ESTIMATING THE SEVERITY LEVEL OF LATE BLIGH DISEASE USING ENHANCED RECOGNITION MODEL.

Zakarya A. AL-Tarawneh
IT Department, Faculty of Information Technology, Mutah university ,P.O.Box7, Mutah 61710, Karak, Jordan
ztarawneh@mutah.edu.jo

***ABSTRACT**: In this paper, shows importance of molding a reliable, precise and accurate method which estimates disease severity of Late Bligh Disease predicting yield loss using image segmentation techniques based on Artificial Neural Networks (ANNs), forecasting and monitoring epidemics, Evaluation of crop germplasm for disease resistance, and to understand fundamental biological processes including coevolution. Late Bligh Disease assessments that are inaccurate might lead to faulty conclusions being drawn from the data, which in turn can lead to wrong actions being taken in disease management decision. In the proposed approach which is a disease-independent approach, a system modeling for Late Bligh Disease will be discussed considering the necessary element that participates in the enhanced model. Therefore, experimental results will be introduced and discussed in detail. The results show that the enhanced automatic recognition image processing model based on ANN gives fast, accurate and efficient severity estimation. More than 93.5 % average accuracy is achieved. Finally, disease severity is categories by calculating the quotient of lesion area and leaf area. The enhanced model can be potentially extended to cope with various kinds of disease and taking into account not only the infected area, but also the number of infected spots and disease degree.*

Keywords: Late Blight Plant disease, Artificial Neural Networks (ANNs), Image Segmentation, K- means clustering

1. INTRODUCTION

Plant diseases can cause significant reduction in both quality and quantity of agricultural products [1]. It has been estimated that approximately $539.74 million was spent due to plant disease in Georgia (USA) in the year 2007. Consequently, Georgia state spared about 185 million USD to control the diseases, while the rest is the value of resulted damage by such diseases. To save money spent on preventing plant from diseases, a decision has to be made upon the severity level of plant diseases. This process is highly required to properly (i) identify the type of medicine and (ii) precisely evaluate the amount of medicine required for treatment to be used [2].

As a result, it is critically required of efficient, automatic, affordable and trusted method to estimate disease severity [3, 4].

Late blight [22], is a disease caused by the fungus Phytopthora infestants,a pathogen specialized for tomatos which can completely destroy the crop within 2-3 weeks of infestation. This is a air born fungus which spores infect the leaves when they land on them and as a result lesions appear on the leaves within 3-5 days. The characteristic symptoms of this disease are water-statured lesions which become black with white sporulation on the margin, frequently occurring on lower surface and occasionally on the upper surface of the leaf shows beige in color when lesions dry out in hot weather. Solutions like “disease forecasting” which enabled farmers to predict the disease when the environmental conditions were highly favorable for the spread of the pathogen where unsatisfactory. Naked eye observation of experts is also a main approach adopted in practice for its detection. Thus, farmers require the aid of experts for the detection of this disease so that the loss of crops can be prevented. This paper proposes an enhanced model software for planet leaf disease detection and analysis of late blight using image processing techniques to help forming a bridge between experts and farmers. Here, the application can take the role of the expert and help the farmer identify the disease instantly at an early stage and take control measures to protect the crop.

It is well known that, one basic requirement for the computer-based solution to be enhanced is that it should be able to estimate disease severity level for the disease symptoms that appear on plant leaves.

2. RELATED WORK

Others divided unhealthy area of plant leaves and separated their surface elements to characterize the disease[5]. Leaves from banana, guava, beans, jackfruit, lemon, mango, tomato and potato were classified for different sorts of malady utilizing bolster vector machines leading to the accuracy of 87.66%. Some others suggested a software based solution to detect and classify the plant based diseases[6]. Their work is based on color transformation structure for the input leaf, without preprocessing method to manipulate the image. The used image was exposed to color space transformation and segmentation using the K-means clustering. The extraction method of texture features was accomplished from segmented infected objects. The analysis results for late Late Blight Disease in Tomato leaves formed the input of a neural to classify the leaves.

Another method was suggested to identify nitrogen and potassium insufficiencies in tomato plants[7]. The algorithm separated a number of features from the color image. Additionally, texture features were separated using three distinct strategies, namely difference operators, Fourier transform and Wavelet packet decomposition. The choice of the elements was done by methods for a hereditary calculation. The precision of this indicative framework was over 82.5% and it could analyze malady around 6–10 days before specialists could decide. Authors in [8] introduced a technique to measure disease manifestations based on Fuzzy logic. The test images were pomegranate leaves. The calculation starts changing over the images to the L*a*b* color space. The pixels are assembled into various classes through K- means clustering. One of the gatherings will relate to the diseased regions, however the paper does not give any data on how the right gathering is distinguished. In the following, the percentage of the leaf that is estimated. Finally, a Fuzzy Inference System is utilized for the last estimation of the disease rating

The strategy proposed elsewhere tried to distinguish five different plant diseases[9]. The authors did not indicate the types of plants part in the tests. After a preprocessing stage to clean up and filtering the image, a K- means clustering algorithm is applied to divide the image into four clusters. As indicated by the author, no less than one of the clusters must relate to one of the diseases. After that, for each cluster a number of color and texture features are obtained by color co-occurrence technique, which deals with images in the HSI format. The final classification was carried out using a MLP Neural Network with ten hidden layers. Some designed a system to classify two types of disease that infect rice leaves[10]. The algorithm convert image from RGB to HSI color space then applying K- means technique to cluster the pixels into a set of groups. Those groups are then compared to a library that relates colors to the respective diseases. This comparison results in values show the probability of every district being influenced by each of the diseases.

3. ENHANCED METHODOLOGY:

The enhanced model in this paper aims to detect the severity level of Late blight disease using image processing techniques, in particular, the leaf disease detection in Al karak area, One of Jordan governorate, the estimation of disease severity degree is an image-processing-based and is composed of multiple phases. The first step is image acquisition; the next step is a pre-processing set of actions that aim at reducing the entropy of the images at hand. The next step is the implementation of a color transformation structure for the RGB leaf image. The CIELAB is used and created the *L*a*b** color space that maintains a color representation that suited to some purposes than XYZ. Next, apply the color space transformation for the color transformation structure using a device-independent. After that, apply image segmentation the K-Means clustering technique. [11,12]. Figure1 introduces the phases of the enhanced model of image processing based disease severity measurement.

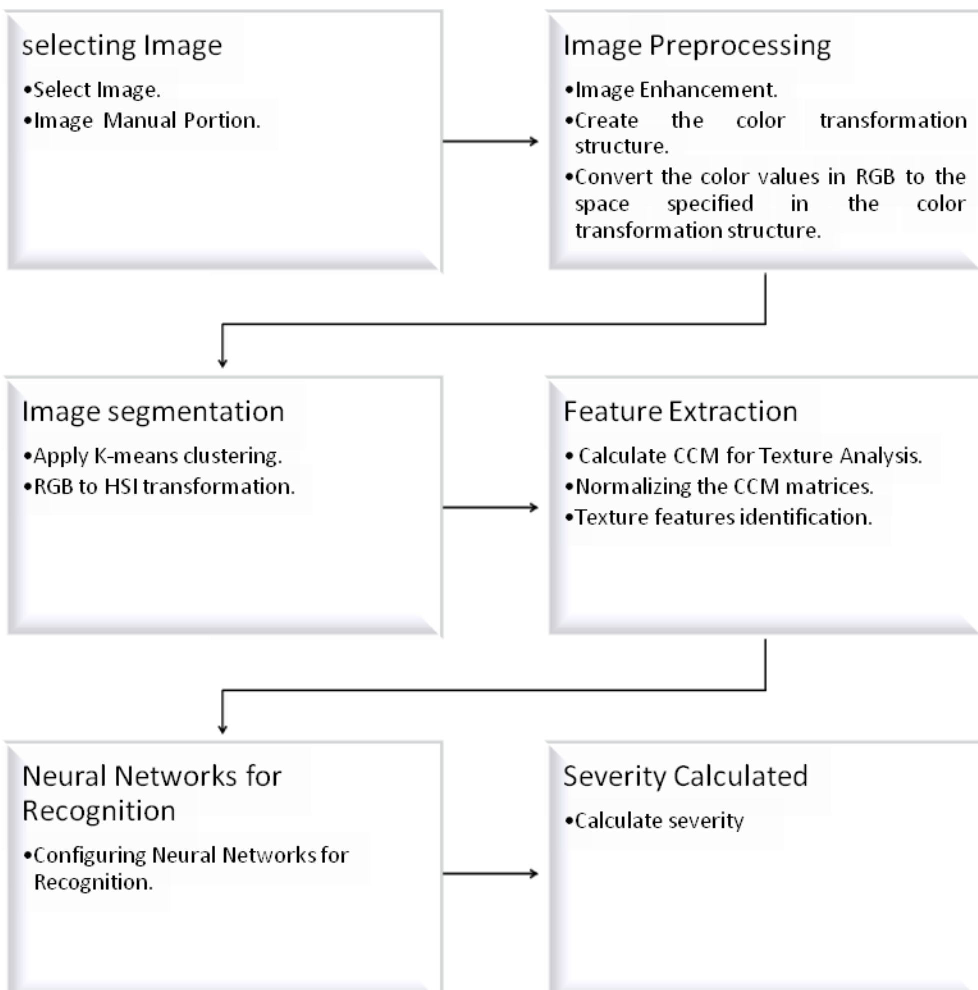

Figure 1: The flowchart enhanced model of image processing based disease severity measurement.

4. IMAGE PORTION

A selected portion of a leaf image has been taken as shown in Figure 2. The selected portion may be either infected or normal. An infected leaf image is processed as will be discussed in the following sections.

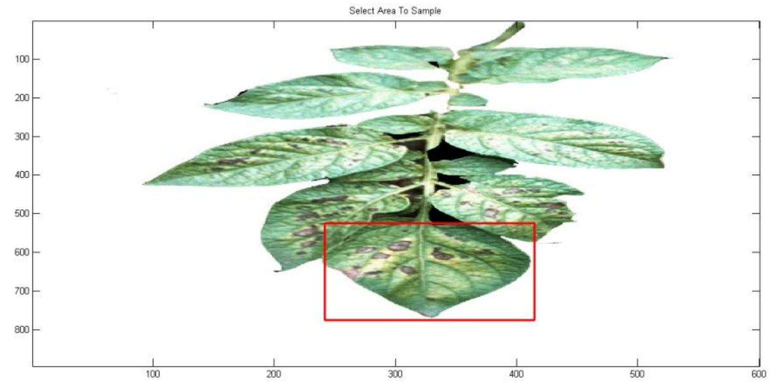

Figure 2: A sample of Image Portion.

5. IMAGE ENHANCEMENT

Image enhancement is a necessary step to improve the quality of the image. In order to make efficient segmentation, an image is enhanced by varying intensities which can be carried out by applying a Gaussian filter. In a similar manner to the mean filter, its effect is to blur the image. The probability distribution of a Gaussian filter is given in the equation below [4]:

$$G(i, j) = I_0(i, j) * e^{\left(\frac{i^2 + j^2}{2\sigma^*(i, j)}\right)} \quad (1)$$

Where (σ) is the standard deviation which determine the smoothing degree.

Once the image enhancement is accomplished, the color transformation algorithm is applied.

5.1 Color Transformation

The processing algorithm of the color transformation lies in converting the RGB image of the diseased plant or leaf, into the H, I3a and I3b color transformations. The I3a and I3b transformations are obtained by altering the original color transformation to fit the requirements of the plant disease data set [4].

5.2

K-means Clustering

Under the consideration that each leaf contains only one disease K-means clustering is used to partition the leaf image into two clusters in which one cluster is infected by a disease, while the other is an intact leaf partition. The K-means clustering algorithm tries to classify objects (pixels) based on a set of features into K number of classes. The classification is carried out by minimizing the sum of squares of distances between the objects and the corresponding cluster or class centroid [11][12]. Typically, K-means clustering is set to use squared Euclidean distances.

5.3 HSI Transformation

Hue Saturation Intensity (HSI) space is also a popular color space because it is based on human color perception [13]. Electromagnetic radiation in the range of wavelengths of about 400 to 700 nanometers is called a visible light because the human visual system is sensitive to this range. The concept of Hue referred to the wavelength of a light and intensity shows the amplitude of a light. Lastly, The “colorfulness” in HSI space is measured by a component called saturation measures [13].

6. FEATURES EXTRACTION

CCM method was used for analyzing images. The use of colored image properties in the visible light spectrum extracts additional image characteristic features over the traditional gray-scale representation [14].

The CCM methodology can be divided into three common mathematical steps. First, the pre-processed RGB image of leaf is converted into HSI color space representation. Then, each pixel map is used to generate a color co-occurrence matrix, resulting in CCM matrices, one for the H and S pixel maps [15] .

The color co-occurrence texture analysis technique was adopted through the use of Spatial Gray-level Dependence Matrices or in short SGDM's [1]. The gray level co-occurrence method is a statistical way to depict shape by statistically sampling of the way specific grey-levels occur in relation to other grey-levels. Gray level co-occurrence matrix (GLCM) created from image I by calculating how often a pixel with gray-level (grayscale intensity) value *i* occurs horizontally adjacent to a pixel with the value *j*. Each element (*i*,*j*) in GLCM specifies the horizontal occurrence of the pixel with value *i* adjacent to a pixel with value *j*.

The CCM matrices are then normalized using the equation given below, where $p(i,j,1,0)$represents the intensity co-occurrence matrix [9].

$$p(i,j) = \frac{p(i,j,1,0)}{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j,1,0)} \quad (2)$$

where N_g is the total number of intensity levels. Next is the marginal probability matrix

$$p_x(i) = \sum_{j=0}^{N_g-1} p(i,j) \quad (3)$$

The sum and difference matrices are

$$p_{x+y}(k) = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j) \quad (4)$$

where $k = I + j$; $\mathrm{k} = 0,1,2,\ldots 2(\mathrm{N_g} - 1)$, and

$$p_{x-y}(k) = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j) \quad (5)$$

where $k = |I - j|$; $\mathrm{k} = 0,1,2,\ldots 2(\mathrm{N_g} - 1)$, and $p(i,j)$ is the image attribute matrix.

In texture Features Identification the angular moment (I_1) is a measure of the image homogeneity and is defined as[15].

$$I_1 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} [p(i,j)]^2 \quad (6)$$

The mean intensity level, I_2, is a measure of image brightness and is derived from the co-occurrence matrix as follows

$$I_2 = \sum_{i=0}^{N_g-1} ip(i) \quad (7)$$

Vriation of image intensity is identified by the variance texture feature (I_3) and is computed by

$$I_3 = \sum_{i=0}^{N_g-1} (i - I_2)^2 \, p_x(i) \quad (8)$$

Correlation (I_4) is a measure of intensity linear dependence in the image and is computed by

$$I_4 = \frac{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} ijp(i,j) - I_2}{I_3} \quad (9)$$

The product moment (I_5) is analogous to the covariance of the intensity co-occurance matrix and is computed by

$$I_5 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (i - I_2)(j - I_2)\, p(i,j) \quad (10)$$

The contrast of an image can be measured by the inverse difference moment

$$I_6 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} \frac{p(i,j)}{1 + (i-j)^2} \quad (11)$$

The entropy feature (I_7) is a measure of the amount of order in an image, and is computed by

$$I_7 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j) \ln(\mathrm{i, j}) \quad (12)$$

The sum and difference entropies (I_8 and I_9) can be computed by

$$I_8 = \sum_{k=0}^{2(N_g-1)} P_{x+y}(k) \ln P_{x+y}(k) \quad (13)$$

$$I_9 = \sum_{k=0}^{2(N_g-1)} P_{x-y}(k) \ln P_{x-y}(k) \quad (14)$$

The information measures of correlation (I_{10} and I_{11}) do not exhibit any aparent physical interpretation [15].

$$I_{10} = \frac{I_7 - HXY1}{HX} \quad (15)$$

$$I_{11} = \left[1 - e^{-2(HXY - I_7)}\right]^{1/2} \quad (16)$$

where

$$HX = -\sum_{i=0}^{N_g-1} P_x(i) \ln P_x(i) \quad (17)$$

$$HXY1 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j)\ln\left[P_x(i)P_x(j)\right] \tag{18}$$

$$HXY1 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_x(i)P_x(j)\ln\left[P_x(i)P_x(j)\right] \tag{19}$$

The energy returns the sum of squared elements in the GLCM and is calculated by

$$I_{12} = \sum_{i,j} p(i,j)^2 \tag{20}$$

Homogeneity sets a value that estimates the nearness of the distribution of elements in the GLCM to the GLCM diagonal and is evaluated by

$$I_{13} = \sum_{i,j}\frac{p(i,j)}{1+|i-j|} \tag{21}$$

Autocorrelation measures the size of the primitive element and its corresponding relationship is given below [19]:

$$Autocorrelation, I_{14} = \sum_{i,j}(i,j)\,p(i,j) \tag{22}$$

7.1 Artificial Neural Networks (ANNs)

The ANN is a parallel-distributed processor whose architecture is inspired by knowledge about biological neural cells (neurons) in the brain [20]. A Back Propagation Artificial Neural Network (BP-ANN) consists of input layer, hidden layer, and output layer as shown in the Figure 3 [21]. Neural net classifier is layered feed forward network classification that incorporates supervised learning. The network learns by minimize the deference between the output node activation and the output. The aim of ANN training process is to perform a basic and most effective weight updating method of ANNs for specific computing tasks. ANNs model are good for complex systems especially when input– output patterns are in quantitative form, ANNs are considered as an approach for black-box models, which need the input and output data sets.

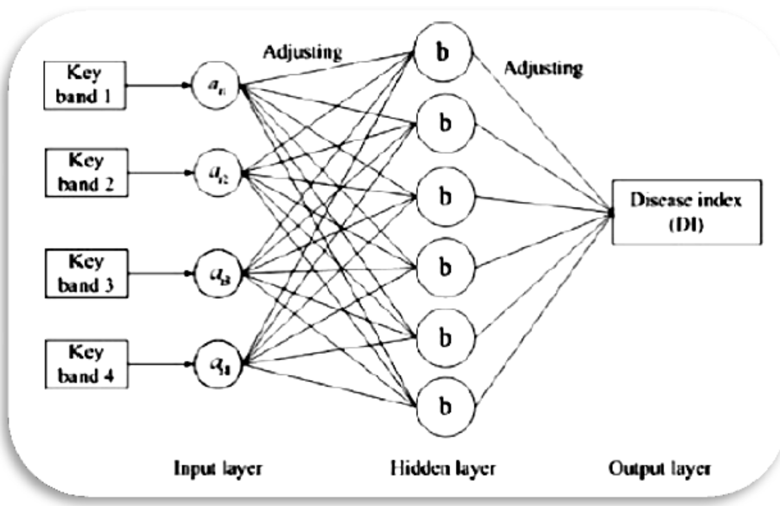

Figure 3 : A simple BP-neural network architecture.

6.1 Severity Evaluation Parameter

The disease severity can be determined according to the development of severity scale as given in 4.29 equation [18]:

Disease severity=(Infected cluster size/Total clustered area) x 100%. (23)

7. EXPERIMENTAL RESULTS:

This part discusses the implemented enhanced model for estimating the severity level of Late Blight based on ANNs. The enhanced model was developed based on MATLAB software. The model was implemented using neural networks technique for Late Blight disease. Next section presents the evaluation criteria that has been used to measure the quality of the enhanced model

8. EVALUATION CRITERIA

The Variance Accounted (VAF) is an efficient measure to evaluate the quality of the enhanced model, by comparing the actual output with the measured output of the model. VAF represents the percentage of the convergence between the value of actual output and the measured output of the enhanced model. It can be understood by simple example if one assume that y_1 and y_2 are arrays and both are equal then the VAF is 100% (Variance is zero). The VAF is computed as follows [16]:

$$VAF = \left(1 - \frac{\operatorname{var}(y_1 - y_2)}{\operatorname{var}(y_1)}\right) x100\% \tag{5.1}$$

where y_1 is the estimated (trained) array and y_2 is the measured array.

Error calculation is another relevant measure which can be used to measure error made by the neural network model. Minimum Absolute Error (MAE) is one of the most commonly used functions which are the summation of absolute difference between the estimated values and the true values divided by the number of samples *n*. That is [17].

$$MAE = \frac{\sum_{1}^{n} \left| Estimated\ value\ i\text{-}Measured\ value\ i \right|}{n} \qquad (5.2)$$

.

9. RESULTS:

The results show that the trained ANN produces a trained MAE of 5.888x10-4 and VAF of 99.83%. The tested MAE is 5.83x10-3 and VAF of 97.67%. Figure 5.17 and Figure 5.19 show that actual and estimated Late Blight plant disease for training and testing cases, respectively. The estimated error for Late Blight plant disease in both training and testing cases are demonstrated in Figures 5.18 and 5.20, respectively. The convergence of the ANN Late Blight plant disease is shown in Figure 5.21. It can be also clearly seen that the network converges when the mean square error (MSE) is approximately 10-6

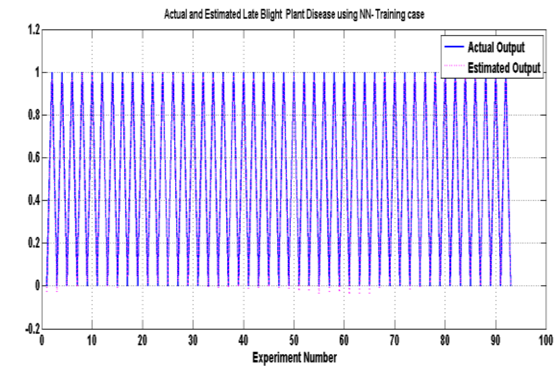

Figure 5.17: Actual and Estimated values of Late Blight Plant Disease using ANN Model / Training Case.

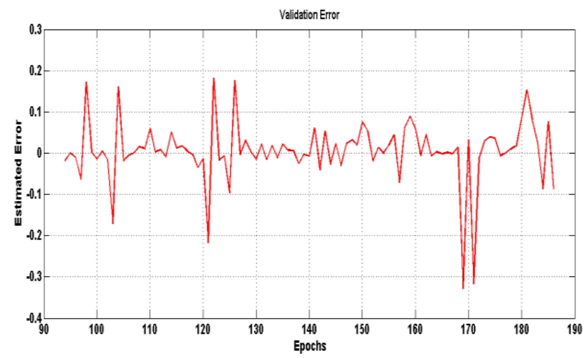

Figure 5.20: Estimated Error of the ANN Late Blight Plant Testing Case.

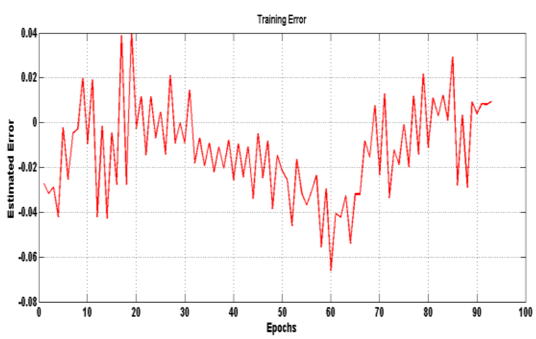

Figure 5.18: Estimated Error of the ANN Late Blight Plant Training Case.

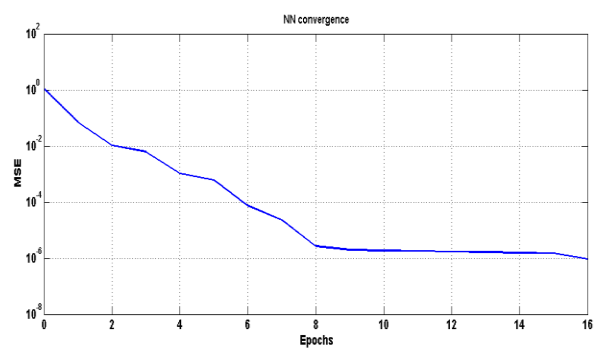

Figure 5.21: The Convergence of the ANN Late Blight Plant Case.

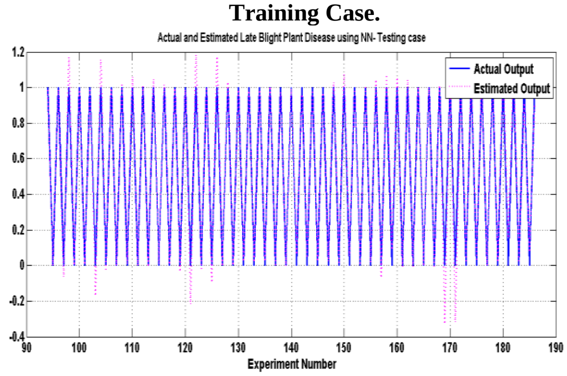

Figure 5.19: Actual and Estimated values of Late Blight Plant Disease using ANN Model / Testing Case.

10. CONCLUSION:

From the above mentioned results presented in the previous section, considering VAF as a measure of severity degree, the enhanced model showed that the minimum MAE and VAF were (5.83x10-3) and (97.67%) respectively .Thus, the ability of ANN to recognize and identify the normal and infected leaves with Late Blight Plant disease is improved by applying the enhanced model for. Therefore, it is evident that the enhanced model gives an excellent match between the actual and estimated values which, in turn, reflected on the high degree measurement of the severity degree.

11. REFERENCES:

[1] Weizheng, S., Yachun, W., Zhanliang, C. & Hongda, W. (2008) Grading Method of Leaf Spot Disease Based on Image Processing. International Conference on Computer Science and Software.

[2] Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. & Alrahamneh, Z. (2011) Fast Accurate Detection and Classification of Plant Diseases and. International Journal of Computer Applications, 17, 31-38.

[3] Al-Bashish, D., Braik, M. & Bani-Ahmad, S. (2011) Detection and Classification of Leaf Diseases Using K-Means-Based Segmentation and Neural-Networks-Based Classification International Technology Journal, 10, 267-275.

[4] Camargo, A. & Smith, J. S. (2009a) An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystem Engineering, I02, 9-21.

Camargo, A. & Smith, J. S. (2009b) Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture, 66, 121-125.

[5] S. Arivazghan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini. Detection of Unhealthy Region of Plant leaves and Classification of Plant Diseases using Texture Features. Agricultural Engineering International: CIGR Journal 2013;15(1):211-217

[6] Niket Amoda, Bharat Jadhav, Smeeta Naikwadi. “Detection and Classification of Plant Diseases using Image Processing”, International Journal of Innovative Science, Engineering and Technology 2014;1- 7 .

Engineering, 2008 WWW.MATHWORKS.COM.

[7] Guili Xu, Fenling Zhang, Syed Ghafoor Shah, Yongqiang Ye, Hanping Mao. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognition Letters 2011;32(11):1584-1590.

[8] Sannakki SS, Rajpurohit VS, Nargund VB, Kumar A. Leaf Disease Grading by Machine Vision and Fuzzy Logic. International Journal. 2011;2(5):1709-1716.

[9] Al Bashish, D., Braik, M. & Bani-Ahmad, S. (2010) A framework for detection and classification of plant leaf and stem diseases. International Conference on Signal and Image Processing (ICSIP).

Al Bashish D, Braik M, BaniAhmad S. A framework for detection and classification of plant leaf and stem diseases. International conference on signal and image processing Chennai: IEEE 2010;113–118.

[10] R. A. D. Pugoy and V. Y. Mariano, “Automated riceleaf disease detection using color image analysis,” in Proc. International Conference on Digital Image Processing. International Society for Optics and Photonics, 2011, pp. 80 090F–80 090F.

[11] Macqueen, J. (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

[12] Hartigan, J. A. & Wong, M. A. (1979) Algorithm AS 136: A K-Means Clustering Algorithm. Series C (Applied Statistics), 28, 100-118.

[13] Stone, M. C. (August 2001). “A Survey of Color for Computer Graphics”. Course at SIGGRAPH 2001.

[14] Aldrich, B. & Desai, M. (1994) Application of spatial grey level dependence methods to digitized mammograms. Proceedings of the IEEE Southwest Symposium.

[15] Pydipati, R., Burks, T. F. & Lee, W. S. (2006) Identification of citrus disease using color texture features and discriminant analysis. Computers and Electronics in Agriculture, 52, 49-59.

[16] Babuška, R., Roubos, J. A. & Verbruggen, H. B. (1998) Identification of MIMO systems by input-output TS fuzzy models. The 1998 IEEE International Conference on Fuzzy Systems Proceedings, 1998. IEEE World Congress on Computational Intelligence.

[17] Lee, L. W., Wang, L. H. & Chen, S. M. (2007) Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms. Science Direct Expert Systems with Applications 33, 539–550.

[18] Masood, A., Saeed, S., Iqbal, N., Malik, M. T. & Kazmi, A. M. R. (2010) Methodology For The Evaluation Of Symptoms Severity of Mango Sudden Death Syndrome In Pakistan. Pak. J. Bot., 42, 1289-1299.

[19] Rowshan, B. R. (2011) Texture Classication in Underwater Images. Department of Computer Architecture and Technology. Universitat de Girona.

[20] Pydipati, R. (2004) Evaluation of Classifiers for Automatic Disease Detection in Citrus Leaves Using Machine Vision. University of Florida.

[21] Zhang, H., Hu, H., Zhang, X.-B., Zhu, L.-F., Zheng, K.-F., Jin, Q.-Y. & Zeng, F.-P. (2011) Estimation of rice neck blasts severity using spectral reflectance based on BP-neural network. Acta Physiol Plant, 33, 2461–2466.

[22] Late blight on tomato.[online], 2014, http://livegpath.cals.cornell.edu/research/late-blight-of-tomato/