Identification of Agarwood Oil qualities using Radial Basis Function (RBF) Neural Network

N.S.A. Zubir[1], M.A.Abas[2], Nurlaila Ismail[1],Nor Azah M.Ali[2], M.H.F.Rahiman[1], K.M.Ng[1], Saiful.N.T[3] and M.N.Taib[1]

[1]Faculty of Electrical Engineering UiTM Shah Alam, Selangor
[2]Natural Product Program Forest Research Institute of Malaysia (FRIM) Kepong, Selangor
[3] Faculty of Industrial and Science Technology (FIST), University Malaysia Pahang (UMP).
*nurulshakilaaz@gmail.com

***Abstract*—Agarwood is known as a valuable non-timber product found in the dark fragrant resin in the stem, branch and roots of certain species of *Aquilaria*. Agarwood oil is one of the popular essential oil that has been used not only in Asian but in the world. The price of the agarwood oil is referring based on the quality of agarwood oil. The agarwood oil have distinct pattern which can be discriminating the qualities of agarwood oil by classification technique such as radial basis function. The Radial Basis Function networks (RBFNs) are commonly used for complex pattern classification. This study examines the performance of radial basis function of identifying the quality of agarwood oil either high or low quality. The dataset consists of the abundances of significant compounds (%) and qualities of the agarwood oil. The result reveals that the classification using RBF technique, performs slightly have a better performance of MSE values depends on the 100 maximum numbers of neurons and 3 number of spread. The hypothesis from this study is the larger number of spread the smoother the function approximation. Besides that, the small number of spread the large number of neurons required to fit a smooth function.**

Index Terms*—Radial Basis Function Networks (RBFNs), pattern classification, agarwood oil, Mean Square Error (MSE),

I. INTRODUCTION

Agarwood oil is one part of the essentials oil and it will be classified as a volatile compound that produces from the extraction process [1]. The factor influences the best quality of agarwood oil are temperature, pressure, particle of raw materials, chemical composition and distillation time. Nowadays, agarwood oil is very common and has very high demand due to its variety of usage in human life. Basically, agarwood oil qualities are based on the odor and color. The quality of agarwood is divided into two types which are high and low[2]. Other than that, quality of agarwood oil can be detect refer to the quantitative difference (or chemical composition) table[3], gas chromatograph (GC) charts and thin layer chromatogram (TLC) [4].

Radial basis function neural networks are commonly applied in robotics [5], medical diagnostic [6]-[7], PID and dynamic system design [8]-[9] Radial basis function is one of the familiar methods to learn task faster and widely used in data classification [10]. It was found that the radial basis function provides a reliable result in detection, categorization and classification [11]. This algorithm is simple, good generalization and powerful distance learning scheme which has been used extensively and great success for numerical data [12]-[14]. This is due to its rapid learning capacity and activation functions.

RBF network is known as function approximation and classifications. In Fig.1 show that the network is supervised 3-layer feed-forward networks that has hidden layer consisting of radial basis activation function [11]. RBF commonly used for complex pattern classification. It is because for high dimensional space non-linearly is more likely to be linearly separable than in a low dimension space. Additionally, linear separable is easier to solve classification problem.

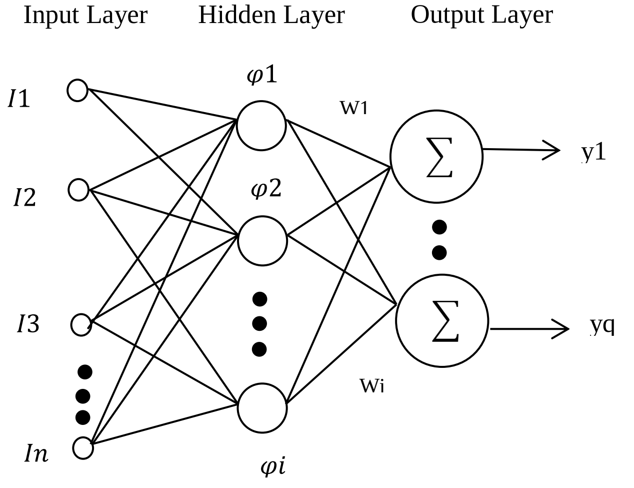

Fig.1: The structure of RBF network

This paper described the analysis of agarwood oil qualities by using Radial Basis Function (RBF) neural networks models. This paper is submitted on 6th June 2017. Accepted on 9th November 2017. This research was support by LESTARI Grant no (600-IRMI/DANA 5/3/LESTARI (0172/2016) for my financial support.

Nurul Shakila Ahmad Zubir is wilth Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. (Email:nurulshakilaaz@gmail.com)

Nurlaila Ismail is lecturer system Faculty of electrical Engineering, Universiti Teknologi MARA (Email: nrk_my@yahoo.com)

Mohd Nasir Taib is a Dean of Electrical Engineering, Faculty of electrical Engineering,Universiti Teknologi MARA (Email:dr.nasir.ieee.org)

Ng kok Mun is lecturer system Faculty of electrical Engineering, Universiti Teknologi MARA (Email: kokmun77@yahoo.com.sg)

The basic three linked note in RBF networks consist of input layer which is composed a signal sources nodes, hidden layer and output layer. The output layer is performed when the sum of the linear weighting the output of hidden layer is develop. The theoretical Gaussian RBF stated that, the sensitivity of hidden neuron can be tuned by adjusting the spread, where the larger spread implies less sensitivity [15]-[16]. Other than that, the hidden layer in RBF is non-linear and the output is linear. Furthermore, neurons in hidden layer consist of Gaussian transfer function. The argument of activation function of each hidden neuron in RBF computes the Euclidean distance between input vector and the center of that unit. The equation of Euclidian distance stated as below:

$$\text{Output} = \sum_{j=1}^{m} W_j f(X_j) \qquad (1)$$

$$f(X_j) = radbas\,(X_j) \qquad (2)$$

Where $X_j = \| I_{n=1 \to n} - C_j \|$ = Euclidian distance (3)

The distance between two point which is connecting two path involved coordinate (xi, yi) (xj, yj). Each x will go into the radbas activation function.

$$dj = \sqrt{(xi - xj)^2 + (yi - yj)^2} \qquad (3)$$

Besides that, RBF neural network is depends on the distance of the input from a given stored vector. The main vector is "interpolated" which is using three vector. In Fig.2, the main vector state as "P" and weight have 3 groups which are weight 1, 2 and 3. Each vector gives a contribution that depends on its weight and on its distance from the main point. RBF represent local receptors, where each point is a stored vector used in one RBF as illustrated in Fig.2.

Note: W1<W3<W2 (weight 2 is highest followed by weight 3 and weight 1)

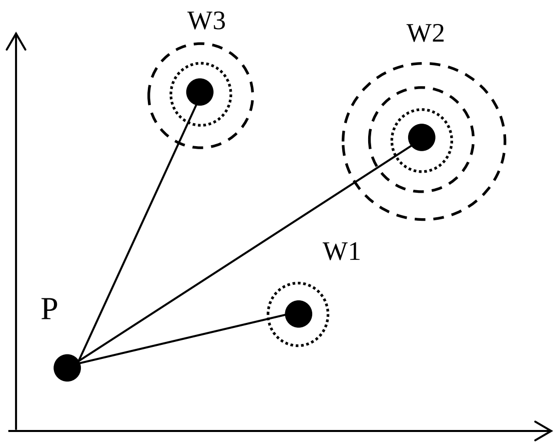

Fig.2: Radial Basis Function (RBF) Networks

II. METHODOLOGY

The database for agarwood oil was obtained from Forest Research Institute Malaysia (FRIM) & University Malaysia Pahang (UMP) [18] which is consist of gas chromatography-mass spectrometry (GC-MS) data. In details, it is consist of 7 different compounds of agarwood oil and their (abundances, %) represent high and low qualities of agarwood oil. The data was normalized by using min-max normalization before apply to RBF algorithms. The function of normalization is to maintain, reduce or eliminate data redundancy and standardized properly the data collection.

Fig.3 illustrates the overall methodology to analyze the performance of input and output data by applied RBF technique. The first stage is started with data collection which is data is divided into training and validation with the ratio 80:20. The data is the significant compounds were extracted by gas chromatography- mass spectrometry (GC-MS) data. Typically, dataset consist of 7 input parameter which are the (abundances, %) of chemical compounds and the output is the quality (high or low) of agarwood oil. Then, the data was divided into two parts, training and validation which are 80% for training and 20% for validation. RBF algorithm was applied by adjusting number of hidden neuron and number of spread in order to analyze the performance by compare the results based on MSE value.

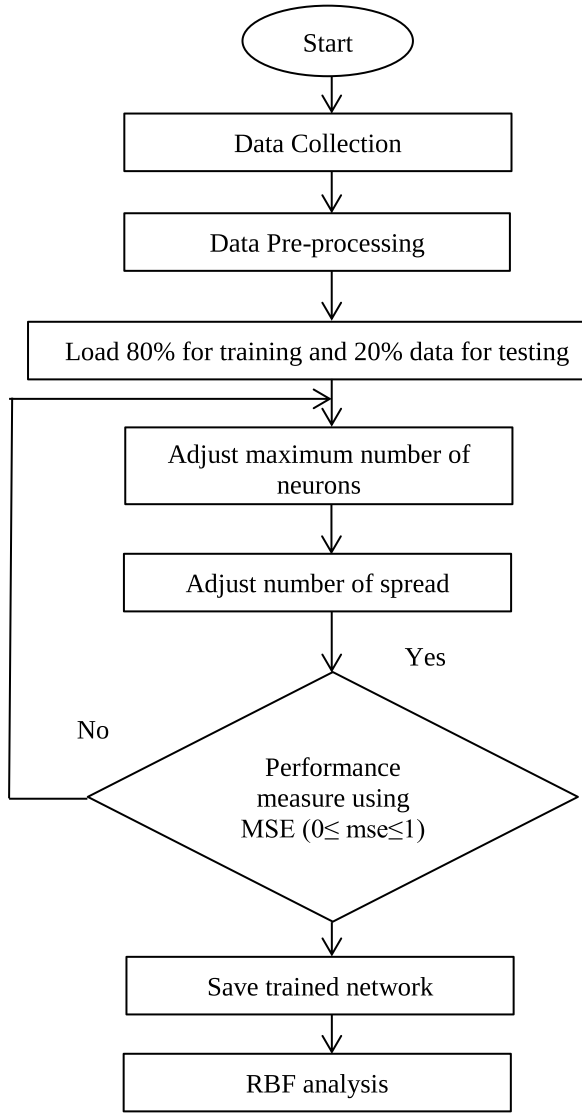

Fig.3: Flowchart of RBF analysis

III. RESULTS

This section describes the radial basis function analysis which is based on the two parameters adjustment which is by adjusting the maximum number of neurons and number of spread.

1) Data Collection

In this study, RBF analysis was used data from Forest Research Institute Malaysia (FRIM) & University Malaysia Pahang (UMP) which is consist of gas chromatography-mass spectrometry (GC-MS) data. The data consist of 7 compounds repsresent the abundances (%) which are C1 = β-agarofuran, C2 = α-agarofuran, C3 = 10, -epi-ϒ-eudesmol, C4 = ϒ-eudesmol, C5 = longifolol, C6 = hexadecanol and C7 = eudesmol. All compounds consist of high and low quality [19].

As shown in Fig.4 compound 1, 2, and 3 are significantly present in high quality and compound 5, 6 and 7 are significantly present in low quality. Compound 4 is neutral. It is because, this compound was found significantly in high and low qualities [20].

The total data is 672 which is 96 agarwood oil samples were used in this experiment. It is consist of 18 samples of low quality and 78 samples of high quality. Meanwhile, the target feature used is and the qualities of the agarwood oil which is 1(Low) or 2 (High). The development of algorithm from radial basis function (RBF) neural network in advanced signal processing will be used to discriminate the qualities of agarwood oil chemical compounds.

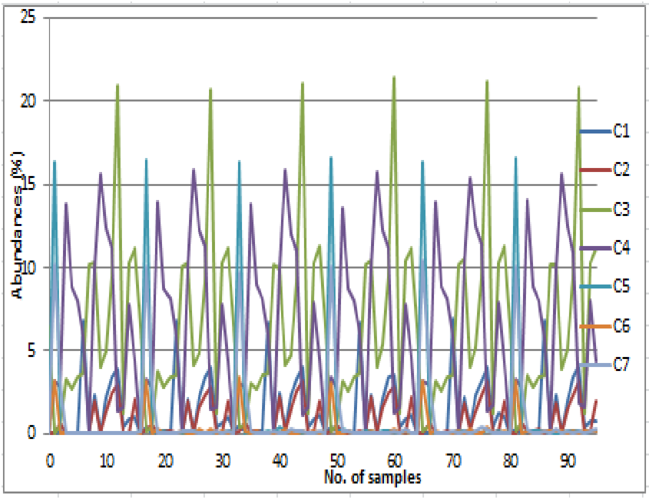

Fig.4: Total data of GC-MS

2) Data pre-processing

Table I: Before Min-Max Normalization

Compound	Sum	Average	Standard Deviation	Mean Square Error (MSE)
C1	130.72	1.367	1.8779	3.4895
C2	73.82	0.789	1.0212	1.0325
C3	617.5	6.432	5.2664	27.446
C4	643.3	6.701	4.8816	23.587
C5	105.92	1.103	3.9789	15.666
C6	52.66	0.2743	0.7611	0.5732
C7	85.40	0.8896	2.4959	6.1647

Table II: After Min-Max Normalization

Compound	Average	Mean Square Error (MSE)
C1	1.367	1.7067
C2	0.789	0.5694
C3	6.432	37.662
C4	6.701	40.865
C5	1.103	1.1398
C6	0.2743	0.0696
C7	0.8896	0.7327

Table I and II shows the comparison of MSE value after normalize the scale of the input with the range 0 to 1, the value of mean square error (MSE) was varies. MSE act as predictor to predict the measure of the difference the actual and the estimation value in terms is error. However, the value of MSE at compound 3 and 4 are higher it is because the value of the standard deviation and have higher maximum value of abundances. The values of standard deviation affect the performance of MSE. Furthermore, the graph in Fig.5 shows clearly the difference between before and after normalization.

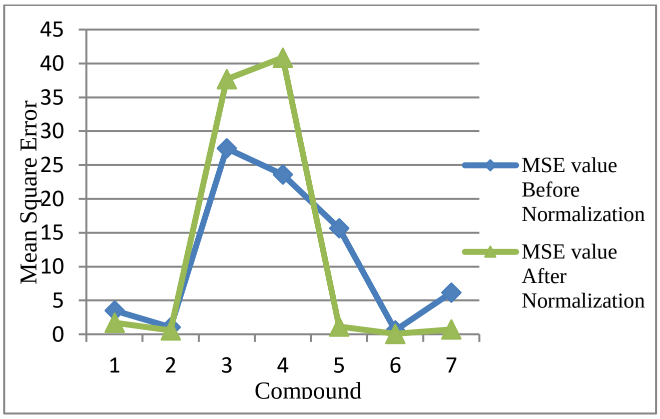

Fig.5: Comparison MSE values between before and after normalization

3) *Number of spread*

In Table III shows the network was trained by adjusting the value of number of spread. The number of spread was in the range of 1 to 100. The comparison with all number of spread showed that, the lowest value of MSE was using 1 and 3 number of spread.

Based on the residuals the number of 3 spread has a better performance because the residual performance show in Fig.8 and Fig.9 was closely related between the actual and the predicted value. However, the residual in Fig.6 and Fig.7 was closely but there are outliers for the output testing. Lastly, the large number of spread will smooth the function approximation but too large number of spread means a lot of number of neurons required to fit a smooth function.

Table III: Selection of number of spread

Number of Spread	**Max Numbers of neurons**	**Epochs**	**Mean Square Error (MSE)**
1*	100	96	1.34e-30
3*	100	96	3.95e-28
10	100	96	7.29e-21
20	100	96	8.73e-16
30	100	96	8.73e-16
40	100	96	8.73e-16
50	100	96	4.33e-10
60	100	96	8.24e-09
70	100	96	3.07e09
80	100	96	4.09e-07
90	100	96	7.58e-07
100	100	96	8.09e-07

*Lowest MSE value

In Fig.6 shows the residuals of training (77 data) by using 1 numbers of spread. The actual and the predicted performance were compared. The actual and predicted are absolutely same or accurate.

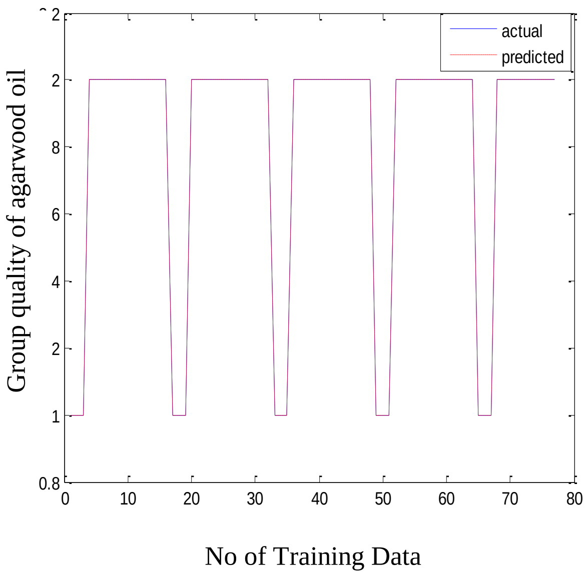

Fig.6: The training residual using 1 number of spread.

In Fig.7 shows the residuals of testing (19 data) by using 1 numbers of spread. The actual and the predicted performance were compared. The actual and predicted are not accurate

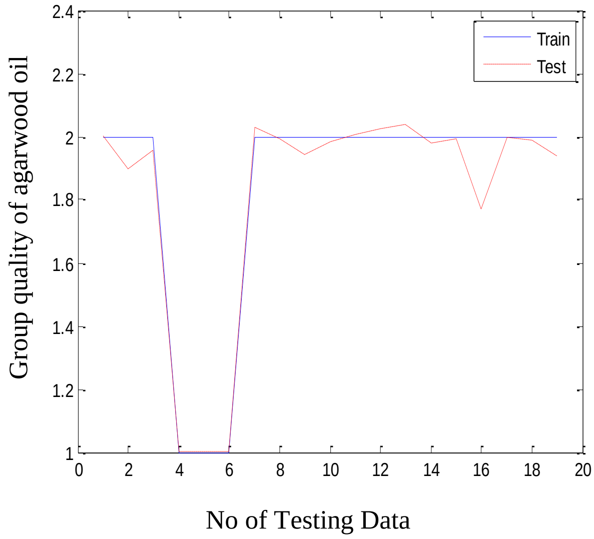

Fig.7: The testing residual using 1 number of spread.

In Fig.8 shows the residuals of training (77 data) by using 3 numbers of spread. The actual and the predicted performance were compared. The actual and predicted are absolutely same or accurate.

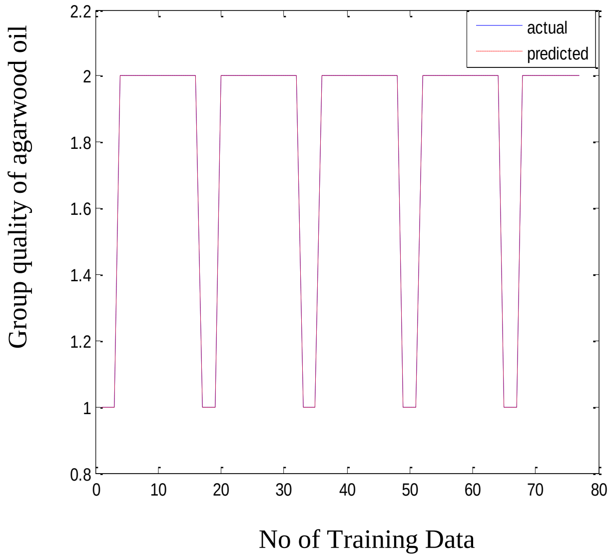

Fig.8: The training residual using 3 number of spread.

In Fig.9 shows the residuals of testing (19 data) by using 3 numbers of spread. The train which is the actual and the test which is the estimation was compared. The actual and estimation are not accurate.

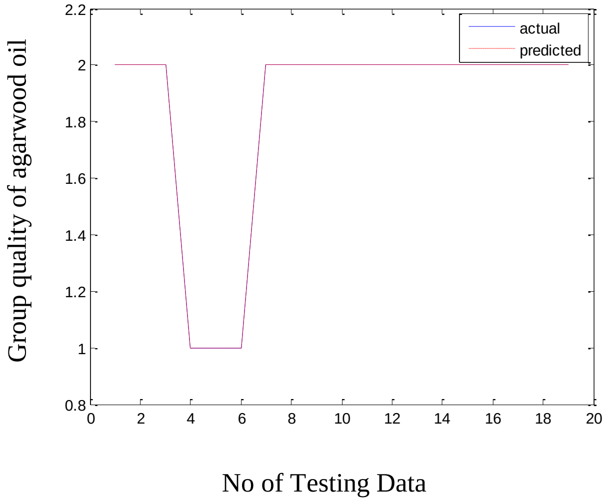

Fig.9: The testing residual using 3 number of spread.

4) Maximum number of neurons

In this section in Table IV, the network is trained until it achieves a relatively small error. The maximum number of neurons was increased gradually from 10 to 200. The mean square error (MSE) was recorded for each of maximum numbers of neurons.

The trained was using 3 number of spread. The epochs performance was analyzed depends on the maximum number of neurons. Starts from 100 numbers of neurons, the epochs run constantly by 96. It is refer to the all number of data set which is the total data is 96. However, the number of neurons which is lower than the numbers of data set (96 data sets) was has same value with epochs respectively. Furthermore, the value of MSE for the greater maximum number of neurons is the smallest error. Starts from 100 to 200 number of neurons the number of epochs and the MSE was constant. It will be conclude that, the numbers of neurons are added to the network until the error fall beneath an error goal or a maximum number of neurons have been reached. Besides that, the smaller number of spread required a lot of number of neuron to fit a smooth function. The performance was shown in Fig.9

Table IV: Selection of maximum number of neurons

Number of Spread	Max Numbers of neurons	Epochs	Mean Square Error (MSE
3	10	10	0.0245
3	20	20	0.0003
3	30	30	5.72e-05
3	40	40	1.24e-05
3	50	50	4.57e-06
3	60	60	3.09e-06
3	70	70	1.94e-06
3	80	80	2.08e -07
3	90	90	3.59e -09
3	100*	96	3.37e -26
3	150*	96	3.37e -26
3	200*	96	3.37e-26

*Lowest MSE value

Fig.10 shows the classification of RBF network training, the result shows 100% of the classification accuracy. It is support in the confusion matrix depends on the TP and TN value. Refer to the total value of training data (77 dataset), the predicted class has successful predicted fifteen numbers to the actual class 1 (low quality) and sixty two numbers to the actual class 2 (high quality). The sensitivity, specificity and precision for the classification behavior are 100% for the training data classification shown in Table V.

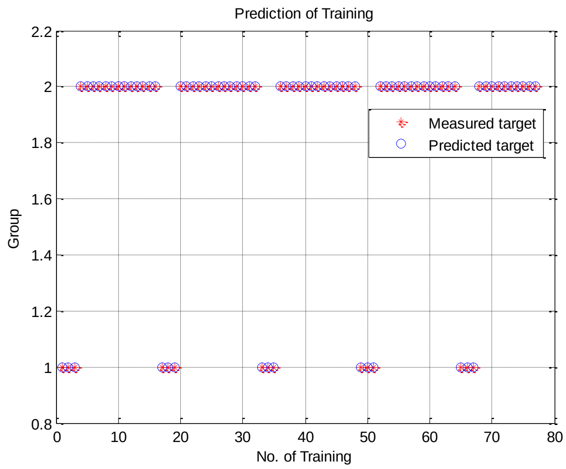

Fig.10: Classification of RBF network training

Table V: The RBF network classification performance measures for training data; (a) confusion matrix and (b) Sensitivity, Specificity and Precision

DATA CLASS	Predicted/Classified as positive	Predicted/Classified as negative
positive	15 (tp)	0 (fn)
negative	0 (fp)	62 (tn)

(a)

Parameter	Value
Accuracy	100.00%
Sensitivity	100.00%
Specificity	100.00%
Precision	100.00%

(b)

Fig.11 shows the classification of testing RBF network, the result shows 100% of the classification accuracy. It is support in the confusion matrix depends on the TP and TN value. Refer to the total value of training data (19 dataset), the predicted class has successful predicted three numbers to the actual class 1 (low quality) and sixteen numbers to the actual class 2 (high quality). The sensitivity, specificity and precision for the classification behavior are 100% for the training data classification shown in Table VI.

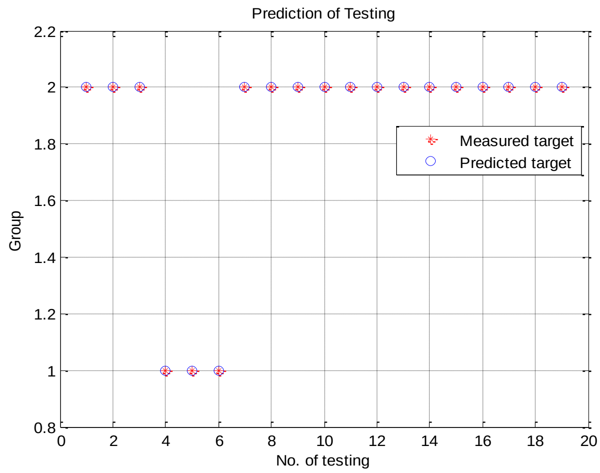

Fig.11: Classification of RBF network testing

Table 5.10: The RBF network classification performance measures for testing data; (a) confusion matrix and (b) Sensitivity, Specificity and Precision

DATA CLASS	Predicted/Classified as positive	Predicted/Classified as negative
positive	3 (tp)	0 (fn)
negative	0 (fp)	16 (tn)

(a)

Parameter	Value
Accuracy	100.00%
Sensitivity	100.00%
Specificity	100.00%
Precision	(a)100.00%

(b)

CONCLUSION

Based on the results obtained from this study, radial basis function (RBF) neural network technique can be used to identify the quality of agarwood oil. The differentiation of agarwood oil quality using intelligent pattern classifier (Radial basis function, RBF) has been successfully implemented. All chemical compounds are very significant to classify the agarwood oil quality by set as input and output, the quality of agarwood oil either high or low can be determined. RBF with 3 numbers of spread and 100 numbers of hidden neurons classification model is selected based on the maximum performance, accuracy and mean square error (MSE).

ACKNOWLEDGMENTS

The author would like to acknowledge PICO and Advanced Signal Processing (ASP) group for the guidance and encouragement in finishing this research. The author would also like to thank Faculty of Electrical Engineering, UiTM Shah Alam, Natural Product Program Forest Research Institute of Malaysia (FRIM) and Faculty of Industrial and Science Technology (FIST) UMP for data collection and analysis. Besides that LESTARI Grant no (600-IRMI/DANA 5/3/LESTARI (0172/2016) for my financial support.

REFERENCES

[1] X. Gao, M. Xie, S. Liu, X. Guo, X. Chen, Z. Zhong, L. Wang, and W. Zhang, "Chromatographic fingerprint analysis of metabolites in natural and artificial agarwood using gas chromatography – mass spectrometry combined with chemometric methods," *J. Chromatogr. B*, vol. 967, pp. 264–273, 2014.

[2] N. Ismail, M. Hezri, N. Azah, and M. Ali, "Application of ANN in Agarwood Oil Grade Classification," pp. 7–9, 2014.

[3] D. Yang, W. Li, W. Dong, J. Wang, W. Mei, and H. Dai, "Fitoterapia Five new 5 , 11-epoxyguaiane sesquiterpenes in agarwood ' Qi-Nan ' from Aquilaria sinensis," *Fitoterapia*, vol. 112, pp. 191–196, 2016.

[4] N. Ismail, M. Hezri, F. Rahiman, and A. A. Malaccensis, "A Review on Agarwood and Its Quality Determination," pp. 103–108, 2015.

[5] T. T. Nguyen, H. Van Nguyen, and T. V. Hoa, "Synthetic tracking controller for robot manipulator with flexible joints, dynamics of executive motors and affect of disturbance based on radial basic function (RBF) neural network," *2014 IEEE Int. Conf. Robot. Biomimetics, IEEE ROBIO 2014*, pp. 1469–1474, 2014.

[6] J. P. Kelwade, "Radial basis function Neural Network for Prediction of Cardiac Arrhythmias based on Heart rate time series," *2016 IEEE First Int. Conf. Control. Meas. Instrum. Radial*, no. Cmi, pp. 454–458, 2016.

[7] B. Dog, "Expert Systems with Applications ECG beat classification using particle swarm optimization and radial basis function neural network," *Expert Syst. with Appl. 37*, vol. 37, pp. 7563–7569, 2010.

[8] D. S. Yeung, B. Sun, W. W. Y. Ng, And P. P. K. Chan, "Construction Of Radial Basis Function Neural Networks Via A Minimization Of Its Localized Generalization Error," *Proc. Ninth Int. Conf. Mach. Learn. Cybern.*, No. July, Pp. 11–14, 2010.

[9] I. Sutrisno, M. A. Jami'In, J. Hu, N. Mariun, and M. H. Marhaban, "Nonlinear model-predictive control based on quasi-ARX radial-basis function-neural-network," *Proc. - Asia Model. Symp. 2014 8th Asia Int. Conf. Math. Model. Comput. Simulation, AMS 2014*, pp. 104–109, 2014.

[10] H. Yu, S. Member, T. Xie, S. Paszczyñski, and S. Member, "Advantages of Radial Basis Function Networks for Dynamic System Design," *5438 IEEE Trans. Ind. Electron. VOL. 58, NO. 12,*

DECEMBER 2011, vol. 58, no. 12, pp. 5438–5450, 2011.

[11] I. Yunan, I. M. Yassin, S. Farid, S. Adnan, M. Hezri, F. Rahiman, and U. T. Mara, "Identification of Essential Oil Extraction System using Radial Basis Function (RBF) Neural Network," *2012 IEEE 8th Int. Colloq. Signal Process. its Appl. Identif.*, pp. 495–499, 2012.

[12] M. A. Slim, A. Abdelkrim, and M. Benrejeb, "An Efficient Handwriting Velocity Modelling for Electromyographic Signals Reconstruction Using Radial Basis Function Neural Networks," *7th Int. Conf. Model. Identif. Control (ICMIC 2015)*, no. Icmic, pp. 3–8, 2015.

[13] M. Mohammadi, R. Ronge, J. R. Chandiramani, and S. Nandy, "An Accelerator for Classification using Radial Basis Function Neural Network," *Indian Inst. Sci. Bangalore, India* -, vol. 560012, pp. 137–142, 2015.

[14] R. A. Husain, A. S. Zayed, W. M. Ahmed, and H. S. Elhaji, "Image Segmentation with Improved Watershed Algorithm Using Radial Bases Function Neural Networks," *16th Int. Conf. Sci. Tech. Autom. Control Comput. Eng. - STA'2015, Monast. Tunis. December 21-23, 2015 STA'2015-PID3836-IFR Image*, 2015

[15] A. Alexandridis, E. Chondrodima, S. Member, N. Giannopoulos, and H. Sarimveis, "A Fast and Efficient Method for Training Categorical Radial Basis Function Networks," pp. 1–6, 2016.

[16] M. Murakami and N. Honda, "Comparison between real-time learning capabilities of the ids method and radial basis function networks," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, pp. 1262–1267, 2007.

[17] N. Ismail, M. Hezri, and F. Rahiman, "The Grading of Agarwood Oil Quality using k- Nearest Neighbor (k-NN)," no. December, pp. 13–15, 2013.

[18] N. Ismail, M. Hezri, and F. Rahiman, "Analysis of Chemical Compounds of Agarwood Oil based on Headspace-Solid Phase Microextraction combined with Gas Chromatography Mass-Spectrometry," pp. 8–10, 2013.

[19] N. Azah and M. Ali, "Analysis of Agarwood Oil (Aquilaria Malaccensis) Based on GC-MS Data," pp. 470–473, 2012.

[20] T. Mahira, N. Imamogh, J. Gomez-Tames, K. Kita, and W. Yu, "Modeling bimanual coordination using back propagation neural network and radial basis function network," *IEEE Int. Conf. Robot. Biomimetics*, pp. 1356–1361, 2014.

Nurul Shakila Binti Ahmad Zubir was born in Malaysia, on January 1992. She received Diploma in Electrical Engineering (Instrumentation) and Bachelor of Engineering (Hons.) from Universiti Teknologi Mara (UiTM) in 2013 and 2016. She currently enrolled as a Master student at Universiti Teknologi MARA, (UiTM).

Dr Nurlaila Ismail was born in Malaysia, Dec 10 1982. She received her PhD in Electrical Engineering from Universiti Teknologi MARA, Malaysia in 2015, MSc in Electrical Engineering from Universiti Teknologi MARA, Malaysia in 2011 and Bachelor in Electrical & Electronics Engineering (Hons.) from Universiti Teknologi MARA, Malaysia in 2005.She currently lectures at Universiti Teknologi MARA, Malaysia. Her research interests are artificial intelligence and system identification.

Prof Mohd Nasir Taib was born in Malaysia. He received his degree in Electrical Engineering from the University of Tasmania, Australia, MSc in Control Engineering from Sheffield University, UK, and PhD in Instrumentation from UMIST, UK. Currently, he is a Professor and Dean of The Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM. Currently, his research interests are system and instrumentatation.

Dr Ng kok mun was born in Malaysia. He graduated from Universiti Teknologi Tun Hussein Onn, Kuittho. Currently, he is a lecturer at Universiti Teknologi MARA, Malaysia. His research interests are system modelling, simulation and optimization computational intelligence and intelligent transportation development.