Classification of Watermelon Leaf Diseases Using Neural Network Analysis

Suhaili Beeran Kutty, Noor Ezan Abdullah, Dr. Hadzli Hashim, A'zraa Afhzan Ab Rahim, Aida Sulinda Kusim,Tuan Norjihan Tuan Yaakub, Puteri Nor Ashikin Megat Yunus, Mohd Fauzi Abd Rahman
Faculty of Electrical Engineering
Universiti Teknologi Mara
Shah Alam, Selangor
noor_ezan@salam.uitm.edu.my,suhaili_bk@yahoo.com

***Abstract*— This paper mainly discussed the process to classify *Anthracnose* and *Downey Mildew*, watermelon leaf diseases using neural network analysis. A few of infected leaf samples were collected and they were captured using a digital camera with specific calibration procedure under controlled environment. The classification on the watermelon's leaf diseases is based on color feature extraction from RGB color model where the RGB pixel color indices have been extracted from the identified Regions of Interest (ROI). The proposed automated classification model involved the process of diseases classification using Statistical Package for the Social Sciences (SPSS) and Neural Network Pattern Recognition Toolbox in MATLAB. Determinations in this work have shown that the type of leaf diseases achieved 75.9% of accuracy based on its RGB mean color component.**

Keywords-Watermelon Leaf Diseases; RGB; SPSS; Neural Network Pattern Recognition

I. INTRODUCTION

Malaysia is a country that rich with tropical fruits. The typical climate that is always hot and humid along the year are very suitable for the growth of these tropical plant especially watermelon. They need hot climate for their growth as well as for good production and the suitable temperature is in between 25^{o}C to 35^{o}C during day time and 18^{o}C to 28^{o}C during night time. The high temperature may decrease the matured time for watermelon and they are not suitable planted at the highland temperature below than 22^{o}C. *Citrullus Lanatus* or known as Watermelons are commonly consumed fruits from a vine-like plant of the cucurbitaceous family [1].

Since the demanding for watermelon production is increased, it needs a lot of commitment and hard work to maintain the quality and quantity of watermelon productions. There are two main factors that need to be considered before planting the watermelon, which are climate factor and the soil factor. The wrong decision in choosing the right place will burden the cultivators. The main problem that ruined the watermelon cultivators' is the diseases that attack on leaves, fruits and seeds. Some of the popular watermelon leaf diseases in Malaysia are *Anthracnose, Powdery Mildew* and Downey *Mildew.*

The main diseases that causes serious affect to the watermelon production are *Downey Mildew* and *Anthracnose*. Between these two diseases, *Downey Mildew* is dominant compare to *Anthracnose*. *Downey Mildew* is caused by fungus called *Pseudoperonospora cubensis.* The disease is most severe during periods of cool temperatures and rain. Irregular yellowish to brown spots, often vague in outline, appear on upper leaf surface near the crown. Brown spots later become more distinct on both sides of the leaves [2]. Likewise, *Anthracnose* is caused by fungus *Colletotrichum orbiculare.* The symptoms appear first on crown leaves as small, brown-black spots usually after vines begin to "run". These lesions are also visible on the underside of leaves. During damp weather, orange-pink masses of spores develop in centers of larger leaf spots. During dry weather the spore masses turn to gray [2]. These diseases will affect the sweetness, size, and freshness of the watermelon.

In this research, two types of watermelon leaf diseases which are *Downey Mildew* and *Anthracnose* have been tested in order to produce an automated model for detecting watermelon leaf diseases. With the advancement of computer software, processing and analysis of the watermelon leaf diseases can be visualized easier [3]. The analysis of RGB color indices extraction will be evaluated through SPSS and Neural Network Pattern Recognition Toolbox in MATLAB.

II. METHODOLOGY

A. Data Collection

A total of 200 leaf samples were taken from a watermelon nursery in Kuala Selangor. The leaf samples were collected carefully to make sure that the symptoms are related to the disease. The leaves samples were then put into the black plastic bags in order to preserve the freshness.

The calibration of the room lighting need to be done prior to photo session in order to make sure the consistency of lux intensity in the studio is under controlled environment. The photo session was taken placed at the Image Capturing Studio Room (ICS Room) at Advanced Signal Processing (ASP) Research Lab, Faculty of Electrical Engineering,

978-1-4673-5968-9/13/$31.00 ©2013 IEEE

UiTM.The lux reading of room light has been recorded using Heavy Duty Light Meter Data Logger (HD 450). The calibration was categorized in two intervals time, morning at 8.00am to 12pm and afternoon at 2.00pm to 6.00pm. For this research, the photo session has been conducted in the morning session. Fig. 1 below shows the samples of watermelon leaf diseases for Downey Mildew and Anthracnose.

(a) Downey Mildew

(b) Anthracnose

Figure 1. Watermelon leaf samples for each disease.

B. Data Capturing Setup

The basic digital color image in Red, Green and Blue (RGB) were obtained using Nikon D80 digital camera with image resolution of 3872 x 2592 in JPEG format. For the camera settings, the International Standards Organization (ISO) was set at 800 to gain the sensitivity of image sensor, the shutter is 50, the aperture is F8 and zoom is 35mm. For capturing the image, the camera was placed 11.5 inches above from the leaf sample at 90 degrees angle.

The lighting source was provided by spotlight which is (Digicolor K-2500) at 45 degrees angle. The room light intensity was maintained at 2331.17±170 lux. Further details on image acquisition setup are illustrated in Fig. 2.

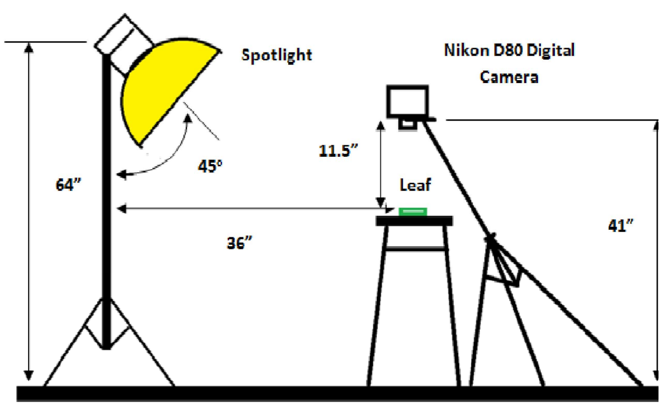

Figure 2. Image acquisition setup.

C. Image Processing

Initially, the Region of Interest (ROI) need to be identified from each infected leaf samples based on its RGB color component. Then, it is ready to crop out and resized into a dimension of 15 x 15 pixel area [4]. Later, all cropped data will be transferred to SPSS for further analysis.

The number of samples obtained from the cropping process was 200 and 186 for Downey Mildew and Anthracnose respectively.

Flowchart in Fig. 3 depicted overall classifying process of the watermelon leaf diseases identification.

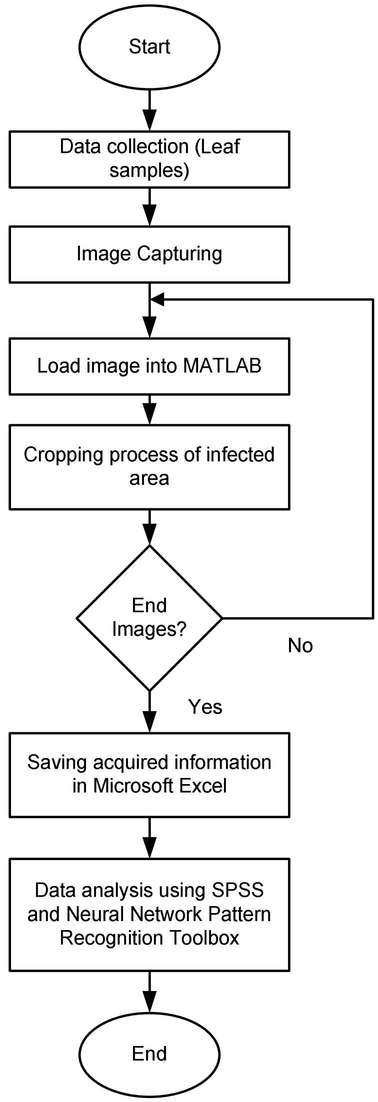

Figure 3. Flowchart of identification watermelon leaf diseases.

D. Preprocessing

Conventionally, a color image of size $M \times N$ can be defined as 3 sequences of pixel level:-

$$r(m, n), \quad 1 \le m \le M \; ; 1 \le n \le N \tag{1}$$

$$g(m, n), \quad 1 \le m \le M \; ; 1 \le n \le N \tag{2}$$

$$b(m, n), \quad 1 \le m \le M \; ; 1 \le n \le N \tag{3}$$

where each element pixel representing R, G and B component respectively [5]. The individual color components are generally represented by 8-bit which means that each element would be an integer in the interval [0,255]. In any processing of 8-bit images, the integer restriction is abandoned and the image is processed in a floating point representation in order to minimize quantization effects.

E. *Median Filtering*

Preliminary step in the process was the preprocessing of the images in order to reducing noise and facilitating image segmentation by using median filtering. The imaging technique may be noisy in terms of small white ellipse lines or dots. This artifact can be considered as impulsive noise and may thus be reduced using a median filter given by:

$$P_{med}\left(m,,n\right) = median\left\{P\left(m-k, n-1\right) \middle| \; -\frac{N_{med}-1}{2} \le k, \; l \le \frac{N_{med}-1}{2} \wedge 1 \le m-k \le m \wedge 1 \le n-1 \le N \right\} \quad (4)$$

Where N_{med} is odd[2] and indicates the size of the two dimensional median filter. P represents all the three color components and only square median filter kernel was considered [6].

F. *Inference Test*

For this project the analysis have been done by using error bar plot in SPSS software as initially inference analysis. From the test, the comparison can be made and the significant value also can be obtained [7, 10].

G. *Neural Network Pattern Recognition Toolbox*

This work develops a classification system for watermelon leaf diseases using Neural Network Pattern Recognition Toolbox (nprtool) in MATLAB. The neural network connection is consists of a two layer feed forward network and trained with Lavenberg-Marquardt backpropagation algorithm (trainlm)[11].

III. Results and Discussion

A. *Error Plot*

Error bar plot is used to show the estimated error or uncertainty in the dependent variable in standard X-Y plot [8]. The 95% confidence interval normally used correlating with the error plot to indicate the range of estimated data or error. The small black circle and square in the middle of each error plot represent the sample mean of the data, while the horizontal bar at each end of the vertical line is represent the upper (UCL) and lower (LCL) limit.

RGB error bar plots will provide better clarification if discrimination is required through observation [8].

From Fig. 4, it is clearly shows that the Red and Green color space for the two types of diseases can be discriminated from each other. However, there is no significance difference in Blue color component for both diseases.

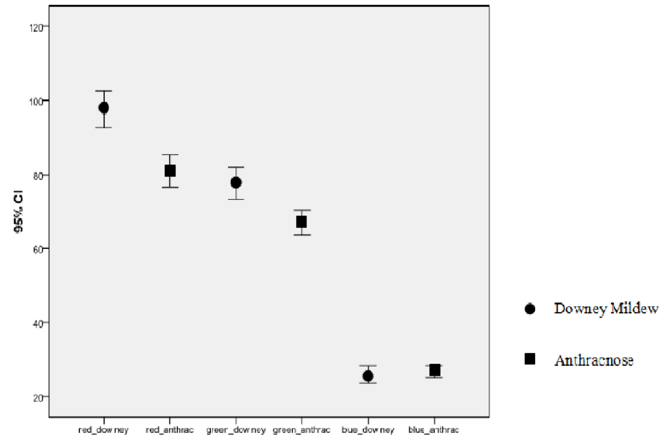

Figure 4. RGB error bar plot for mean

B. *Neural Network Pattern Recognition*

The second analysis is made by using Neural Network Pattern Recognition Tool in order to enhance the previous analysis with the best 20 neurons in its hidden layer.

The plots of confusion matrix for each training, validation, testing and overall performance in Neural Network were shown in Fig 5 until Fig 8.

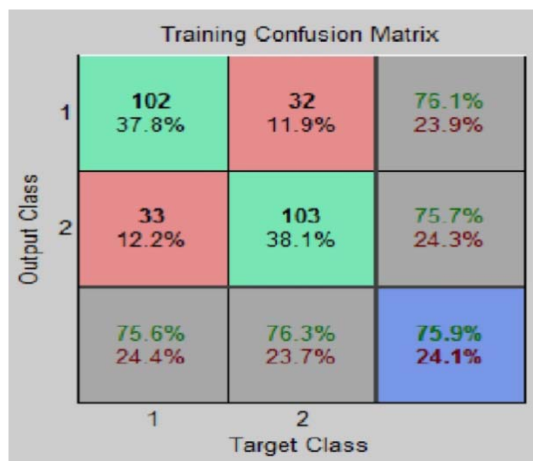

Figure 5. Confusion matrix for training data.

The training confusion matrix plot is consists of True Positive Rate (TPR) represent by Green box, True Negative Rate (TNR) represent by Red box and overall accuracy (Blue box). While row number 1 for Downey Mildew and row number 2 is Anthracnose. Based on Fig. 5 the training data is set to be 70% (270 samples). From the confusion plot it is observed that the TPR and TNR for Downey Mildew are 37.8% and 11.9% respectively while, the Anthracnose achieved TPR (12.2%) and TNR (38.1%). The overall accuracy for training is 75.9%.

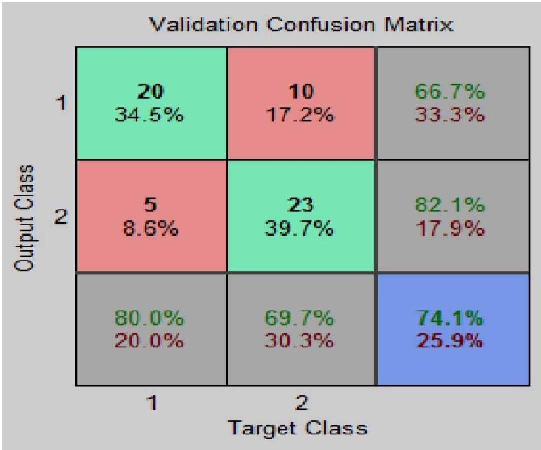

Figure 6. Confusion matrix for validation data.

Fig. 6 represents the 15% (58 samples) for the validating data. The TPR and TNR for Downey Mildew are 34.5% and 17.2% respectively. Whereas for the Anthracnose the TPR is equal to 8.6% and the TNR is 39.7%. The accuracy for the validating data is 74.1%.

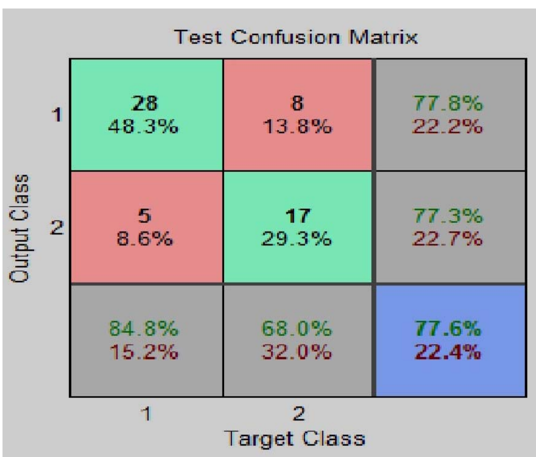

Figure 7. Confusion matrix for testing data.

Fig. 7 depicts the confusion matrix for 15% (58 samples) test data. From the plot it is observed that the TPR and TNR for the Downey Mildew are 48.3% and 13.8% respectively while the Anthracnose achieved TPR (8.6%) and TNR (29.3%). It is indicated that the accuracy of testing is 77.6%.

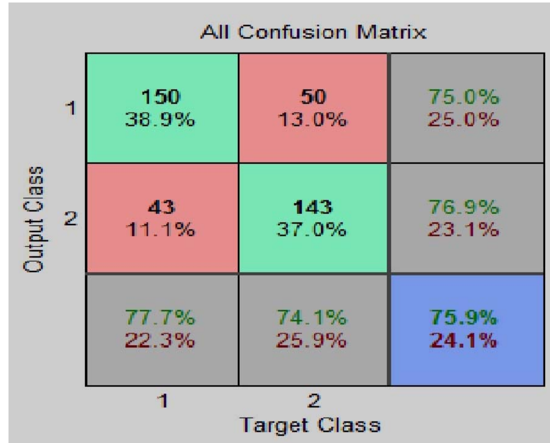

Figure 8. Confusion matrix for overall data.

According to Fig. 8, the confusion matrix it is found that the TPR and TNR for Downey Mildew are 38.9% and 13.0% respectively while, the Anthracnose carried out TPR (11.1%) and TNR (37.0%). The accuracy for the overall confusion matrix achieved 75.9% of accuracy.

Subsequent to these results, the Receiver Operating Characteristic (ROC) is plotted as shown in Fig. 9 and 10.

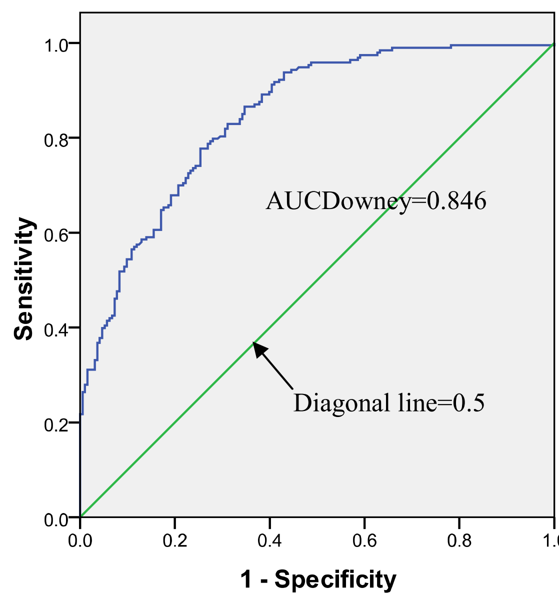

Figure 9. ROC curve for Downey Mildew data.

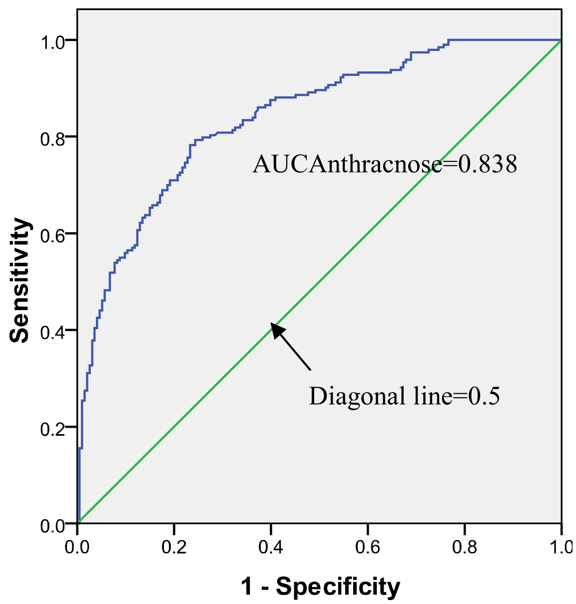

Figure 10. ROC curve for Anthracnose data

Fig. 9 and 10 depict the ROC curve for the overall performance of Neural Network model for both diseases. The performance of ROC curve is indicated by:

a) The curves leaning very closely to the upper left corner of the plot and far from the diagonal line.
b) If the Area Under Curve (AUC) is close to 1 (area of unit square) , AUC indicates that the system is a very good test [9].

From the acquire curves, it is clearly shows that both of them are far from the diagonal line (0.5). It is implies that the tests are in fair and acceptable. Based on the ROC curves, it is indicated that the AUC for Downey Mildew is 0.846 while for the Anthracnose is 0.838.The summarization of the Neural Network system is tabulated in Table I.

TABLE I. OVERALL PERFORMANCE FOR BOTH TWO DISEASES

Diseases	Downey Mildew	Anthracnose
Color Pixel Component	RGB mean	RGB mean
Accuracy of Training (%)	76.1	75.7
Accuracy Validation (%)	66.7	82.1
Accuracy of Testing (%)	77.8	77.3
Overall (%)	75.0	76.9
AUC (%)	84.6	83.8

Based on the Table I, the Neural Network system are able to achieve overall performance of 75.0% accuracy for Downey Mildew and 76.9% accuracy for Anthracnose. These results indicated that the intelligent system using Neural Network is success to classify the diseases.

IV. CONCLUSION

This paper describes the analysis of watermelon leaf diseases by using image processing technique with respect to its mean value of RGB color component. There are two types of watermelon leaf diseases used in this work named Downey Mildew and Anthracnose. Each group consists of images captured under specific requirement. The cropped data from each leaves images then were analyzed using SPSS software and then tested their performance by using Error Bar plot and Neural Network Pattern Recognition Toolbox.

As a conclusion, the overall results show that the identification for true classification of Downey Mildew and Anthracnose accuracy is the 75.9% and through Receiver Operating Characteristic (ROC) showed that they lied above the diagonal line (AUC>0.5). This implies that the results can be used as references in the future.

V. FUTURE RECOMMENDATION

In order to improve the effectiveness of this classification system for the watermelon leaf diseases, it is recommended to use a high pixel of digital camera to get the best images. Also recommended to increase the number of data for the training and testing to get the best result. In addition, the lighting setup must be in proper position because it also can affect the image captured. Besides that, the other color model can be used as the input in order to increase the efficiency such as Cyan, Magenta, Yellow (CMY), Hue Saturation Value (HSV) and Hue Saturation Lightness (HSL).

ACKNOWLEDGMENT

The author would like to express appreciation to Haji Yunus bin Saluki and Mr. Azim, owner and supervisor at the watermelon nursery, Dr. Rozieta Laboh, Pest and Disease Managemant Program at Malaysian Agricultural Research and Development Institute (MARDI) for sharing their knowledge. And not to forget to UiTM Research Management Institute (RMI) and Faculty of Electrical Engineering, UiTM for supporting this work under Excellent Fund (Research Intensive Faculty) code (81/2012).

REFERENCES

[1] Dane, F. and J. liu, *Diversity and Origin of Cultivated and Citron Type watermelon (Citrullus Lanatus).* Genetic Resources and Crop Evolution International Journal, September 2007. **54**(6): p. 1255-1265.

[2] http://aggiehorticulture.tamu.edu/vegetable/watermelon/foliar.text.html view on January 2012

[3] Abdullah, N.E., et al., *Classification of Rubber Tree Leaf Diseases using Multilayer Perceptron Neural Network*, in *The 5th Student Conference on Research and Development (SCOReD 2007).* 11-12 December 2007: Kuala Lumpur.

[4] Paliwal, J., N.S. Shashidhar, and D.S. Jayas, *Grain Identification using Kernel Signature.* Transaction of the ASAE, 1999. **42**(6): p. 1921-1924

[5] A.K.Jain, Fundamentals of Digital Image Processing: Prentice Hall, 1989.

[6] Hassan, S.L.M., et al., *A Study on Palmistry Color Reflectance Related to Personality of Subject*, in *2011 International Conference on electronic Devices, systems & Applications (ICEDSA 2011)*. 25-27 April 2011: Boulevard Hotel, Kuala Lumpur Malaysia.

[7] R. Levesque, "Programming and Data Management for SPSS Statistics 17 . 0," *Data Management*.

[8] Denan, N., *"Ultilising 3rd Level Haar Wavelet Using Algorithm on Psoriasis Lession Images*, in *Faculty of Electrical Engineering*. 2007, Universiti Teknologi MARA.

[9] Slaby, A., *ROC Analysis with Matlab*, in *Information Technology Interfaces International Conference (ITI 2007)*. 25-28 June 2007: Cavtat. p. 191-196.

[10] Abdullah, M.M., et al., *Various Grades of Red Flesh Watermelon Ripeness based on NIR and VIS Reflectance Measurement*, in *2012 IEEE Student Conference on Research and Development (SCOReD 2012)*. 4 - 5 December 2012: Batu Ferringhi, Pulau Pinang.

[11] MATLAB, *Neural Network Pattern Recognition Toolbox*. 2012