases on a significantly large scale. These algorithms were tested on a set of 38 images which were used for Neural Training. By increasing the number of features and the number of inputs to the Neural Network the algorithms can be enhanced. If this technique is developed into a sophisticated interface in the form of a Website or Android Application it may prove to be great asset to the agricultural sector. In the future this methodology can be integrated with other yet to be developed methods for disease identification and classification. The use of other algorithms can be explored to enhance the efficiency of the system in future.

REFERENCES

[1] Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011.

[2] 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. Inform. Technol. J., 10: 267-275. DOI: 10.3923/itj.2011.267.275.

[3] https://www.britannica.com/editor/TheEdito s-of-Encyclopaedia-Britannica/4419.

[4] Biological management of Yellow Sigatoka Disease of banana caused by Mycospharella Musicola by V.Shanthiya.

[5] https://www.pioneer.com/home/site/us/agronoy/cropmanagement/soybeaninsect-disease/frogeye leaf-spot.

[6] Detection of Diseases in Different Plants Using Digital Image Processing by K.Gowthami, M.Pratyusha, B.Somasekhar.

[7] Disease Detection Using Different Algorithms by Trimi Neha Tete (Department of Electrical & Electronics) and Sushma Kamlu.

[8] Color Space Transformations Philippe Colantoni and Al 2004.

[9] The Component Median Filter for Noise Removal in Digital Images Harish 1 and M.R.Gowtham2 1Department of Electronics and Communication Engineering, Priyadarshini College of Engineering, NELLORE – 524 004, A. P., INDIA

[10] R.Nathan," Spatial Frequency Filtering," in Picture Processing and Psychopictrotics,B.S .Lipkin and A.Roswnfeld,Eds., Academic Press, New York

[11] https://www.datascience.com/blog/k-means-clustering.

[12] https://towardsdatascience.com/activation-functions-and-its-types-which-isbetter-a9a5310cc8f.

[13] The Influence of the Sigmoid Function Parameters on the Speed of Back propagation Learning Jun Han Claudio Moraga Research Group Computational Intelligence Dept. of Computer Science, University of D

[14] https://brilliant.org/wiki/feedforward-neural-networks/Performance Measures, Consistency, and Power for Artificial Neural Network Models* J. M. TWOMEY AND A. E. SMITH+ Department of Industrial Engineering, University of Pittsburgh 1031 Benedum Hall, Pittsburgh, PA 15261, U.S.A.