Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

AENSI Journals

Journal of Applied Science and Agriculture

ISSN 1816-9112

Journal home page: www.aensiweb.com/JASA

Early Detection of White-Root Disease for Rubber Tree Based on Leaf Discoloration with Neural Network Technique

[1]Mohd Suhaimi Sulaiman, [2]Hadzli Hashim, [2]Noor Ezan Abdullah, [2]Siti Sarah Rahmat, [1]Nurbaiti Wahid, [2]Azrie Faris Mohd Azmi, [2]Faridatul Aima Ismail

[1]Faculty of Electrical Engineering, Universiti Teknologi Mara, 23000, Dungun, Terengganu, Malaysia
[2]Faculty of Electrical Engineering, Universiti Teknologi Mara, 40450, Shah Alam, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form
8 July 2014
Accepted 10 August May 2014
Available online 30 August 2014

Keywords:
Artificial Neural Network
Spectrometer Experimental data (percentage of leaf discoloration)
SPSS Software

ABSTRACT

White-root disease is the most serious disease of all rubber diseases in Malaysia. This disease is caused by a fungus called Rigidoporus Lignosus. It will attack young rubber tree and can spread out faster through transferring by root-to-root of the nearest tree. This disease is not only can damage the tree, but it also involves heavy expenditure for its treatment. Normally, the symptom of this disease can be seen through the drastic discoloration of the rubber leaf. Once discoloration occurs to the leaf, it's already too late to give treatment. The only treatment is to cut off the infected root. Recognizing this disease is done by expert people who are well trained to recognize the various symptoms. Therefore outcome of this project can help in doing early detection of white-root disease for rubber tree without any supervising by an expert person. This project was about the implementation of Artificial Neural Network (ANN) techniques in order to classify a decision of early detection of rubber tree white-root disease based on leaf discoloration. An ANN model was developed using quantitative optical measurements of leaf samples using spectrometer where the green colour of the leaf was the reference wavelength. There were three selected regions of interest to be measured: main vein, vein and petiolule. After tested with statistical analysis using SPSS software, only one selected region was recommended to be used as input in order to develop the ANN model for classification decision. The model was trained iteratively so that the output would categorize the input optical measurement into three cases: healthy, medium or worst. The performance of the optimized model was decided after observing the Receiver Operating Curve (ROC) plot analysis. The overall percentage accuracy for the optimized model is found to be 84.4% while the sensitivity for healthy, medium and worst cases are 83.3%, 73.3% and 96.7% respectively.

To Cite This Article: Mohd Suhaimi Sulaiman, Hadzli Hashim, Noor Ezan Abdullah, Siti Sarah Rahmat, Nurbaiti Wahid, Azrie Faris Mohd Azmi, Faridatul Aima Ismail, Early Detection of White-Root Disease for Rubber Tree Based on Leaf Discoloration with Neural Network Technique. ***J. Appl. Sci. & Agric.**, 9(11): 341-351, 2014*

INTRODUCTION

Rubber has been known to the indigenous people of the Americas long before the arrival of European explorers (Mohamad Johari, Mohamad Hassan, 2005). Rubber is one of the most important products to come out of the rainforest. Though indigenous rainforest dwellers of South America have been using rubber for generations, it was not until 1839 that rubber had its first practical application in the industrial world (Wade Davis, 1996). Natural rubber tree in Malaysia began in 1877 at Kuala Kangsar, Perak with the Havea Brasilliensis species.

Most of the disease problems of rubber in Malaysia are caused by indigenous fungal parasites. There are several types of diseases of rubber tree for example leaf spot, bird's eye spot, root disease and etc. There are six types of root diseases which are white, red, brown, ustulina, poria and stinking root diseases (Rao, B. Sripathi Hilton R. N., 1975).

This paper focus on the investigation of white-root disease because it is the most serious disease of all rubber diseases in Malaysia. This disease is caused by the fungus called *Rigidoporus Lignosus* (Rao, B. Sripathi Hilton R. N., 1975). White-root disease attacked young rubber trees and can spread out faster by touching root-to-root. This disease not only can damage the tree, but it involves money when it comes to control. One must rely on symptoms appearing on the roots themselves in order to recognize the disease. Figure 1.1 below shows the differential of discoloration that occurs in rubber leafs.

Corresponding Author: Mohd Suhaimi Sulaiman. Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000, Dungun, Terengganu, Malaysia.
Tel: +60173662952 suhaimisulaiman_2003@yahoo.com

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

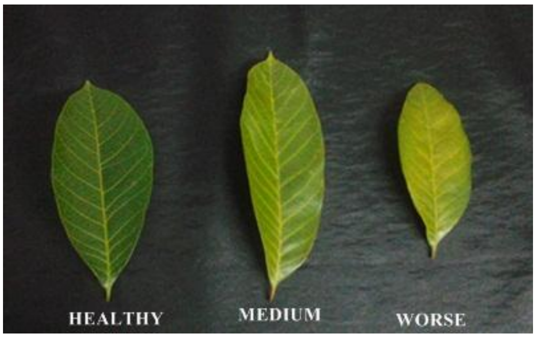

Fig. (1.1): The discoloration of white-root disease on rubber leafs.

The visible light spectrum is the section of the electromagnetic radiation spectrum that is visible to the human eye. The various color ranges indicated in Table 1.1 are an approximation: the spectrum is continuous, with no clear boundaries between one color and the next (Bruno, T.J., & Svoronos, P.D.N., 2005). The wavelength of green color (495-570nm) will be used as the color reference for this project.

This visual spectrum was used in order to see through the sample experiments and it can be related to the electrical instrument to be used. This wavelength will be used by the spectrometer in order to collect data in terms of percentage when the reflected light is sent back to the spectrometer (Hashim, H., 2010).

Table 1.1: Range of various colors

Color	Frequency	Wavelength
violet	668–789 THz	380–450 nm
blue	631–668 THz	450–475 nm
cyan	606–630 THz	476–495 nm
green	526–606 THz	495–570 nm
yellow	508–526 THz	570–590 nm
orange	484–508 THz	590–620 nm
red	400–484 THz	620–750 nm

The instrument that was used to measure the percentage of green colour in rubber leafs for this project is ZEISS Spectrometer. Figure 1.2 shows the ZEISS spectrometer which is an instrument operated based on application optical sensor system. The instrument is used to measure properties of light over a specific portion of the electromagnetic spectrum, typically used in spectroscopic analysis to identify materials. Usually the variable measured is the reflectance of light's intensity. The independent variable is usually the wavelength of the light directly proportional to the reflectance. The instrument also interfaces with ASPECT PLUS software to extract spectrum after the process was done by the instrument.

Fig. 1.2: ZEISS Spectrometer.

Artificial neural network (ANN) was selected as the Artificial Intelligence (AI) that was used to develop a model for this project. ANN is inspired by the early models of sensory processing by the brain (Jain, Anil K, Mao, Jianchang, & Mohiuddin, KM., 1996). It can be created by simulating a network of model neurons in a computer (Valentin Pupezescu, Felicia Ionescu, 2008). By applying algorithms that mimic the processes of real neurons, it can solve a variety of problems in pattern recognition, prediction, optimization, associative memory, and control (Jain, Anil K, Mao, Jianchang, & Mohiuddin, KM., 1996).

Methodology:

Figure 2.1 shows the flowchart of the overall methods that have been applied throughout the project.

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

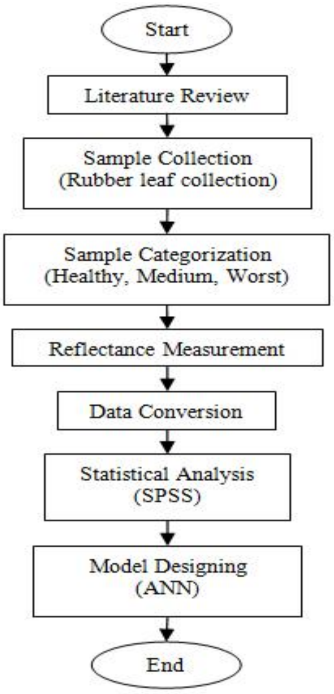

Fig. 2.1: flowchart of Overall Methods.

From figure 2.1, it shows that the development of this project only included algorithm design. It started with the literature review of the project with the main objective and ended with designing the AI model (ANN) after being done with statistical analysis using SPSS software of the experimental data.

A. *Sample Collection:*

The rubber leaf samples used in this project were collected from the nursery at Rubber Research Institute Malaysia (RRIM), Sg. Buloh. Three different stages of white-root disease were selected: healthy, medium and worst. The selection was done by an expert worker from the (RRIM). A database containing 25 inputs representing green spectrum for 6 selected rubber leaf clones, which is PB350, PB260, RRIM928, RRIM2004, RRIM2024 and RRIM 2025 was built. For this project, 100 leaf samples were taken for each clone and plucking was done in the morning in order to maintain the colour level of the leaf before it got exposed too much to the sunlight. The samples were sealed in plastic bags with appropriate labels and brought to the Advanced Signal Processing (ASP) laboratory at the Faculty of Electrical Engineering UiTM Shah Alam. Figure 2.2 shows samples of rubber leaf collected representing the three cases.

Fig. 2.2: Rubber leaf samples for each case; healthy, medium and worst.

B. *Reflectance Measurement:*

The maximum reflectance percentage of green colour from rubber leaf samples were measured using ZEISS Spectrometer. The measurements involved three selected regions of interest (ROI); main vein, vein and petiolule. Figure 2.3 shows the ROI of a rubber leaf where measurements were taken in this project. For each rubber leaf, three measurements were collected for each ROI representing each case of healthy, medium and worst. Thus, the total experimental data being measured are 5400 for all the 6 identified clones. The scanning process was done at the ASP Lab. The scanned light reflectance information was directly extracted using Aspect Plus software. All these data were then transferred to Microsoft Office Excel.

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

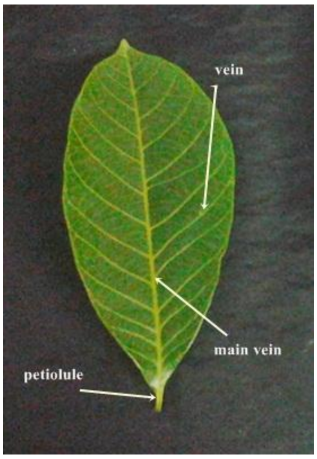

Fig. 2.3: ROI of rubber leaf.

C. *Statistical Analysis:*

Data extracted from the spectrometer were then analyzed using statistical method. For this project, SPSS software was utilized as the statistical measurement tool in order to analyze the optical reflectance of rubber leaf discoloration. The data were analyzed using three major tests which are the normality test, error plot and independent-T test (for multiple comparisons).

Normality test was used to determine whether a data set is well-modeled by a normal distribution or not, or to compute how likely an underlying random variable is to be normally distributed (Ralph B. D'Agostino, 1986). If the data set is normal or likely to be normal distributed, then further parametric tests are used for conclusive findings. In this project, error bar plots and independent-T test or multiple comparisons were recommended for observing any population discrimination.

D. *Model Designing:*

The project outcome is to develop an ANN model that can differentiate between healthy, medium and worst case of white root disease that has infected a rubber tree. For the ANN model, multilayered perceptron (MLP) network with one hidden layer was recommended in diagnosing these cases based on the fact that it has been widely applied by many projectors (Adam Lund, Mark Lund, 2013). In addition, the network that suits this diagnosing data is Lavenberg Marquardt (LM) network with one hidden layer.

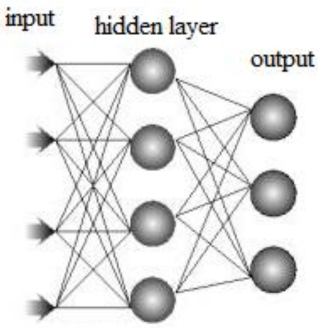

Fig. 2.4: A typical architecture feedforward of simple neural network for 3 outputs.

Optimization of the trained models was decided using a confusion matrix (Kohavi and Provost, 1998). Confusion matrix is a matrix for a two-class classifier, contains information about actual and predicted classifications done by a classification system (D.S.V.G.K.Kaladhar et. al., 2010). Table 2.1 describes the confusion matrix for a two cases classifier (Valentin Pupezescu, Felicia Ionescu, 2008).

Table 2.1: confusion table.

		ACTUAL CLASS	
		D+	D-
PREDICTED CLASS	T+	TP	FP
	T-	FN	TN

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

The entries in the confusion matrix have the following meanings:

- *D*+ (diagnosis positive) - the correct predictions
- *D*- (diagnosis negative) - the incorrect predictions
- *T*+(diagnosis positive) - the actual correct number
- *T*- (diagnosis negative) - the actual incorrect number
- *TN* (true negative) - the number of correct predictions that an instance is negative
- *FP* (false positive) - the number of incorrect predictions that an instance is positive
- *FN* (false negative) - the number of incorrect of predictions that an instance negative
- *TP* (true negative) - the number of correct predictions that an instance is positive

Sensitivity and specificity are commonly used terms that generally describe the accuracy of a test. Sensitivity is a measure of the ratio or percentage of true positive (*TP*) and a positive diagnosis test (*D*+). It is also known as true positive rate (*TPR*) which is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$FPR = \frac{TP}{TP + FN}$$

Specificity is a measure of the ratio or percentage of true negative (*TN*) and a negative diagnosis test (*D*-). It is also called as true negative rate (*TNR*) which is the proportion of negative cases that were correctly identified, as calculated using the equation:

$$FNR = \frac{FN}{TP + FN}$$

Percentage accuracy is calculated from the ratio of *TPR* plus *TNR* with the total of all samples. The equation is shown below.

$$Accuracy(\%) = \frac{TPR + TNR}{TP + FN + TN + FP} \; x \; 100$$

Performances of all the trained ANN models were analyzed by observing the Receiver Operating Curve (ROC) plot. The best threshold level for each plot was selected by calculating the minimum Euclidean Distance (ED) from the ideal point (0,1) of the ROC plot (Krogh, Anders., 2008). In addition, the best optimized model could be decided from the respective calculated area under the curve (AUC). Table 2.2 shows the set of training and testing data that were used in this project. The output of the model is the classification of healthy, medium and worst case and were defined as 100, 010 and 001 respectively.

Table 2.2: proposed set of training and testing data for each classes and ANN output.

CLASSES	TRAINING	TESTING	OUTPUT
HEALTHY	70	30	100
MEDIUM	70	30	010
WORST	70	30	001

RESULTS AND DISCUSSIONS

A. *Reflectance Measurement:*

The optical light reflectance measurements of the leaf samples were directly extracted using Aspect Plus software in terms of a displayed graph. One of the graphs from the experiment is shown in Figure 3.1 below.

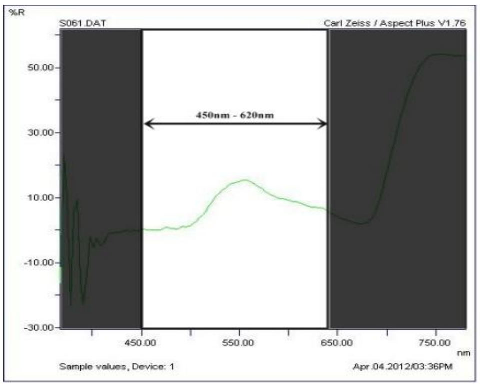

Fig. 3.1: Example of Aspect Plus graph.

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

Figure 3.2 shows an example when comparing reflectance of leaf discoloration spectrum for the three cases; healthy, medium and worst for RRIM2004 clone type. In this project, only 450nm-570nm wavelength ranges would be used for statistical analysis since the range actually represents green colouration. From the plot, it is observed that each case has different signature form of reflectance spectrum.

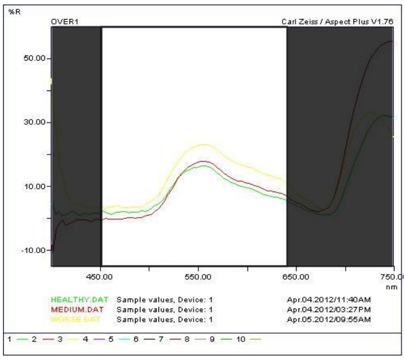

Fig. 3.2: Comparison of spectrum reflectance of leaf discoloration for three classes.

A. *SPSS analysis:*

An analysis on the data has been done by using the SPSS software. The first step before continuing with the error bar plot is the analysis of normality test. Table 3.1 shows the normality test results for each ROI with respect to each case for all clones. It is observed that only the ROI groups that belong to vein and main vein are normally distributed for every healthy, medium and worst data set used. The rule for normal distribution is if the Sig. p-value for both Kolmogorov-Smirnov and Shapiro-Wilk tests is greater the 0.05, then the data set implies having normal distribution. Otherwise, if it is below than 0.05, then the data significantly deviate from a normal distribution form (Ralph B. D'Agostino, 1986).

Table 3.1: Normality test of groups.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	condition	Statistic	df	Sig.	Statistic	df	Sig.
vein	healthy	.060	100	.200*	.982	100	.196
	medium	.048	100	.200*	.991	100	.733
	worse	.053	100	.200*	.993	100	.915
mian_vein	healthy	.052	100	.200*	.988	100	.479
	medium	.058	100	.200*	.982	100	.187
	worse	.062	100	.200*	.985	100	.309
peteolute	healthy	.073	100	.200*	.971	100	.028
	medium	.085	100	.070	.971	100	.025
	worse	.089	100	.049	.980	100	.125

Figure 3.3 depicted the error plots representing the three ROI with respect to each of the three cases for RRIM 2004. It also shows that only vein and main vein groups show significant discrimination. As a result, this test was able to prove that only rubber leafs from RRIM 2004 clones show the highest significance among all the clones and were selected to be the reference of this model. This fact would be reinforced by the numerical independent T-test result in the next paragraph.

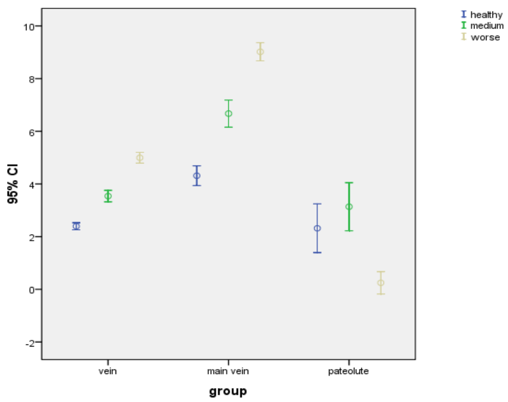

Fig. 3.3: Error bar plot.

Table 3.2 shows the comparison of discoloration of rubber leaf between cases in terms of mean of maximum percentage of reflectance and significant values. The result shows that only the mean significance of vein and main vein are below than 0.5 which means that the evidence is proven to assume that only the vein and main vein are significantly different between each other. From the overall observations below, it is shown that main vein has the highest mean difference among both condition (I) and (J). Therefore as a result, the main vein was chosen as the significant region to distinguish the various cases of rubber leaf disease for the next step in designing ANN model.

Table 3.2: Multiple comparisons of independent T-test

Multiple Comparisons

LSD

Dependent Variable	(I) condition	(J) condition	Mean Difference (I-J)	Std. Error	Sig.
vein	healthy	medium	-1.14686*	.13450	.000
		worse	-2.60165*	.13450	.000
	medium	healthy	1.14686*	.13450	.000
		worse	-1.45479*	.13450	.000
	worse	healthy	2.60165*	.13450	.000
		medium	1.45479*	.13450	.000
mian_vein	healthy	medium	-2.35775*	.29869	.000
		worse	-4.70536*	.29869	.000
	medium	healthy	2.35775*	.29869	.000
		worse	-2.34761*	.29869	.000
	worse	healthy	4.70536*	.29869	.000
		medium	2.34761*	.29869	.000
peteolute	healthy	medium	-.81922	.56263	.146
		worse	2.07073*	.56263	.000
	medium	healthy	.81922	.56263	.146
		worse	2.88995*	.56263	.000
	worse	healthy	-2.07073*	.56263	.000
		medium	-2.88995*	.56263	.000

B. *ANN Model Designing:*

In the designing of the ANN model, the focus of the training is for the model to differentiate healthy, medium and worst cases from the given training set. The trained model would then be validated using the testing set for performance analysis as being described in Table 2.2. The distribution of healthy, medium and worst used in the training and testing set were 70:70:70 and 30:30:30 respectively. 10 models were designed with different number of neurons in each hidden layer which are; 3,7,15,18,21,25,32,36,44 and 50. The input has 25 nodes representing the range of green wavelength spectrum.

Figure 3.4 describes the sum square error (SSE) performance for each designed model versus 100 epochs. All the models started to produce converging error during training phase beginning 2 to 6 epoch cycles.

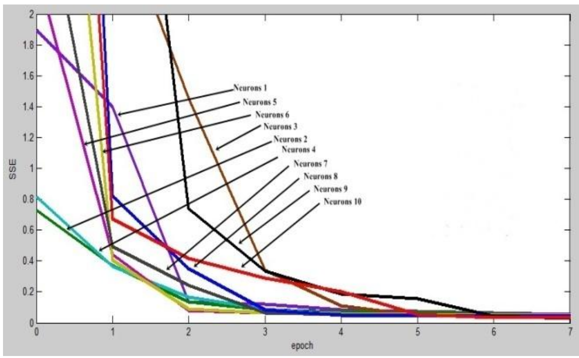

Fig. 3.4: ANN Model showing the performance of 10 hidden layer sizes with respect to SSE performance.

Figure 3.5 shows the percentage of accuracy performance calculated from the available confusion matrix for each model. It is observed that only models with 3 and 44 neurons in its hidden layer have the highest percentage of accuracy (84.4%). Therefore, only these two models are shortlisted for the final selection of the best optimized model.

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

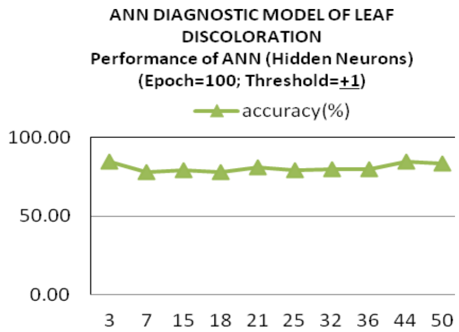

Fig. 3.5: ANN model showing the 3 neurons selected because they have the highest accuracy (84.4%).

The equation below shows the total number of connections used when designing the ANN network model. Thus, the above models (3 and 44 neurons in the hidden layer) would then produce 106 and 1172 number of connections as shown below.

$$\text{No.of connection} = \text{input} + (\text{input} + \text{hidden size}) + (\text{hidden size} + \text{output})$$

$$\text{No.of connection}_3 = 25 + (25 \times 3) + (3 + 3) \quad = 106$$

$$\text{No.of connection}_{44} = 25 + (25 \times 44) + (44 + 3) \quad = 1172$$

From the results above, it is then decided that the optimized model with hidden layer of 3 neurons would be selected for further detailed analysis. Table 3.3 shows its confusion table representing the performance of true positive rates (TPR) for healthy, medium and worst (83.3%, 73.3% and 96.7% respectively). It is proven by the equation (7), (8) and (9) below:

$$\text{TPR}_H = \frac{25}{25+5} \times 100\% = 83.3\%$$

$$\text{TPR}_M = \frac{22}{22+8} \times 100\% = 73.3\%$$

$$\text{TPR}_W = \frac{29}{29+1} \times 100\% = 96.7\%$$

The green boxes or cell (1,1), cell (2,2) and cell (3,3) of the table represents the result of true positive (TP) test applied for each case with 25 for healthy, 22 for medium, 29 for worst. Meanwhile the blue color box or cell (4, 4) represents the overall percentage accuracy of 84.4%.

Table 3.3: Confusion matrix for optimized ANN model with hidden layer of 3 neurons.

Output Class \ Target Class	1	2	3	
1	25 27.8%	8 8.9%	1 1.1%	73.5% 26.5%
2	4 4.4%	22 24.4%	0 0.0%	84.6% 15.4%
3	1 1.1%	0 0.0%	29 32.2%	96.7% 3.3%
	83.3% 16.7%	73.3% 26.7%	96.7% 3.3%	**84.4%** **15.6%**

Table 3.4 below shows the confusion matrix if only 2 cases are analysed which are healthy and non-healthy.

Table 3.4: Confusion matrix for healthy.

		ACTUAL CASE	
		D+	*D-*
PREDICTED CASE	*T+*	*25* *(TP)*	*9* *(FP)*
	T-	*5* *(FN)*	*51* *(TN)*

From the table, the *TP*, *FP*, *FN* and *TN* rate are calculated as:

Journal of Applied Science and Agriculture, 9(11) Special 2014, Pages: 341-351

$$\mathrm{TP_H} = \frac{25}{25+5} = 0.883$$
$$\mathrm{FP_H} = \frac{9}{9+51} = 0.15$$
$$\mathrm{FN_H} = \frac{5}{25+5} = 0.167$$
$$\mathrm{TN_H} = \frac{51}{51+9} = 0.85$$

Meanwhile, Table 3.5 below shows the confusion matrix if only medium and non-medium cases are analysed.

Table 3.5: Confusion matrix for medium and non-medium.

		ACTUAL CASE	
		D+	*D-*
PREDICTED CASE	*T+*	*22* *(TP)*	*4* *(FP)*
	T-	*8* *(FN)*	*56* *(TN)*

From the table, the *TP*, *FP*, *FN* and *TN* rate are calculated as:

$$\mathrm{TP_M} = \frac{22}{22+8} = 0.733$$
$$\mathrm{FP_M} = \frac{4}{4+56} = 0.067$$
$$\mathrm{FN_M} = \frac{8}{8+22} = 0.267$$
$$\mathrm{TN_M} = \frac{56}{56+4} = 0.933$$

Table 3.6 below shows the confusion matrix if only worst and non-worst case is being analyzed.

Table 3.6: Confusion matrix for worst and non-worst.

		ACTUAL CASE	
		D+	*D-*
PREDICTED CASE	*T+*	*29* *(TP)*	*1* *(FP)*
	T-	*1* *(FN)*	*59* *(TN)*

From the table, the *TP*, *FP*, *FN* and *TN* rate are calculated as:

$$\mathrm{TP_W} = \frac{29}{29+1} = 0.967$$
$$\mathrm{FP_W} = \frac{1}{51+9} = 0.167$$
$$\mathrm{FN_W} = \frac{1}{1+29} = 0.033$$
$$\mathrm{TN_W} = \frac{59}{59+1} = 0.983$$

Additional information can be deduced from the comparison of the available stages which are healthy, medium and worst case using ROC plot as described in Figure 3.6. ROC analysis is an established method of measuring diagnostic performance in various domains (Xiao Hu, Neil Eklund, Kai Goebel., 2006). This figure can show the best threshold point recommended for each case where the point with the minimum Euclidean Distance (ED) for healthy is 0.4 while the ED point for medium is 0.33 and worst is 0.19. From the overall observation of ROC graph below, it is proven that worst case has the best sensitivity and specificity than others. Classifiers with ROC curves located in the upper-left corner in ROC space are better in sensitivity and specificity because they represent classifiers that have lower false positive rate and higher true positive rate than the classifiers below them (Xiao Hu, Neil Eklund, Kai Goebel., 2006).

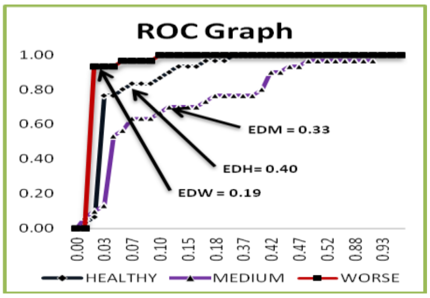

Fig. 3.6: the ROC graph of neurons 3.

The total area under curve (AUC) with respect to each case is calculated as:

$$AUC_{healthy} = 83\ \%$$

$$AUC_{medium} = 81.17\ \%$$

$$AUC_{worse} = 93.83\ \%$$

It is a priority that medium case should be selected for this project. Medium is the stage where the infection of white-root disease for rubber tree is at its early stage. Therefore, the specification for threshold selection is observed at 0.33. The sensitivity and specificity is 73.3% and 93.3% respectively.

Conclusion:

In this project, an experimental investigation into the discoloration of rubber leaf has been presented. Six types of rubber leaf clones RRIM2004, RRIM2025, RRIM 928, RRIM 2024, PB350 and PB 260 were tested. From the overall observation of statistical analysis done by SPSS software, the major tests (normality test, error bar plot and independent-T test) are able to prove that rubber leaf from RRIM2004 can be selected as the reference for this project. It is proven by the selected region of this clone is the most significant among the selected region from other clones. Result from the confusion table proves that models with neurons 3 and 44 have the highest accuracy, 84.4% and hidden layer with number of neurons 3 were selected because of the number of connection factor. True Positive Rate (TPR) for each case are 83.3% for healthy, 73.3% for medium and 96.7% for worst. Medium case is given the priority for this project since infection of white-root disease for rubber tree is at its early stage at medium. Therefore, the specification threshold for this project is 0.30. The sensitivity and specificity are 73.3% and 93.3% respectively. Thus, ANN was successfully applied to classify the cases of healthy, medium and worst using the percentage of optical leaf discoloration on rubber leafs based on green wavelength range.

ACKNOWLEDGEMENT

I am thankful to my supervisor; Prof. Madya Dr. Hadzli bin Hashim, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding and achieving the goals of this project. I would like to thank Kementerian Pengajian Tinggi (KPT) and Research Management Institute (RMI) UiTM for giving the opportunity and support by granting RAGS grant [600-RMI/RAGS 5/3 (33/2013)] in this research. I also want to thank to all RRIM staff especially the Director of Rubber Research Institute of Malaysia (RRIM) at Sg. Buloh , Dr. Mohd Nasaruddin Bin Mohd Aris and also the Crop Improvement and Protection Unit staff such as Ms. Nurmi Rohayu Bt Abdul Hamid, Ms. Azimah Bt. Izhar, Mr. Soni Bin Othman and Mr. Segar A/L Meenah @ Raman for their valuable information, suggestion, guidance and also for providing the data prior to the experiment.

REFERENCES

Adam Lund, Mark Lund, 2013. *Testing for Normality using SPSS by Laerd Statistics*. Retrieved from the Testing for Normality using SPSS website: https://statistics.laerd.com/aboutus.php

Bruno, T.J., P.D.N. Svoronos, 2005. *CRC Handbook of Fundamental Spectroscopic Correlation Charts*: Taylor & Francis.

D'Agostino, R.B., 1986. *Goodness-of-Fit-Techniques*: Taylor & Francis.

Davis, W., 2010. *One River*: Simon & Schuster.

D.S.V.G.K.Kaladhar, 2010. *Confusion Matrix Analysis for Evaluation of Speech on Parkinson Disease Using Weka and Matlab*. International Journal of Engineering Science and Technology, 2(7): 2734-2737.

Hashim, H., F.N. Osman, S.A.M. Al Junid, M.A. Haron, H.M. Salleh, 2010, 27-29 Jan. 2010. *An Intelligent Classification Model for Rubber Seed Clones Based on Shape Features through Imaging Techniques.* Paper presented at the Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference, pp: 25-31.

Jain, K. Anil, Mao, Jianchang, KM. Mohiuddin, 1996Artificial neural networks: A tutorial. *Computer, 29*(3): 31-44.

Kohavi, Ron, Provost, Foster, 1998. Glossary of terms. *Machine Learning, 30*(2-3): 271-274.

Krogh, Anders, 2008. What are artificial neural networks? *Nature biotechnology, 26*(2): 195-197. DOI: 10.1038/nbt1386

Mohamad Johari Mohamad Hassan, 2005. Economics of rubber forest plantation. Malaysian Rubber Board Monograph.

Rao, B., R.N. Sripathi Hilton, 1975. *Maladies of hevea in Malaysia*. Kuala Lumpur: Rubber Research Institute of Malaysia.

Valentin Pupezescu, Felicia Ionescu, 2008. *Advances In Knowledge Discovery In Databases*. Journal of Applied Economic Sciences, 3(4): 6.

Xiao Hu, Neil Eklund, Kai Goebel, 2006. *Using Rank Permutation For Aircraft Engine Fault Diagnostics*. Proceedings of the 60th Meeting of the Society For Machinery Failure Prevention Technology (MFPT), pp: 291-300, 6167, 396-403.