

k-nearest neighbor modelling of agarwood oil samples available in capital of Malaysia market

Erny Haslina Abd Latib¹, Nurlaila Ismail², Saiful Nizam Tajuddin³, Jasmin Jamil⁴,
Zakiah Mohd Yusoff⁵

¹Faculty Science and Technology Industry, Universiti Malaysia Pahang, Kuantan, Malaysia

²School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia

³Bioaromatic Research Centre, Universiti Malaysia Pahang, Kuantan, Malaysia

⁴Discovery Agarwood Research Sdn Bhd, Balik Pulau, Pulau Pinang

⁵School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Cawangan Johor, Malaysia

Article Info

Article history:

Received May 7, 2021

Revised Jan 24, 2022

Accepted Feb 7, 2022

Keywords:

Agarwood oil

Classification

k-nearest neighbor modelling

Machine learning

Malaysia markets

ABSTRACT

Agarwood oil is consumed during traditional ceremonies and even in medicinal purposes due to its effective therapeutic characteristic. As a part of ongoing research on agarwood oil, this paper presented a k-nearest neighbor (k-NN) modelling of agarwood oil samples available in the capital of Malaysia market. The work involved agarwood oil samples from three sources which are lab, local manufacturer and market. The inputs are the chemical compounds and the output is the oil's resources. The input-output was divided into training and testing dataset with the ratio of 80% to 20%, respectively, before they were fed to the k-NN for model development as well as model validation. During the model development, the k-value was varied from 1 to 5, and their accuracy was observed. The result showed that the k=1 and k=2 shared the similar accuracy for training and testing datasets, which are 98.63% and 100.00%, respectively. This study revealed the capabilities of the k-NN model in classifying the agarwood oil samples to the three sources: lab, local manufacturer and market. It was a significant study and contributed to further work especially those related to agarwood oil research area.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nurlaila Ismail

School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA

Shah Alam, Selangor

Email: nurlaila0583@uitm.edu.my

1. INTRODUCTION

Agarwood essential oil is widely used for perfumes and medication purposes. Usually, agarwood is resin-impregnated heartwood originating in the tree trunk of the genus *Aquilaria*. It is referred to as dense, heavy, or fragrant resinous wood. Formation of resin is produced due to natural infection of microorganism as well as insects [1]–[3]. In order to protect itself, genus *Aquilaria* secretes substances enriched with high volatile organic compounds called resin around the infected area. The resin that is produced in response to the infection is normally dark in color. These substances are known as secondary metabolites which play a role in plant defense mechanisms [4]–[6]. Due to the high demand, the prices of the pure agarwood essential oil become expensive. The price of agarwood is reflected by its quality. High quality agarwood may reach up to US\$ 30, 000 per kilogram. However, low quality agarwood may only cost a few dollars per kilogram [4]. Color and odor of agarwood affect its quality [7]. Generally, high quality agarwood features in dark brown color and long-lasting aroma. Light color agarwood and easily faded odor is indicated as low quality.

Literature showed that agarwood oils are mainly composed of the terpenoids group, which is sesquiterpenes, sesquiterpene alcohols, oxygenated compounds and chromone derivatives. A number of compounds have been extracted from *Aquilaria* species such as alkaloids, tannins, phenols, terpenoids, quinones and flavonoid [5]–[8]. Terpenoids that are also called isoprenoids are one of the largest groups of natural products found in nature with more than 30,000 known examples and their example is growing. Terpenoids are based on the isoprene unit, which is 2-methylbuta-1, 3-diene. As claimed previously, the highest quality of agarwood oil has a lot of resin which means that it contains various kinds of oxygenated sesquiterpenes and chromone derivatives [8]. Reflective to that, the presence of sesquiterpenes differentiated the premium of agarwood oil in the market. Some of the important sesquiterpene compounds which are agarospirol, jinkohol-eremol, jinkohol and kusunol may contribute to the characteristic aroma of agarwood [7], [8]. While in Malaysia agarwood oils, some of chemical compounds detected are 3-phenyl-butanone, α -guaiene, β -agarofuran, α -agarofuran, agarospirol and jinkoh-eremol as in previous study by [9], [10]. Unfortunately, only a few reports of agarwood oil quality studies have been conducted in Malaysia [9], [11]–[13]. Therefore, the previous paper reported the differences in the oil compositions between agarwood essential oil that was obtained from the Kuala Lumpur market and from local traders with the aim to enrich knowledge about the oil [14], [15]. As a part of ongoing research from the previous research [16], this paper presented k-NN Modelling of agarwood oil samples available in the capital of Malaysia market as one of the study cases.

k-nearest neighbours (k-NN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). k-NN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique. It is a supervised machine learning algorithm, and has been proven its capability for both classification and regression problems as well as for both binary and multi-class classifications [17]–[20]. Consider the following binary classification as in Figure 1. After a new point is added to the data set, the new point is determined to belong to which class it was grouped. After the number of neighbours (k) is selected, the nearest neighbours are identified based on the Euclidean distance of the new point to those neighbours. Calculation of Euclidean distance or Euclidean metrics is shown in Figure 2 [15], [16].

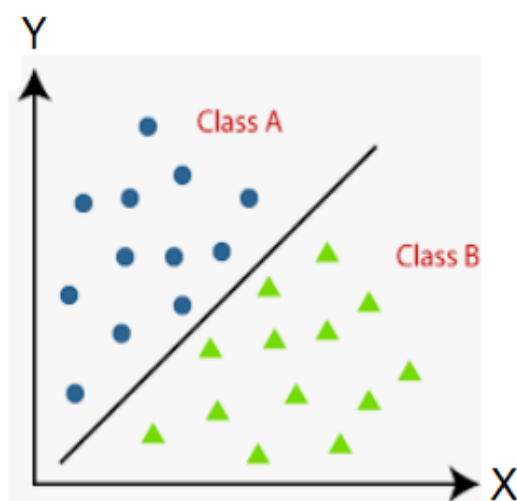


Figure 1. Binary classification [20]

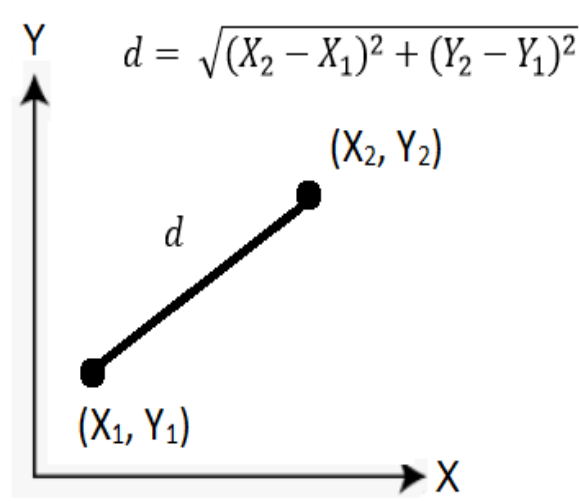


Figure 2. Euclidean distance concept [20]

There are several distance metrics that can be used for k-NN modelling such as Euclidean, City-block, Cosine, and Correlation and for this paper Euclidean distance was chosen [21]. The data needs to be scaled during the training stage. For example, for k value equal to 7 is selected, so the seven nearest neighbors are identified. The new point belongs to a class which has the highest neighbour. The most excellent esteem for K is chosen in odd numbers, and in case an indeed number is chosen it may conclude up with the same number of neighbors. As the number of K increments the chosen locale might alter as on Figure 3, in the event that k=3 is selected at that point Class A (red) is planning to be the Class that the unused point has a place, in any case by expanding the k to 7, the chosen locale would be Class B (green) [22]–[25].

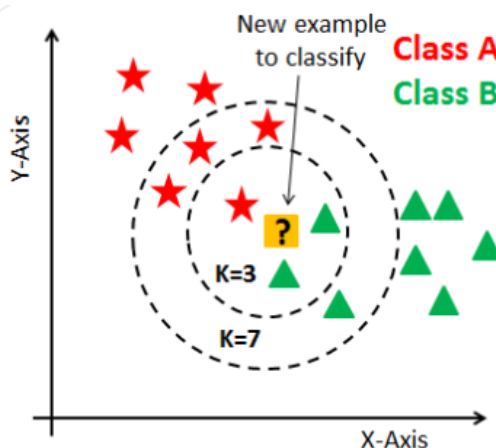


Figure 3. k=3 and k=7 and neighbours' new point [23]

2. MATERIALS AND METHOD

2.1. Agarwood oil samples and sample preparation

Samples of agarwood oil were collected from three different sources: i) extracted in the laboratory by using lab scale hydrodistillation (L), ii) obtained from industrial set-up hydrodistillation in factory directly from local manufacturer (LM), and iii) Kuala Lumpur market (M). As for lab scale hydrodistillation samples (L), they are agarwood chips, obtained from a nature forest located in Rompin, Pahang. The extraction was done by using hydrodistillation (HD) method. Several parameters of each extraction were applied to obtain the best and optimum condition for agarwood oil extract. Prior to extraction, ground agarwood chip was soaked in water for several days sequentially to break down the parenchymatous and oil glands. The samples were diluted in dichloromethane (DCM) with analytical grade for gas chromatography mass spectrometer (GC-MS) and gas chromatography system (GC-FID) determination. Details on the extraction procedure, analysis of GC-MS and GC-FID and identification of the compounds can be found in [15].

2.2. Development of k-NN modelling

Figure 4 shows the flowchart of the k-NN model development carried out in this study. The whole process started by getting all the raw data, samples preparation and identification from the previous researcher [14]. Next, pre-processing data has been carried out, which means the data are normalized, randomized and split into training and testing datasets with the ratio of 80% to 20%, respectively. The next stage is where the k-NN model was developed using the training dataset and distance variation of Euclidean was implemented. Here, the k value is varied from 1 to 5. After that, the remaining 20% belongs to the testing dataset that were used to test the built k-NN network. The model will be accepted once it achieves the accuracy above than 80% (as standard practice in system identification [16]). However, if it is not, it will undergo data pre-processing again.

The k-NN model will separate agarwood oil samples to three groups extracted in the laboratory by using lab scale hydrodistillation (L), obtained from industrial set-up hydrodistillation in the factory directly from local manufacturer (LM) and Kuala Lumpur market (M). In mathematics, a distance implies the idea between two points that represent length between them. It allows the separation of similar and separate compounds. Usually, the ratio between training and testing is equal to 80%:20%. The step of the k-NN algorithm [19]–[21]:

Step 1: Information entropy is calculated,

Step 2: Find value k for training set,

Step 3: Training set split into group of clusters,

Step 4: Obtain mean of cluster so center of every cluster can be obtained,

Step 5: The cluster nearest to test sample is determined with Euclidean distance formula,

Step 6: Weighted Euclidean distance formula will calculate interval between sample in cluster and test sample. Then k-NN is formed,

Step 7: Class label for test sample is chosen from the highest probability.

Using k-NN method, the k' neighbours from 1 to 5 is used to classify the input to its nearest neighbours that is the compound in agarwood oil samples. The interval metric introduced by Euclidean has been reviewed and adjusted during the execution.

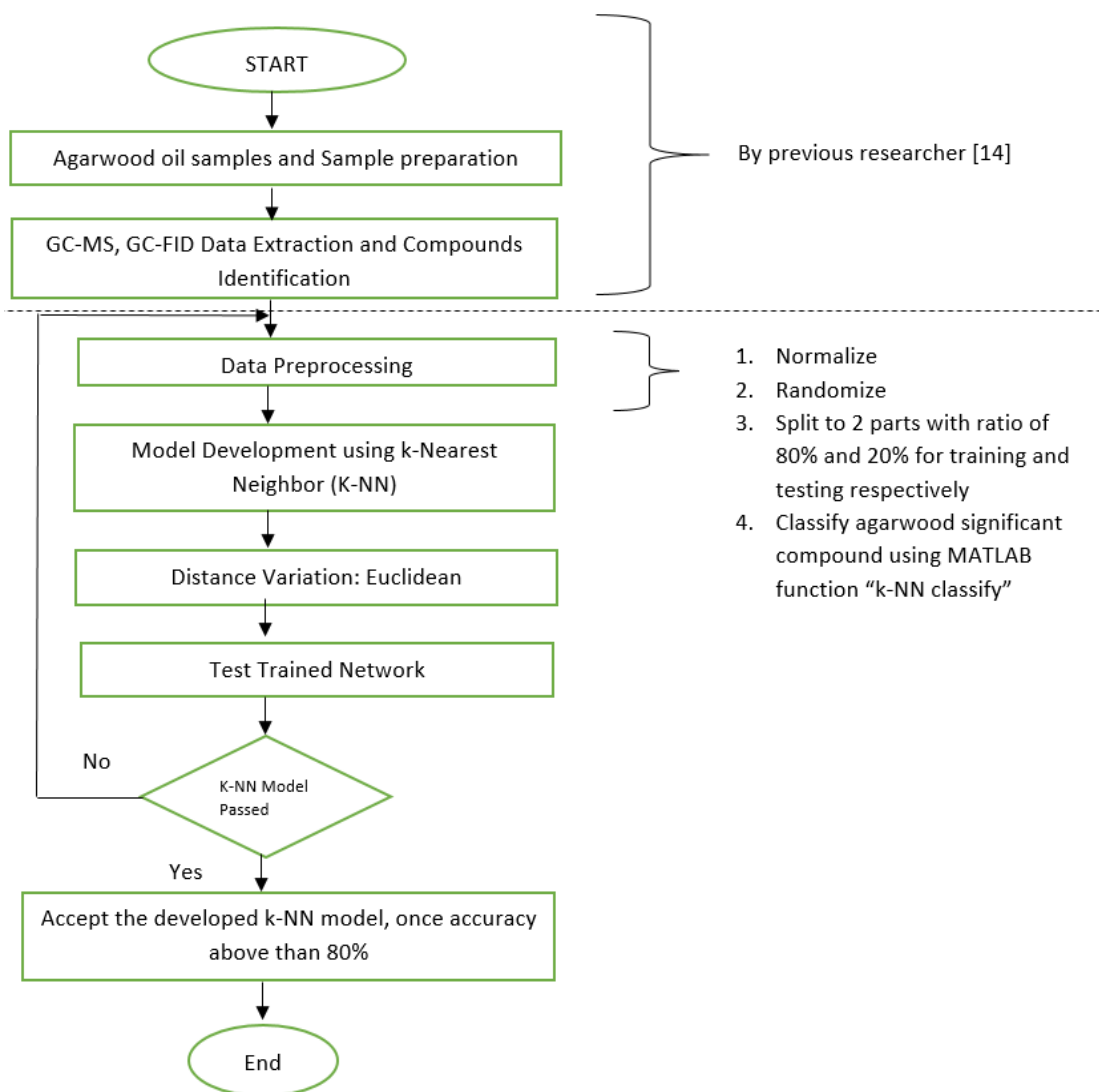


Figure 4. k-NN model development

3 RESULTS AND DISCUSSION

3.1 Agarwood Oil Extraction

In this study, a total of 86 agarwood samples were used that consists of seven samples from extracted in the laboratory by using lab scale hydrodistillation (L), 30 samples from industrial set-up hydrodistillation in factory directly from local manufacturer (LM) and 49 samples from Kuala Lumpur market (M). Figures 5 (a), 5(b) and 5(c) show the graphical observation for three groups of sample sources which are extracted in the laboratory by using lab scale hydrodistillation (L), obtained from industrial set-up hydrodistillation in factory directly from local manufacturer (LM) and Kuala Lumpur market (M), respectively. By taking the chemical composition (%) maximum to 30.00%, several observations can be made. Two of them are the range of chemical composition and the highest peak of chemical composition. For the first observation i.e., the range of chemical composition; the agarwood oil samples from lab (L) are between 0.00% to less than 10.00%, the agarwood oil samples from local manufacturer (LM) are between 0.00% to less than 13.00% and the agarwood oil samples from 0.00% to less than 28.00%. Here, it can be said that different sources of samples produced different peaks and amounts of composition (%) for the agarwood oil. The second observation is the highest peak for each group. It can be seen that for agarwood oil samples from lab (L), the highest peak belongs to compound C40 (hexadecanol) with a composition of 8.08%. For agarwood oil samples from local manufacturer (LM), the highest peak is belonging to compound C5 (aromadendrene) with the composition of 12.05% and for the agarwood samples from Kuala Lumpur market (M), the highest peak belongs to compound C14 (tridecanol) with the composition of 26.01%.

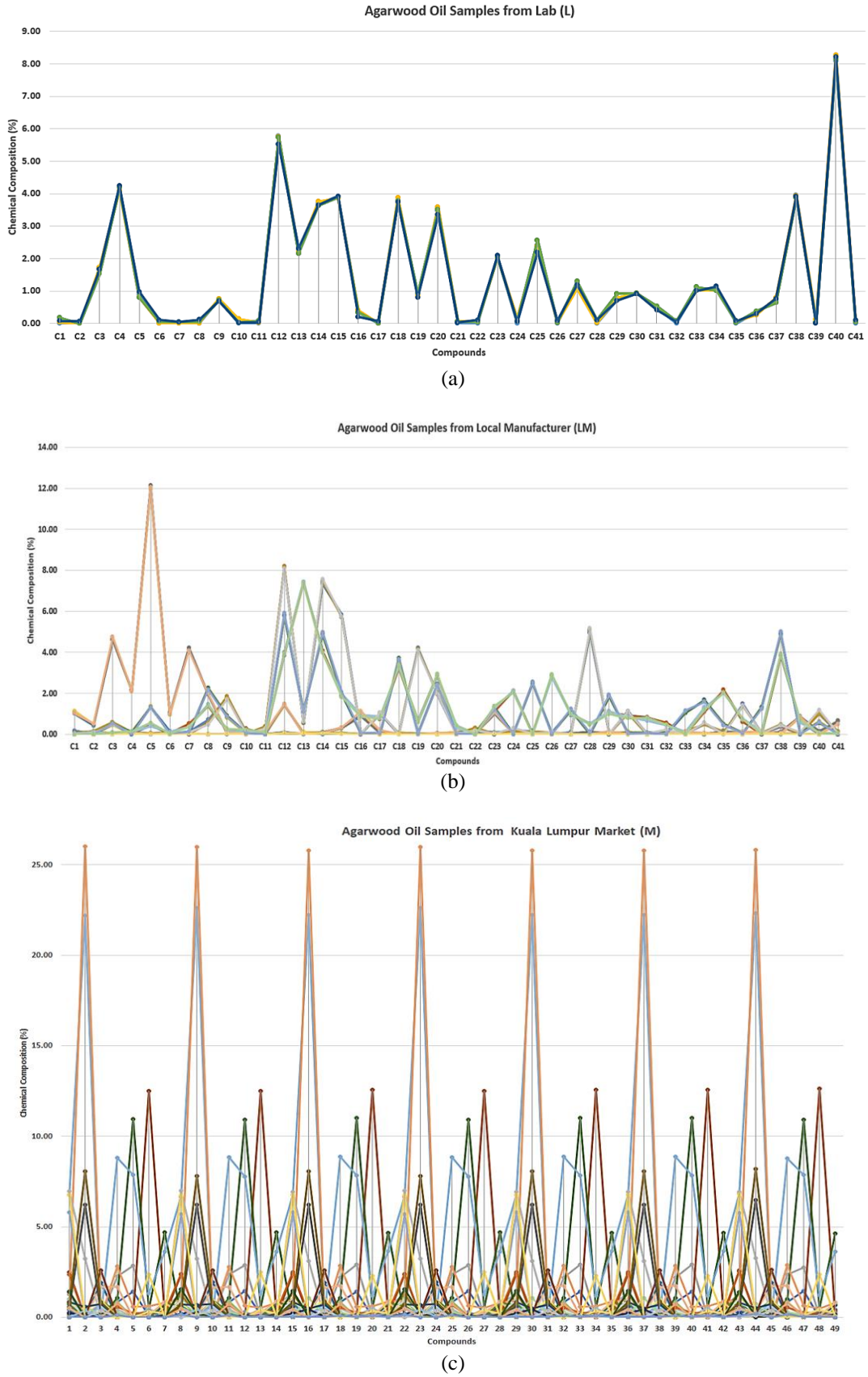


Figure 5. Graphical observation for three groups of sample sources (a) extracted in the laboratory by using lab scale hydrodistillation (L), (b) industrial set-up hydrodistillation in factory directly from local manufacturer (LM), and (c) Kuala Lumpur market (M)

Table 1 tabulated the accuracy of k-NN modelling (training and testing dataset) from k=1 to k=5. In general, the accuracy obtained is from 50.00% to 100.00%. At k=1 and k=2, the training and testing accuracies are similar which is 98.63% and 100.00%, respectively. At k=3, the training accuracy is 80.82% and testing accuracy is 83.33%. At k=4, the training accuracy is 72.60% and the testing accuracy is 72.22%. Lastly, at k=5, the training accuracy is 63.01% and the testing accuracy is 50.00%. It was found that at k=1 and k=2, the developed k-NN model obtained the highest accuracy among all with the training is 98.63% and testing is 100.00%. Thus, it can be said that, at k=1 and k=2, the developed k-NN model able to differentiate the agarwood oil samples to three groups; extracted in the laboratory by using lab scale hydrodistillation (L), industrial set-up hydrodistillation in factory directly from local manufacturer (LM) and Kuala Lumpur market (M), successfully. The result also is supported with the graphical modelling as shown in Figure 6.

Table 1. Accuracy of k-NN modelling (training and testing dataset) from k=1 to k=5

k-Value	Training (%)	Testing (%)
1	98.63	100.00
2	98.63	100.00
3	80.82	83.33
4	72.60	72.22
5	63.01	50.00

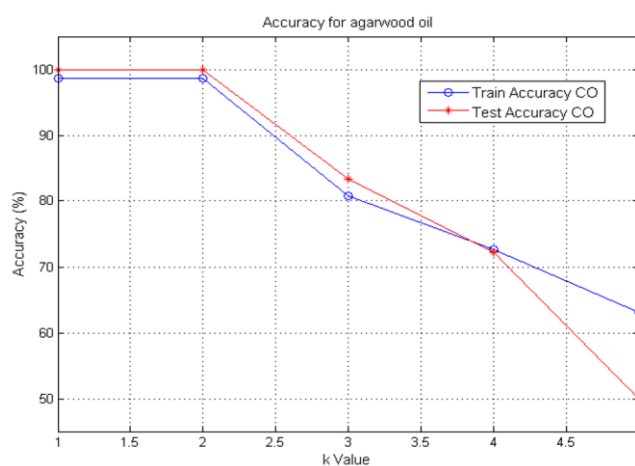


Figure 6. Graphical k-NN modelling for accuracy of training and testing for k=1 to k=5

4. CONCLUSION

The k-NN modelling of agarwood oil samples available in the capital of Malaysia market was successfully presented in this paper. The results revealed the capability of the k-NN model with variation of its k-values from 1 to 5. Among the k-values, k=1 and k=2 obtained similar accuracy by having 98.63% for training dataset and 100.00% for testing dataset. The finding in this study is important as it contributed to the agarwood oil samples from different sources especially to grading related research areas. For future work, it is recommended that more samples should be analyzed with the k-value above than 5 and their performances can be observed, together with other performance criteria such as confusion matrix.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank to Bioaromatic Research Centre, Universiti Malaysia Pahang (UMP) for providing the facilities. This research was supported by a grant from MTUN, Ministry of Higher Education (MOHE) Malaysia, under grant titled: Development of High Quality Agarwood (Gaharu) Formation based on Effective Inoculant Technology – RDU192814/UIC190820).




REFERENCES

- [1] Y. Z. H.-Y. Hashim, P. G. Kerr, P. Abbas, and H. Mohd Salleh, "Aquilaria spp. (agarwood) as source of health beneficial compounds: A review of traditional use, phytochemistry and pharmacology," *Journal of Ethnopharmacology*, vol. 189, pp. 331–360, Aug. 2016, doi: 10.1016/j.jep.2016.06.055.
- [2] B. Tohidi, M. Rahimmalek, and A. Arzani, "Essential oil composition, total phenolic, flavonoid contents, and antioxidant activity of k-nearest neighbor modelling of agarwood oil samples available in capital ... (Erny Haslina Abd Latib)




- Thymus species collected from different regions of Iran,” *Food Chemistry*, vol. 220, pp. 153–161, Apr. 2017, doi: 10.1016/j.foodchem.2016.09.203.
- [3] K. Jayachandran, I. Sekar, K. T. Parthiban, A. Damodarasamy, and K. K. Suresh, “Analysis of different grades of agarwood (*Aquilaria malaccensis* Lamk.) oil through GC-MS,” *Indian Journal of Natural Products and Resources*, vol. 5, no. 1, pp. 44–47, 2014.
- [4] C. S. Tan, N. M. Isa, I. Ismail, and Z. Zainal, “Agarwood induction: current developments and future perspectives,” *Frontiers in Plant Science*, vol. 10, Feb. 2019, doi: 10.3389/fpls.2019.00122.
- [5] M. Ishihara, T. Tsuneyama, and K. Uneyama, “Fragrant sesquiterpenes from agarwood,” *Phytochemistry*, vol. 33, no. 5, pp. 1147–1155, Jul. 1993, doi: 10.1016/0031-9422(93)85039-T.
- [6] A. Hidayat *et al.*, “Bioactive composition, antifungal, antioxidant, and anticancer potential of agarwood essential oil from decaying logs (*Gyrinops* spp.) of Papua Island (Indonesia),” *Journal of Applied Pharmaceutical Science*, vol. 11, pp. 70–78, 2021, doi: 10.7324/JAPS.2021.1101010.
- [7] A. López-Sampson and T. Page, “History of use and trade of agarwood,” *Economic Botany*, vol. 72, no. 1, pp. 107–129, Mar. 2018, doi: 10.1007/s12231-018-9408-4.
- [8] Y. Liu *et al.*, “Whole-tree agarwood-inducing technique: an efficient novel technique for producing High-quality agarwood in cultivated *Aquilaria sinensis* trees,” *Molecules*, vol. 18, no. 3, pp. 3086–3106, Mar. 2013, doi: 10.3390/molecules18033086.
- [9] 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,” in *2015 IEEE 6th Control and System Graduate Research Colloquium (ICSGRC)*, Aug. 2015, pp. 103–108, doi: 10.1109/ICSGRC.2015.7412473.
- [10] M. Dash, J. K. Patra, and P. P. Panda, “Phytochemical and antimicrobial screening of extracts of *Aquilaria agallocha* Roxb.,” *African Journal of Biotechnology*, vol. 7, no. 20, pp. 3531–3534, 2008, doi: 10.5897/AJB08.623.
- [11] D. T. Ahmaed and A. D. Kulkarni, “Sesquiterpenes and chromones of agarwood: a review,” *Malaysian Journal of Chemistry*, vol. 19, no. 1, pp. 33–58, 2017.
- [12] E. Haslina, M. S. Najib, C. M. A. C. Mohd, and S. N. Tajuddin, “Analysis of different quality agarwood oil (*aquilaria malaccensis*) and sensory study,” *Journal of Telecommunication*, vol. 10, pp. 1–3, 2018.
- [13] P. Pripdeevech, W. Khummueng, and S.-K. Park, “Identification of odor-active components of agarwood essential oils from Thailand by solid phase microextraction-GC/MS and GC-O,” *Journal of Essential Oil Research*, vol. 23, no. 4, pp. 46–53, Jul. 2011, doi: 10.1080/10412905.2011.9700468.
- [14] N. A. M. Ali *et al.*, “Comparison of chemical profiles of selected gaharu oils from peninsular Malaysia,” *The Malaysian Journal of Analytical Sciences*, vol. 12, no. 2, pp. 338–340, 2008.
- [15] S. N. Tajuddin, N. S. Muhamad, M. A. Yarmo, and M. M. Yusoff, “Characterization of the chemical constituents of agarwood oils from Malaysia by comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry,” *Mendeleviev Communications*, vol. 23, no. 1, pp. 51–52, Jan. 2013, doi: 10.1016/j.mencom.2013.01.019.
- [16] N. Ismail, N. A. M. Ali, M. Jamil, M. H. F. Rahiman, M. N. Taib, and S. N. Tajuddin, “Major volatile chemical compounds of agarwood oils from Malaysia based on Z-score technique,” *Chemistry of Natural Compounds*, vol. 51, no. 4, pp. 776–779, Jul. 2015, doi: 10.1007/s10600-015-1409-2.
- [17] S. Ismail *et al.*, “Discriminative analysis of different grades of gaharu (*Aquilaria malaccensis* Lamk.) via 1H-NMR-based metabolomics using PLS-DA and random forests classification models,” *Molecules*, vol. 22, no. 10, p. 1612, Sep. 2017, doi: 10.3390/molecules22101612.
- [18] S. Lias, N. A. M. Ali, M. Jamil, M. H. Zainal, and S. H. Ab Ghani, “Classification of pure and mixture agarwood oils by electronic nose and discriminant factorial analysis (DFA),” in *2015 International Conference on Smart Sensors and Application (ICSSA)*, May 2015, pp. 7–10, doi: 10.1109/ICSSA.2015.7322500.
- [19] M. M. R. Khan, R. B. Arif, M. A. B. Siddique, and M. R. Oishe, “Study and observation of the variation of accuracies of KNN, SVM, LMNN, ENN algorithms on eleven different datasets from UCI machine learning repository,” in *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEICT)*, Sep. 2018, pp. 124–129, doi: 10.1109/CEEICT.2018.8628041.
- [20] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient kNN classification with different numbers of nearest neighbors,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774–1785, May 2018, doi: 10.1109/TNNLS.2017.2673241.
- [21] A. Kataria and M. Singh, “A review of data classification using K-nearest neighbour algorithm,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 6, pp. 354–360, 2013.
- [22] L. M. Zouhal and T. Denoeux, “An evidence-theoretic k-NN rule with parameter optimization,” *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 2, pp. 263–271, May 1998, doi: 10.1109/5326.669565.
- [23] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, 2009, Art. no. 1883, doi: 10.4249/scholarpedia.1883.
- [24] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, “When Is ‘Nearest Neighbor’ Meaningful?,” in *International conference on database theory*, 1999, pp. 217–235.
- [25] K. Yu, L. Ji, and X. Zhang, “Kernel nearest-neighbor algorithm,” *Neural Processing Letters*, no. 15, pp. 147–156, 2002, doi: 10.1023/A:1015244902967.

BIOGRAPHIES OF AUTHORS






Erny Haslina Binti Abd Latib    received BSc. (Analytical and Environmental Chemistry) in 2009 from Universiti Malaysia Terengganu and Master of Environment (Water Chemistry) from Universiti Putra Malaysia in 2011. She is currently undergoing a postgraduate program for her Ph.D. (Industrial Chemistry) in Universiti Malaysia Pahang. Her research interest in the field of analytical chemistry that focuses on authenticity analysis of agarwood oil. She is presented in various academic as well as research-based papers at several national and international conferences including the 4th International Conferences on Electrical, Control & Computer Engineering 2017 – Langkawi, Malaysia; 25th International Invention, Innovation & Technology Exhibition ITEX 2014 – Kuala Lumpur, Malaysia. ChM. Erny Haslina Binti Abd Latib is also endorsed as a registered chemist from Malaysian Institute of Chemistry in 2019. She can be contacted at email: ernyhaslina87@gmail.com.






Nurlaila Ismail    is a senior lecturer at School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor. She obtained her B.Sc., M.Sc., and Ph.D. in Electrical Engineering from UiTM Shah Alam. She joined UiTM during her postdoctoral service in 2016. She is a professional engineer in discipline of teaching recognized by Board of Engineer Malaysia (BEM), an active member in several organizations including IEEE Malaysia especially Control System Society, Malaysian Society for Computed Tomography and Imaging Technology (MyCT) and Institute of Engineer Malaysia (IEM) and Malaysia Board of Technologists (MBOT) as well as ASEAN Engineering Register (AER). She has published more than fifty technical papers, locally and at international level. Her research interests are in advanced signal processing and artificial intelligence. She can be contacted at email: nurlaila0583@uitm.edu.my.






Saiful Nizam Tajuddin    is a senior lecturer at Faculty of Industrial Science & Technology (FIST). He has worked at Universiti Malaysia Pahang since 2005 and was one the members who first founded FIST Faculty in 2007. Later in 2013, he established the BioAromatic Research Center (BIOAROMATIK) and is currently appointed as Director of COE. His main research area is on natural products, focusing on aromatic plants, agarwood (locally known as Gaharu), spanning the upstream to downstream development. His research projects include development of methods for extraction of essential oil, and isolation and purification of complex mixtures of terpene compounds that are predominantly present in the essential oil. Chiral separation technique for separation of complex mixtures of isomers is one of the most integral components in these projects. These studies comprise various separation techniques using chromatography and spectroscopy. The application from these fundamental works, can be applied in flavor, fragrance, pheromone, and pharma industries. He can be contacted at email: saifulnizam@ump.edu.my,



Jasmin Jamil    is a founder of Discovery Agarwood Research. His company is based in Pulau Pinang, provides all those businesses and research related to agarwood covering plantation, inoculant, extraction, and marketing. He can be contacted at email: discovery.agarwood@gmail.com.



Zakiah Mohd Yusoff    is a senior lecturer who is currently working at UITM Pasir Gudang. She received the B. ENG in Electrical Engineering and PhD in Electrical Engineering from UITM Shah Alam, in 2009 and 2014, respectively. In Mei 2014, she joined UITM Pasir Gudang as a teaching staff. Her major interests include process control, system identification, and essential oil extraction system. She can be contacted at email: zakiah9018@uitm.edu.my.