Comparison of ANN Performance Towards Agarwood Oil Compounds Pre-processing Based on Principal Component Analysis (PCA) and Stepwise Regression Selection Method

Noratikah Zawani Mahabob, Aqib Fawwaz Mohd Amidon, Zakiah Mohd Yusoff, Nurlaila Ismail, Saiful Nizam Tajuddin, NorAzah Mohd Ali, and Mohd Nasir Taib

Abstract—**This paper presents the performance of Artificial Neural Network (ANN) application towards the agarwood oil quality classification. The works involved the uses of agarwood oil compounds based on two different feature selection techniques. The compounds were are selected based on using Principal Component Analysis (PCA) and Stepwise Regression. The compounds identified by PCA (three compounds) were β-agarofuran, α-agarofuran, and 10-epi-ϒ-eudesmol while the compounds identified by stepwise regression (four compounds) were β-agarofuran, ϒ-Eudesmol, Longifolol, and Eudesmol. These compounds were fed into ANN separately as input features and the output was the quality of the oil either high and low. The Resilient Backpropagation as classifier algorithm was used and 1 to 10 hidden neuron in the hidden layer were varied. The performance of ANN using three and four compounds was measured and compared using confusion matrix, mean square error (mse) value and number of epoch. The work was done using software application, Matlab R2017a by using 'patternet' network. The finding showed that the ANN using four compounds of agarwood oil as input feature obtained greater performance with good accuracy, lower mse value and lower number of epoch in one hidden neuron.**

Index Terms*— *agarwood oil, artificial neural network, stepwise regression, resilient backpropagation, confusion matrix

I. Introduction

THERE is a precious tree in the world comes from the species Aquilaria [1]–[4]. This species are mostly found at Asia especially South Asia such as Thailand, Indonesia and Malaysia [3], [4]. The very valuable and pricey heartwood can be found in this species and it is known as agarwood or aloeswood [5].

This manuscript is submitted on 19th April 2021 and accepted on 3th August 2021. Noratikah Zawawi Mahabob, Aqib Fawwaz Mohd Amidon and Nurlaila Ismail are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor whereas Zakiah Mohd Yusoff is with campus of Pasir Gudang.

Saiful Nizam Tajuddin is with Bioaromatic Research Centre of Excellence, Universiti Malaysia Pahang. Nor Azah Mohd Ali is with Forest Research Institute Malaysia and Mohd Nasir Taib from MITRANS, Universiti Teknologi MARA, 40450 Shah Alam, Selangor (e-mail: zakiah9018@uitm.edu.my)

In Malaysia, agarwood is dark resinous heartwood which comes from the main genus Aquilaria malaccensis. In order to produce resin content from agarwood, it is undergo the process of inducement either by injury, disturbance from insects, and microbial invasions [2], [5]–[7].

Agarwood is demanding due to its special applications. The agarwood is widely used as perfume, fragrances and in soap manufactures due to its unique scent. In Middle East, the agarwood is used in wedding ceremony while in India produces an oil known as "minyak attar" for muslim purposes [6], [8]. Furthermore, in Malaysia, some ethnics used agarwood oil as insect repellent, treat liniment or other body pain [6], [9]. The agarwood oil is traded according to its grading classification either high quality or low quality. High quality agarwood oil has dark color and long lasting aroma as well as the much pricey compared to low quality [3], [4]. There is a Malaysian researcher stated that the price for grade A agarwood oil can reach up to RM20,000 per kg [5].

Grading the agarwood oil into its quality need to be emphasized. Traditionally, grading agarwood oil is based on its color, odor and infection level. Unfortunately, this method is not recommended due to inefficient and cause fatigueless to human [8], [10], [11]. In recent years, the grading is invented using the chemical properties and it is proven can enhance the accuracy and produce an accurate results on the quality of agarwood oil [3], [4], [12]. Intelligent technique such as Multilayer Perceptron (MLP), Support Vector Machine (SVM), Self Organizing Map (SOM) and k-Nearest Neighbour (k-NN) has been used in grading agarwood oil by most researchers [12]–[16].

doi.org/10.24191/jeesr.v19i1.007

The famously used algorithm in ANN is backpropagation (BP) that performs using a simple gradient descent method with sigmoid functions in the hidden layers has contributed to some problems [17], [18]. The small magnitude of the gradient leads to a small change of weight and biased and taking a longer time in the learning process are some of the issues in BP. As a supervised learning algorithm, Resilient Back-propagation is being proposed to abolish the harmful effects of the magnitude of the partial derivatives [17]–[20]. It is also known as the local adaptive learning scheme that functions in performing a direct adaptation of the weight step based on local gradient information [21], [22]. During updating weight in hidden layers, the only parameter that required is a sign of derivative which is used to determine the direction of weight update, while the magnitude of the derivative gives no effect on weight update [17], [19].

This study purposes the use of Artificial Neural Network (ANN) to perform the grading of agarwood oil based on the dataset of chemical compounds of agarwood oil by two different feature selection techniques. The significant agarwood oil compounds are taken from the previous study based on Principal Component Analysis (PCA) and the proposed selection technique using Stepwise Regression. The compounds are fed into ANN as input features and the output is the quality of the oil. The ANN using input from PCA and stepwise regression is compare and contrast using the performance criteria such as confusion matrix, accuracy, sensitivity, specificity, precision, mean square error (mse) value and number of epoch. Finally, the ANN that obtained the best performance is chosen as it is based on which pre-processing technique either from PCA or stepwise regression can effects the performance of ANN.

II. Methodology

At first, the methodology starts with the data acquisition which was the process of obtaining the chemical compounds of agarwood oil. The agarwood oil compounds were obtained from the Forest Research Centre Institute Malaysia (FRIM) and Universiti Malaysia Pahang (UMP). The process of obtaining the data had been conducted by previous researcher based on GCMS analysis and Z-score technique [23].

A. Compounds identified from Principle Component Analysis (PCA) [24].

A researcher has been done a study using the Principle Component Analysis (PCA) technique towards agarwood oil. The detail of the study can be found at paper in reference [24]. The chemical compounds of agarwood oil were reduced into three compounds from 106 compounds after the PCA technique implemented. The compounds found were β-agarofuran, α-agarofuran and 10-epi-ϒ-eudesmol. These compounds have been stated to be important compound for high quality agarwood oil [24].

B. Compounds identified from Stepwise Regression.

Seven significant compounds of agarwood oil obtained from the previous researcher has been fed into stepwise regression technique [23]. The compounds were β-agarofuran, α-agarofuran, 10-epi-ϒ-eudesmol, ϒ-Eudesmol, Longifolol, Hexadecanol and Eudesmol and they were assigned as compound 1 (C1) to compound 7 (C7) sequentially. The stepwise regression was the proposed method for this study. The significant compounds were then identified by forward selection and backward elimination technique of the stepwise regression using p-value analysis. The summarize of the experiment is explained in the flowchart in Figure 1. The feature selection of stepwise regression has identified four significant compounds which were β-agarofuran, ϒ-Eudesmol, Longifolol and Eudesmol.

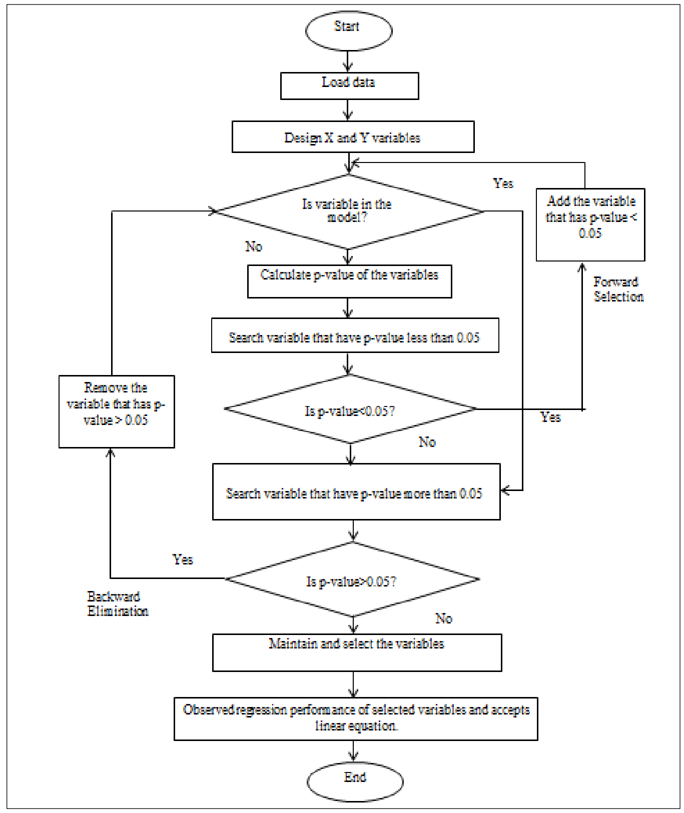

Figure 1 Detail experiment for stepwise regression

C. Classification using Artificial Neural Network (ANN)

At this part, the data of agarwood oil compounds obtained from the PCA and stepwise regression technique were fed into ANN as the input feature for classification process. The output feature was the grade of the agarwood oil either low or high quality. The data was pre-processed using data normalization, data randomization and data division. The data was divided into three separate datasets; training, validation and testing with the ratio 70:15:15, respectively. In the ANN development, the hidden neurons were varied from 1 to 10 with the implementation of Resilient Backpropagation training algorithm. Some criteria must be met before the ANN's performance can be accepted. The criteria were confusion matrix, accuracy, sensitivity, specificity, precision, mean square error (mse) and number of epochs.The ANN performance using three and four compounds from PCA and stepwise regression were compared to each other. The ANN method was summarized below in Figure 2.

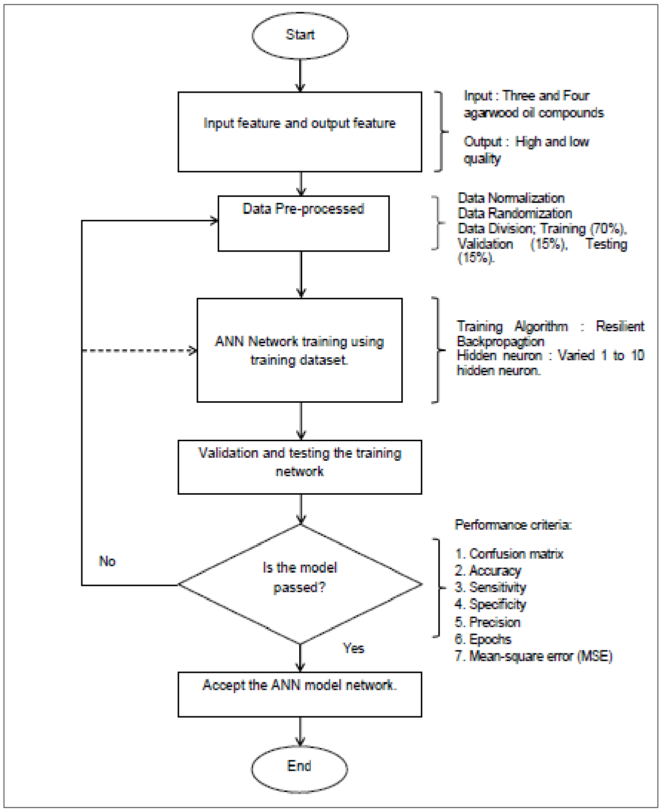

Figure 2 Flowchart of detail experiment of ANN classifier

III. Results and Discussion

The results were divided into three sections; *(A) ANN with three compounds*, *(B) ANN with four compounds* and *(C) Comparison of ANN performance for subsections (A) and (B)*.

A. ANN technique using Three compounds selected from PCA technique

Table 1 shows the results of training, validation and testing dataset using three compounds of agarwood oil. For training accuracy, the results obtained were varied from 79.4% to 95.6%. For the validation, the accuracy obtained was from 71.4% to 100%. The testing accuracy varied from 85.7% to 100%. The minimum mse value was found at hidden neurons four which was 0.044 while the maximum mse value found at hidden neuron two which was 0.113. Finally, the hidden neuron three was chosen as the best hidden neuron as getting higher performance of accuracy at early stage of training, validation and testing with the lower mse value.

TABLE I
Training Dataset Using Three Compounds Based on PCA Technique

Hidden neurons	Accuracy			MSE value
	Training	Validation	Testing	
1	82.4	71.4	85.7	0.102
2	79.4	85.7	85.7	0.113
***3**	**91.2**	**100**	**100**	**0.088**
4	95.6	85.7	92.9	0.044
5	80.9	78.6	85.7	0.101
6	92.6	100	92.9	0.075
7	88.2	92.9	100	0.101
8	94.1	100	85.7	0.059
9	92.6	92.9	100	0.074
10	92.6	100	92.9	0.074

**best hidden neuron*

B. ANN technique using Four compounds selected from Stepwise Regression

Table 2 tabulated the results of three dataset using four compounds of agarwood oil. The training accuracy was varied from 86.8% to 95.6%. For validation, the accuracy obtained were 92.6% (hidden neuron two, four and six) and 100% for others. While the testing dataset, the accuracy was from 85.7% to 100%. For the mse value, the minimum value found at hidden neuron three which was 0.026 while the maximum value found at hidden neuron one which was 0.047. Finally, the hidden neuron one was chosen as the best hidden neuron as all dataset performed good accuracy at early stage of training with a lower mse value.

TABLE II
Training Dataset Using Four Compounds Based on Stepwise Regression Technique

Hidden neurons	Accuracy (%)			MSE value
	Training	Validation	Testing	
***1**	**86.8**	**100**	**100**	**0.047**
2	92.6	92.9	92.9	0.035
3	95.6	100	85.7	0.026
4	95.6	92.9	85.7	0.027
5	95.6	100	85.7	0.030
6	92.6	92.9	92.9	0.034
7	92.6	100	92.9	0.039
8	92.6	100	92.9	0.034
9	89.7	100	100	0.039
10	94.1	100	85.7	0.032

**best hidden neuron*

C. Comparison of ANN performance

The ANN performance in subsection (A) and (B) was compared to each other using performance criteria and the best ANN performance was chosen. The comparison was made using the selected hidden neuron that has been chosen in *(A)* and *(B)*. Figure 3 and figure 4 shows the overall confusion matrix for ANN technique using three and four compounds of agarwood oil respectively. In the confusion matrix, the data is classified into group 0 (low quality) and group 1 (high quality). Figure 3 show that from 96 numbers of samples, the predicted group successfully predicts 12 and 6 samples into group 0 and 78 samples into group 1. Besides, Figure 4 shows predicted group successfully predicted 14 and 4 samples into group 0 while 5 and 73 samples into group 1.

Table 3 clearly summarized the parameters from the confusion matrix. The accuracy obtained by three compounds was 93.8% while for four compounds were 90.6%. The sensitivity for three compounds was 100% and 73.7% obtained by four compounds. The specificity was 92.9% and 94.8% for three and four compounds respectively. The precision obtained by three compounds were 66.7% and 77.8% obtained by four compounds.

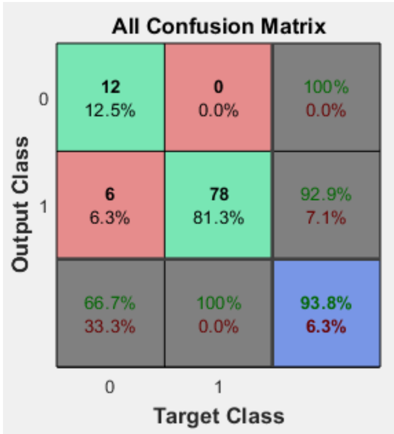

Figure 3 Confusion matrix using Three compounds agarwood oil

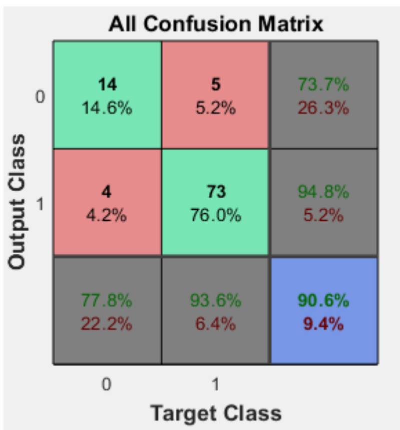

Figure 4 Confusion matrix using Four compounds agarwood oil

TABLE III
ACCURACY, SENSITIVITY, SPECIFICITY, PRECISION

Parameters	**3 Compounds**	**4 Compounds**
Accuracy	93.8%	90.6%
Sensitivity	100%	73.7%
Specificity	92.9%	94.8%
Precision	66.7%	77.8%

Table 4 shows the mse value and epoch value for three and four agarwood oil compounds of the chosen hidden neuron, respectively. Based on the table, the mse value obtained by the four compounds is lower than the mse value obtained by the three compounds which was 0.044. For the epoch value, the ANN using four compounds obtained lower epoch value which was 42 compared to 58 that obtained by the three compounds. This proved that, it performed fastest convergence time as it stop the training during epoch 42 that's mean the model obtained lowest validation error at this stage.

TABLE IV
COMPARISON OF MSE AND EPOCH VALUE USING THREE AND FOUR COMPOUNDS AGARWOOD OIL

	ANN with input of Three compounds	ANN with input of Four compounds
Choosen hidden neuron	Hidden neuron 3	Hidden neuron 1
MSE value	0.088	0.044
Epochs	58	42

Table 5 shows the final design parameter for the ANN architecture and ANN parameters. The ANN performance using four compounds of agarwood oil has been chosen by having better performance compared to ANN performance using three compounds of agarwood oil. The input feature was four referring to four significant compounds selected by stepwise regression. The number of hidden neuron was one due to at the early stage of training using one hidden neuron, the accuracy of training, validation and testing dataset obtained good value with the lower mse value. The choosen ANN performance using four compounds are due to the theory of economical reason, the least hidden neuron used in training the ANN was good to avoid long computational training time and overfitting problems [25], [26]. Besides, the minimum mse value indicated that the data had fitted the ANN model very well [27].

TABLE V
FINAL DESIGN PARAMETER

Parameters	Value
Number of input neuron	4
Number of hidden neuron	1
Number of output neuron	1
Algorithm	RBP
Mse value	0.044
Epoch value	42

IV. CONCLUSION

The performance of ANN using different compounds of agarwood oil has been presented in this research study. This study showed that the ANN perfomed well using the compounds selected from the stepwise regression. The compounds are β-agarofuran, ϒ-Eudesmol, Longifolol, and Eudesmol. Based on the results, one hidden neuron is sufficient to classify the agarwood oil into high and low quality compared with the ANN using the compounds from the PCA. The mse value using four compounds is also lower which is 0.044 at epoch 42. Hence, the computational load is reduced.

ACKNOWLEDGEMENT

This research is funded by Institute of Research Management and Innovation (IRMI), Universiti Teknologi MARA Shah Alam, Selangor and Universiti Teknologi MARA, Cawangan Johor, Kampus Pasir Gudang under Grant No: 600-IRMI/FRGS 5/3 (224/2019).

REFERENCES

[1] N. A. M. Ali, N. Ismail, and M. N. Taib, "Analysis of Agarwood oil (Aquilaria Malaccensis) based on GC-MS data," *Proc. - 2012 IEEE 8th Int. Colloq. Signal Process. Its Appl. CSPA 2012*, pp. 470–473, 2012.

[2] Y. Liu, J. Wei, Z. Gao, Z. Zhang, and J. Lyu, "A Review of Quality Assessment and Grading for Agarwood," *Chinese Herb. Med.*, vol. 9, no. 1, pp. 22–30, 2017.

[3] N. Ismail, N. A. Mohd Ali, M. Jamil, M. H. F. Rahiman, S. N. Tajuddin, and M. N. Taib, "A Review Study of Agarwood Oil and Its Quality Analysis," *J. Teknol. (Sciences Eng.*, vol. 1, no. 68, pp. 37–42, 2014.

[4] N. Ismail, M. H. F. Rahiman, M. N. Taib, M. Ibrahim, S. Zareen, and S. N. Tajuddin, "A review on agarwood and its quality determination," *Proc. - 2015 6th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2015*, pp. 103–108, 2016.

[5] M. A. Nor Azah, S. Saidatul Husni, J. Mailina, L. Sahrim, J. Abdul Majid, and Z. Mohd Faridz, "Classification of agarwood (gaharu) by resin content," *J. Trop. For. Sci.*, vol. 25, no. 2, pp. 213–219, 2013.

[6] S. Akter, M. T. Islam, M. Zulkefeli, and S. I. Khan, "Agarwood Production - A Multidisciplinary Field to be Explored in Bangladesh," *Int. J. Pharm. Life Sci.*, vol. 2, no. 1, pp. 22–32, 2013.

[7] M. H. Haron, M. N. Taib, N. Ismail, N. A. Mohd Ali, and S. N. Tajuddin, "Statistical analysis of agarwood oil compounds based on GC-MS data," *2018 9th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2018 - Proceeding*, no. August, pp. 27–30, 2019.

[8] M. Aqib *et al.*, "Agarwood Oil Quality Classification using Support Vector Classifier and Grid Search Cross Validation Hyperparameter Tuning," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 6, pp. 6–11, 2020.

[9] L. T. Wyn and N. A. Anak, *Wood for the Trees : a Review of the Agarwood (Gaharu) Trade in Malaysia*. 2010.

[10] N. Ismail, M. H. F. Rahiman, M. N. Taib, M. Ibrahim, S. Zareen, and S. N. Tajuddin, "Direct Thermal Desorption (DTD) extraction for different qualities of agarwood incense analysis," *Proceeding - 2016 IEEE 12th Int. Colloq. Signal Process. its Appl. CSPA 2016*, no. March, pp. 291–294, 2016.

[11] N. S. Zubir *et al.*, "Pattern Classifier of Chemical Compounds in Different Qualities of Agarwood Oil Parameter using Scale Conjugate Gradient Algorithm in MLP," *2017 IEEE 8th Control Syst. Grad. Res. Colloquium, ICSGRC 2017 - Proc.*, no. March, pp. 10–12, 2017.

[12] N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, M. Jamil, and S. N. Tajuddin, "Application of ANN in agarwood oil grade classification," *Proc. - 2014 IEEE 10th Int. Colloq. Signal Process. Its Appl. CSPA 2014*, pp. 216–220, 2014.

[13] M. S. Najib, N. A. Mohd Ali, M. N. Mat Arip, M. Abd Jalil, and M. N. Taib, "Classification of Agarwood Oils Using ANN," *Int. J. Electr. Electron. Syst. Res.*, vol. 5, 2012.

[14] M. H. Haron, M. N. Taib, N. Ismail, N. Azahmohdali, and S. N. Tajuddin, "Grading of agarwood oil quality based on its chemical compounds using self organizing map (SOM)," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 7, pp. 3728–3736, 2020.

[15] I. M. Yassin, A. Zabidi, N. Ismail, and H. Jantan, "Differentiation of Agarwood Oil Quality Using Support Vector Machine (SVM)," *J. Eng. Appl. Sci.*, vol. 12 (15), pp. 3810–3812, 2017.

[16] K. A. A. Kamarulzaini, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, and S. N. Tajuddin, "Evaluation of RBF and MLP in SVM kernel tuned parameters for agarwood oil quality classification," *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, no. March, pp. 250–254, 2018.

[17] Ö. Ki and E. Uncuo, "Comparison of three back-propagation training algorithms for two case studies," vol. 12, no. October, pp. 434–442, 2005.

[18] M. Riedmiller, "A Direct Adaptive Method for Faster Backpropagation Learning : The RPROP Algorithm," 1993.

[19] N. Coskun and T. Yildirim, "The effects of training algorithms in MLP network on image classification," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2, no. see 17, pp. 1223–1226, 2003.

[20] S. Kaur, A. S. Sharma, H. Kaur, and K. Singh, "Gene selection for tumor classification using resilient backpropagation neural network," 2016.

[21] M. Misra and B. Warner, "Understanding Neural Networks as Statistical Tools," *Am. Stat.*, vol. 50, no. 4, pp. 284–293, 1996.

[22] N. S. A. Zubir, M. A. Abas, N. Ismail, N. A. M. Ali, M. H. F. Rahiman, and N. K. Mun, "Analysis of Algorithms Variation in Multilayer Perceptron Neural Network for Agarwood Oil Qualities Classification," *2017 IEEE 8th Control Syst. Grad. Res. Colloquium, ICSGRC 2017 - Proc.*, no. August, pp. 4–5, 2017.

[23] N. Ismail, "ANN Modeling for Agarwood Oil Significant Compounds for Oil Quality Discrimination," PhD Dissertation, Faculty of Electrical Engineering, Universiti Teknologi Mara (UiTM), 2014.

[24] M. H. Haron, M. N. Taib, N. Ismail, N. A. Mohd Nor, and S. N. Tajuddin, "Determination of Substantial Chemical Compounds of Agarwood Oil for Quality Grading," *J. Electr. Electron. Syst. Res.*, vol. 17, no. DEC 2020, pp. 50–59, 2020.

[25] A. Moghadassi, S. M. Hosseini, F. Parvizian, I. Al-Hajri, and M. Talebbeigi, "Predicting the supercritical carbon dioxide extraction of oregano bract essential oil," *Songklanakarin J. Sci. Technol.*, vol. 33, no. 5, pp. 531–538, 2011.

[26] A. Zabidi, L. Y. Khuan, W. Mansor, I. M. Yassin, and R. Sahak, "Detection of infant hypothyroidism with mel frequency cepstrum analysis and multi-layer perceptron classification," *Proc. - CSPA 2010 2010 6th Int. Colloq. Signal Process. Its Appl.*, vol. 2010–Janua, pp. 140–144, 2010.

[27] Z. A. Bakar, N. M. Tahir, and I. M. Yassin, "Classification of Parkinson's disease based on Multilayer Perceptrons Neural Network," *Proc. - CSPA 2010 2010 6th Int. Colloq. Signal Process. Its Appl.*, pp. 232–235, 2010.