Malaysian Journal of Chemistry, 2024, Vol. 26(1), 324-331
(Special Issue: *International Conference on Natural Products, ICNP 2023*)

Agarwood Oil Grade Clustering of *Aquilaria malaccensis* Species using Extraction by GC-MS Analysis: Efficient KNN Algorithm Based on Patterns Visualization of Two-dimensional Graph

Siti Mariatul Hazwa Mohd Huzir[1], Anis Hazirah 'Izzati Hasnu Al-Hadi[2], Amir Hussairi Zaidi[2], Nurlaila Ismail[2,*], Zakiah Mohd Yusoff[1], Saiful Nizam Tajuddin[3] and Mohd Nasir Taib[2]

[1]Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Johor, Kampus Pasir Gudang, Malaysia
[2]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
[3]Bioaromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang (UMP), Gambang, Pahang, Malaysia
*Corresponding author (e-mail: nurlaila0583@uitm.edu.my)

Data visualization pattern is an essential task in data analysis. A two-dimensional graph (2D graph) is one of the graphical presentations for data visualization. Over the past decades, Agarwood Oil grade clustering is still at a disadvantage since there is no official standard grading system. Most of the time, an expert grades the agarwood oil manually based on oil appearances such as resin color, smell, texture and intensity. The importance of the agarwood oil grading system will help the seller to stabilize the oil price based on its approximate quality. Besides, Agarwood oil got high requests from big buyers and traders due to its benefits as medicine, cosmetics, perfume and incense. This paper attempts to formulate a better Agarwood oil grading system based on its chemical properties, develops an artificially intelligent k-Nearest Neighbor (KNN) and trained using Matlab version R2015a. The data acquisition process of investigating the chemical compounds was conducted using GC-MS analysis. From 103 chemical compounds extracted, four significant compounds; 10-epi-ɤ-eudesmol, α-agarofuran, ɤ-eudesmol and β-agarofuran were chosen to model the agarwood oil quality. The agarwood oil sample data were categorized into low, medium-low, medium-high and high grades. The findings show that KNN yielded 100% accuracy. Then, 2D graph was applied to plot the sample visualization pattern parallel with KNN accuracy. The KNN 2D plot revealed a distinct separation between the four groups. The accuracy of 100% proved the potential of the KNN model as a good supervised learning classifier towards four different grades of Agarwood oil. In conclusion, the Agarwood oil quality grading technique based on KNN and 2D graph was successful with the ability of KNN to confirm these qualities into 4 grades.

Keywords: Artificial intelligent KNN; pattern visualization; agarwood quality; grading technique

Received: September 2023 ; Accepted: November 2023

Agarwood oil holds immense value as a natural product with valuable applications in various areas such as incense, fragrances, shampoos, traditional medicine for treating stomach complaints, diarrhea, lung and liver pain, as well as in perfumery (especially for darker hues) [1-6]. Notably, the most expensive oil on the market is derived from *Aquilaria malaccensis* species infected heartwood. The volatile aromatic compounds concentrated in agarwood oil are also one of the factors of consumer attraction [7-10]. Agarwood oil's popularity has recently risen with market statistics revealing a significant increase in its purchase in Middle Eastern countries (United Arab Emirates, Saudi Arabia), China and Japan [9]. Malaysia also experienced an increase in demand for agarwood oil, prompting a rapid expansion of agarwood plantations across the country including Perak, Terengganu, Kelantan and Pahang [11, 12].

Currently, manual sensory evaluation based on physical properties is used to assess and grade essential oil quality. The presence of high-content resins, a dark oil color, a strong odor and a long-lasting aroma are all indicators of the highest grade of essential oil. Fragrant agarwood resin is formed as a result of complex biotic, abiotic and physical stress on *Aquilaria* trees [13-15]. However, due to different perceptions and decisions among different individuals, this subjective sensory evaluation method suffers from inaccuracies. As a result, there is no guarantee that this human-based grading process can ensure essential oil purity or overall quality [16-18]. Innovative technologies have emerged to address the limitations of traditional grading systems with the goal of improving the stability and availability of essential oils [4, 18-22]. Modern quality assessment and grading systems have been developed, including novel indexes that include

†Paper presented at the International Conference on Natural Products 2023 (ICNP 2023)

both qualitative and quantitative analysis. The scientific method provides an alternative solution to the grading problem. Advances in data analysis have paved the way for platforms that use intelligent methods to classify agarwood oil quality solely based on its chemical profiles [23-27]. This enables essential oils to be accurately classified into the various classes which range from low to medium-low, medium-high, or high quality.

Several researchers have proposed machine learning techniques to verify the quality of agarwood oil, including artificial neural networks (ANN), linear regression, k-Nearest Neighbour (KNN), self-organizing map (SOM) and OVO multiclass support vector machine (SVM) [23-26]. These cutting-edge methods enable more reliable and precise measurements of essential oil quality, paving the way for a more objective and efficient grading process. The correlation coefficient, R, achieves a perfect value of 1 when the number of hidden neurons is 2 according to research on the seven chemical compounds present in agarwood oil extracted from GC-MS analysis (β-agarofuran, 10-epi-ɤ-eudesmol, ɤ-Eudesmol, eudesmol, hexadecanol, α-agarofuran and longifolol) [24]. This ANN configuration outperforms others with the lowest mean squared error (MSE) value of 7.69×10^{-15} when compared to other neurons. Another literature review on the SOM model was used to identify three significant chemical compounds in agarwood oil: α-agarofuran, β-agarofuran, and 10-epi-ɤ-eudesmol [26]. Notably, previous studies on the quality of agarwood oil has typically only included two categories, namely low and high quality.

Hence, this study emphasises the use of k-Nearest Neighbor (KNN) classification as the primary model for categorising agarwood oil into four distinct qualities (low, medium low, medium high, and high quality), as recommended [25]. The k-Nearest Neighbor (KNN) approach has the advantage of increasing the efficacy of agarwood grading based on the most significant chemical compounds and producing high percentage of accuracy for four quality categories. Furthermore, in terms of efficiency, the intelligent model outperforms traditional methods by reducing time-consuming processes and ensuring greater consistency in the classification process. This study suggests the use of KNN in grading the quality of agarwood essential oil into high, medium-high, medium-low and low quality as a good and future classifier.

EXPERIMENTAL METHODOLOGY

Data Acquisition

The dataset of agarwood oil used in this research was collected by the Forest Research Institute Malaysia (FRIM) in collaboration with the Bioaromatic Research Centre of Excellence (BARCE) and the Universiti Malaysia Pahang (UMP) team [26]. The raw data was gathered by measuring the physical properties of real agarwood oil and converting the findings into digital quantitative values suitable for computer-based analysis. This procedure involved examining the agarwood oil samples and chemical constituents using GC-MS analysis, which provides information about the relative abundance of each compound. It included a total of 660 samples representing four different qualities; low, medium low, medium high and high qualities and 103 extracted chemical compounds. Among these samples, 210 belonged to the low quality category, 90 to medium low qualities, 30 to medium high qualities and 330 to high quality classes as indicated in Table 1. The dataset consisted of four significant chemical compounds selected from preprocessing technique were 10-epi-ɤ-eudesmol, α-agarofuran, ɤ-eudesmol and β-agarofuran. To conduct the simulations, MATLAB software version R2015a was utilized.

Table 1. The number of samples for four different qualities.

Grades	The number of samples	The eleven chemical compounds
Low	210	10-epi-ɤ-eudesmol, α-agarofuran, ɤ-eudesmol and β-agarofuran
Medium-low	90	
Medium-high	30	
High	330	

The Gas Chromatography-mass spectrometry (GC-MS) analysis was employed to conduct chemical analysis and carry out the extraction process for identifying the chemical constituents of agarwood oil. Agilent Technologies' 7890A/5975C Series Mass Spectrometry Detector (MSD) with an HP-5MS column (30 meters in length, 0.25 mm inner diameter and 0.25-micron film thickness) was employed for this purpose. The program initiated at 60 degrees Celsius for a duration of 10 minutes and then gradually

increased to 180 degrees Celsius within 1 minute, at a rate of 3 degrees Celsius per minute. Helium was utilized as the carrier gas, flowing at a rate of 1.0 ml/min and the temperature was maintained at 280 degrees Celsius. To identify the chemical constituents, mass spectral libraries (HPCH2205.L, Wiley7Nist05.L and NIST05a.L) were employed. The results were expressed in terms of peak counts to represent the peak areas.

Framework of Experimental Set-up

The experiment commences with loaded the input and output data into the Matlab workspace, as indicated in the framework presented in Figure 2. Subsequently, the input data for this study comprised the percentage of abundance of eleven significant chemical compounds present in agarwood oil, whereas the output data corresponded to the grades assigned to each of the data samples. Afterward, the data was split into two groups: one for the KNN training phase and the other for the KNN testing process. The training dataset consisted of 80% of the data, while the remaining 20% was reserved for testing purposes. The k-value is set from 1 to 10 and Euclidean (EU) formula is the distance metric for KNN algorithm. For the performance of the KNN classification model, the performance measures by examining its accuracy, sensitivity, specificity and precision in classifying the test data and then comparing it with other classification models to determine its effectiveness. If the accuracy not less than 80%, it is evaluated to have good high accuracy towards agarwood oil qualities classification. Additionally, the KNN algorithm were accepted and consider successful if the model passes all the performance criteria of accuracy, sensitivity, precision and specificity. Equation (1) show the proposed calculation for the Euclidean distance metric [28].

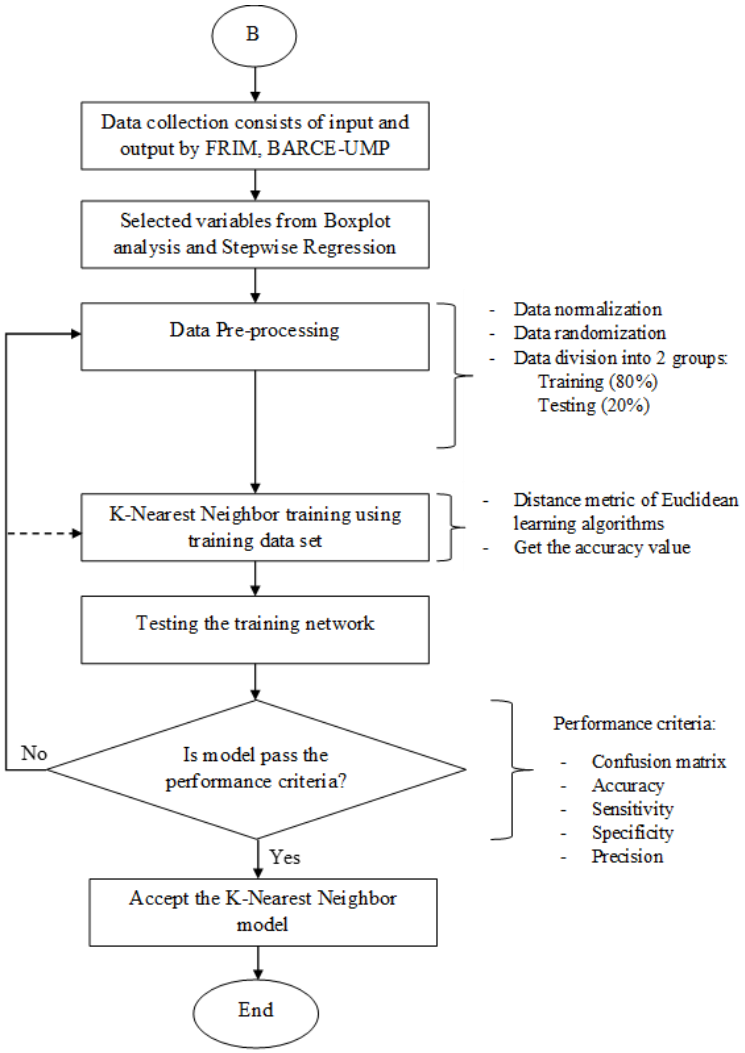

Figure 2. The structure of the KNN classification model development framework.

$$d_{st} = \sqrt{\sum_{j=1}^{n} (x_{sj} - y_{tj})^2} \quad (1)$$

Where,
x_{sj} = an object at coordinate sj
y_{tj} = another object at coordinate tj
d_{st} = distance between both of the coordinates

Table 2. The comparison in relative abundance of chemical compounds.

Quality groups	Relative abundance groups
Low	CKE, EO4, LG, M, MN, MS and R5 group
Medium-low	CM, EO3 and HD group
Medium-high	EO2 group
High	HG, JBD, KB, LA, MA, MA1, MA2, MNS, MPE, RG and T group

RESULTS AND DISCUSSION

Evaluating the Relative Abundance of Chemical Compounds in Agarwood Oil Samples

The GC-MS analysis results revealed that the collected data could not be directly utilized for the study's intended purpose. Therefore, it became essential to preprocess the data in order to acquire a comprehensive and suitable dataset. For a better understanding, the relative abundance of chemical compounds was analyzed to identify trends and patterns of the oil's quality. The boxplots result displayed the data distribution and spread for each class of compounds, offering valuable information regarding the abundance levels and variations across various samples. By doing so, the extracted compositions of agarwood oil such as HG, JBD, KB, LA, MA, MA1, MA2, MNS, MPE, RG, and T group, were identified for the high-quality standard as tabulate in Table 2. The other quality groups of samples, including medium-high (EO2), medium-low (CM, EO3, and HD) and low-quality (CKE, EO4, LG, M, MN, MS and R5) samples from the group.

The redata samples were then evaluated and developed into a KNN classification model based on its performance across four grades. In addition, scatter plots of 2D graphs were created to visualize the patterns based on the KNN results. The following steps involve summarizing and discussing the evaluation process's findings in order to assess the model's effectiveness and success.

KNN Classification Model and Performance Measure of Four Grades

The KNN algorithm was built using Agarwood oil sample data from four different grades. The data was distributed as follows: grade low contributed for 31.82% (210 samples), grade medium-low represented for 13.64% (90 samples), grade medium high represented for 4.55% (30 samples) and grade high contributed 50% (330 samples). Figure 3 and Figure 4 shows the results of the Euclidean confusion matrix for the KNN model, demonstrating that all predicted data precisely align with the actual data, with no misclassifications occurring during the model's testing phase.

The KNN model performed well on the confusion matrix. The standard calculation was used to calculate the accuracy in model effectiveness, sensitivity in quality classification, specificity in making comparisons and precision in clustering by grade for Euclidean distance, as described in the methodology section. The results of the model's performance evaluation are shown in Table 3. By achieving perfect accuracy, sensitivity, specificity, and precision during testing, no errors were found in this model.

The 2D Graph of Four Grades of Agarwood Oil

Figure 5 shows a 2D graph of four qualities of Agarwood oil. This 2D graph's analysis is limited to data from low and medium high grades. As shown in the graph, there is a disparity between these two data grades. While low grade data appears scattered, all medium high grade data points are plotted at the same position for y-axis with abundances of 0.01. Despite this disparity in data distribution, the gap is visible in the graph. When compared to other grades, all of the data from high grade shows the balance of percentages for the significant chemical compounds. This gap benefits the KNN model in accurately classifying the data into its respective grades.

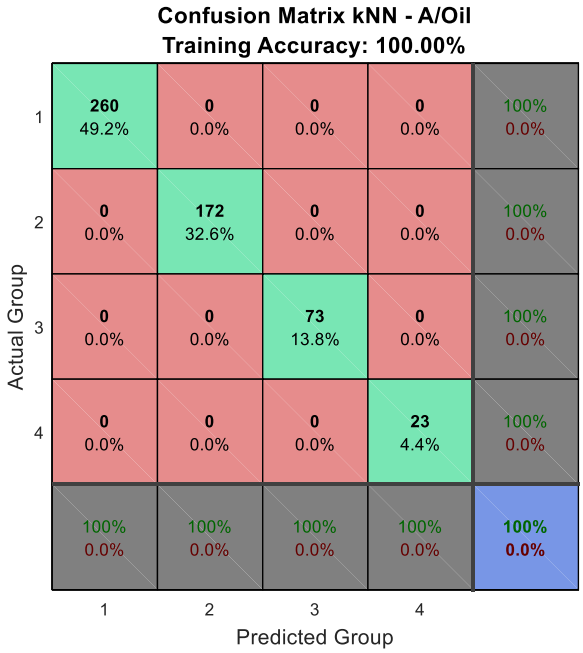

Figure 3. Confusion matrix of four different grades of Agarwood oil data for training dataset.

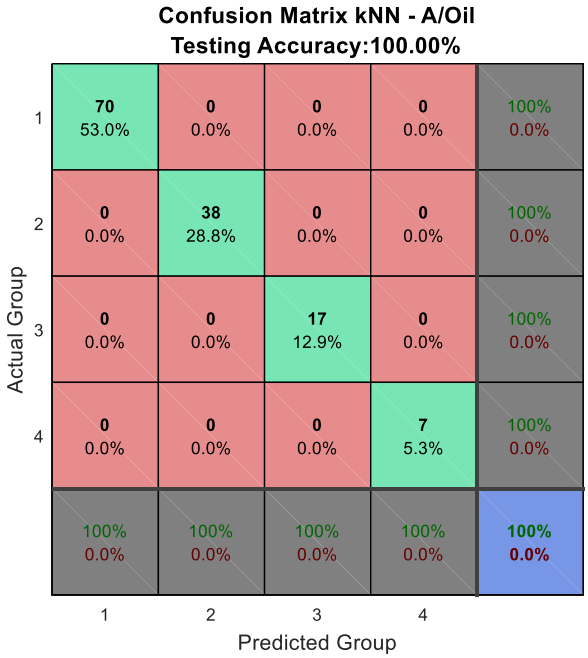

Figure 4. Confusion matrix of four different grades of Agarwood oil data for testing dataset.

Table 3. Result of performance for four grades KNN model.

Performance Criteria	Percentage (%)
Accuracy	100
Sensitivity	100
Specificity	100
Precision	100

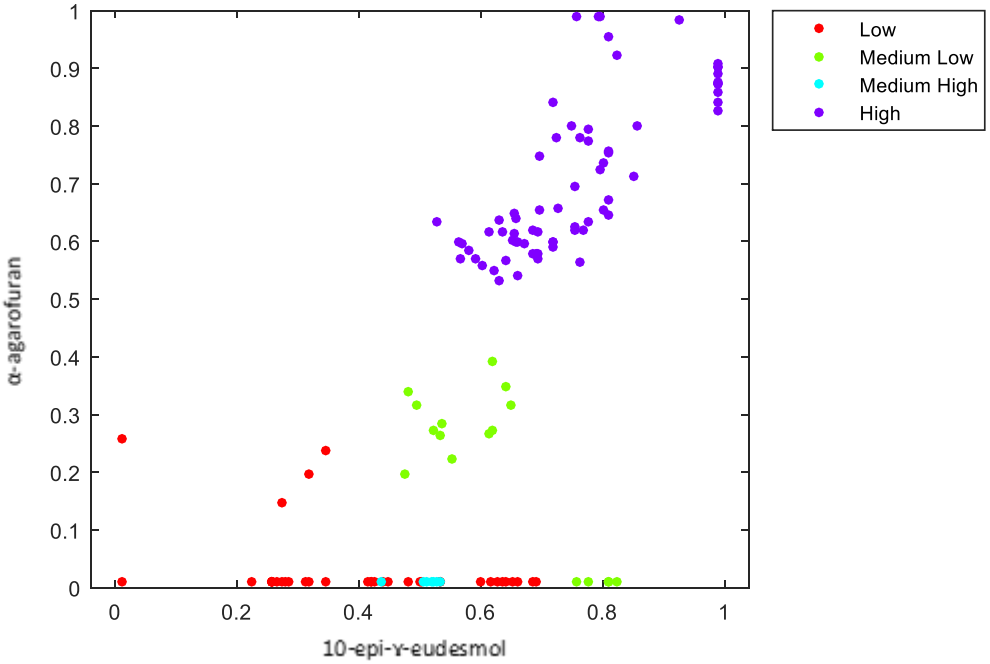

Figure 5. The 2D graph of four different grades of Agarwood oil.

CONCLUSION

The results of the KNN classification model development by utilizing all four significant chemical components as inputs were thoroughly discussed. Based on these findings, the KNN model successfully classified and validate the grade of Agarwood oil into four qualities: low, medium low, medium high and high. The 4 × 4 confusion matrix table was used to classify the four grades prior to conducting the performance measure. Both testing and training confusion matrix showed no misclassifications, indicating that the predicted and actual classification processes were correctly aligned during testing. From four grades, this success was further validated by achieving 100% accuracy, specificity, sensitivity and precision for ten nearest neighbors (k = 10) because of zero misinterpretation during the model testing process. Furthermore, the pattern visualization results based on the 2D plot graph supported the efficacy and validity of this new grading approach, reaffirming its suitability for future implementation. The findings of this paper support the KNN classification model's effectiveness in precisely grading agarwood oil into four distinct grades. The comparison in relative abundance of chemical compounds also identified that HG, JBD, KB, LA, MA, MA1, MA2, MNS, MPE, RG, and T group for the high-quality standard, EO2 for medium-high, CM, EO3, and HD group for medium-low and CKE, EO4, LG, M, MN, MS and R5 group for low quality. Future research should look into the classification of agarwood from various species including those from other producing countries, in order to improve the novelty of agarwood oil classification and advance its chemical analysis. In conclusion, the standardized agarwood oil classification model developed in this study holds great promise for global adoption and implementation.

ACKNOWLEDGEMENTS

The authors would like to express the gratitude to School of Electrical Engineering, UiTM Shah Alam, Selangor and excellent fund from Universiti Teknologi

MARA (UiTM), Cawangan Johor for providing the financial support named GERAN BESTARI FASA 1/2023, 600-TNCPI 5/3/DDN (01) (001/2023) throughout this research.

REFERENCES

1. Yadav, V. K., Choudhary, N., Heena Khan, S., Khayal, A., Ravi, R. K., Kumar, P., Modi, S. and Gnanamoorthy, G. (2020) Incense and incense sticks: types, components, origin and their religious beliefs and importance among different religions. *J. Bio. Innov*., **9**, 1420–1439.

2. Naziz, P. S., Das, R. and Sen, S. (2019) The scent of stress: Evidence from the unique fragrance of agarwood. *Frontiers in plant science*, **10**, 840.

3. Shivanand, P., Arbie, N. F., Krishnamoorthy, S. and Ahmad, N. (2022) Agarwood-The Fragrant Molecules of a Wounded Tree. *Molecules*, **27(11)**, 3386.

4. Sun, W., Wu, S., Zhu, P., Deng, S., Zhou, X., Liu, X., Zhao, J., Zhang, W., Gao, X. and Chen, X. (2022) Rapid Discrimination of Aroma Components in Natural and Artificial Agarwood Incense by Static Headspace Sampling–Gas Chromatography–Mass Spectrometry (SHS-GC-MS) with Principal Component Analysis (PCA) and Partial Least Squares–Discrimination Analysis (PLS-DA). *Analytical Letters*, **55(18)**, 2829–2845.

5. Battaglia, S. (2021) Agarwood/Oud.

6. Cheung, S. C. (2020) The sublime in scent: a comparative study of Japanese Kodo and Chinese incense tradition in the 21st century. *Asian Education and Development Studies*, **10(1)**, 95–104.

7. Tan, C. S., Isa N. M., Ismail, I. and Zainal, Z. (2019) Agarwood induction: current developments and future perspectives. *Frontiers in plant science*, **10**, 122.

8. Syazwan, S. A., Lee S. Y., Ong, S. P. and Mohamed, R. (2019) Damaging insect pests and diseases and their threats to agarwood tree plantations. *Sains Malaysiana*, **48(3)**, 497–507.

9. Sumadiwangsa, S. (2020) The enhancement of added value of agarwood by diversification product. In *IOP Conference Series: Earth and Environmental Science*, *IOP Publishing*.

10. Buniyamin, I., Akhir, R. M., Asli, N. A., Khusaimi, Z. and Mahmood, M. R. (2021) Biosynthesis of SnO2 nanoparticles by aqueous leaves extract of Aquilaria malaccensis (agarwood). In *IOP Conference Series: Materials Science and Engineering*, *IOP Publishing*.

11. Lee, S. L., Zakaria, N. -F., Tnah, L. H., Ng, C. H., Ng, K. K.,S., Lee, C. T., Lau, K. H. and Chua, L. S. L. (2022) DNA databases of a CITES listed species Aquilaria malaccensis (Thymelaeaceae) as the tracking tools for forensic identification and chain of custody certification. *Forensic Science International: Genetics*, **57**, 102658.

12. Hamzah, H. N. and Mohammad, A. (2021) A study of wild mushrooms utilised for food and medicine by the orang asli communities of Sungai Berua, Terengganu. *Resource Use and Sustainability of Orang Asli: Indigenous Communities in Peninsular Malaysia*, 75–84.

13. Wang, Y., Hussain, M., Jiang, Z., Wang, Z., Gao, J., Ye, F., Mao, R. and Li, H. (2021) Aquilaria species (Thymelaeaceae) distribution, volatile and non-volatile phytochemicals, pharmacological uses, agarwood grading system, and induction methods. *Molecules*, **26(24)**, 7708.

14. Azren, P. D., Lee, S. Y., Emang, D. and Mohamed, R. (2019) History and perspectives of induction technology for agarwood production from cultivated Aquilaria in Asia: a review. *Journal of Forestry Research*, **30(1)**, 1–11.

15. Mohamad, M., Najib, M. S., Daud, S. M., Zahed, N., Zahari, M. F., Majid, N. F.,H., Zaib, S. and Manap, H. (2020) Classification of agarwood types (Malaccensis and Crassna) between oil and smoke using e-nose with CBR Classifier. in *InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019*. *Springer*.

16. Mahabob, N. Z., Yusoff, Z. M., Ismail, N. and Taib, M. N. (2020) Preliminary Study on Classification of Cymbopogon Nardus Essential Oil using Support Vector Machine (SVM). In *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, *IEEE*.

17. Hu, Y., Kong, W., Yang, X., Xie, L., Wen, J. and Yang, M. (2014) GC–MS combined with chemometric techniques for the quality control and original discrimination of C urcumae longae rhizome: Analysis of essential oils. *Journal of separation science*, **37(4)**, 404–411.

18. Gerhardt, N., Schwolow, S., Rohn, S., Pérez-Cacho, P. R., Galán-Soldevilla, H., Arce, L. and Weller, P. (2019) Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: Comparison of different processing approaches by LDA, kNN, and SVM. *Food Chemistry*, **278**, 720–728.

19. Wangiyana, I. G. A. S. (2020) Development of induction technology on agarwood cultivation-

a review. *Jurnal Sangkareang Mataram*, **6(1)**, 29–39.

20. Huang, X., Li, H., Ruan, Y., Li, Z., Yang, H., Xie, G., Yang, Y., Du, Q., Ji, K. and Yang, M. (2022) An integrated approach utilizing raman spectroscopy and chemometrics for authentication and detection of adulteration of agarwood essential oils. *Frontiers in Chemistry*, **10**.

21. Desa, A. P., Lee, S., Mustapa, M. Z., Mohamed, R. and Emang, D. (2021) Trends in the agarwood industry of Peninsular Malaysia. *Malays. For*, **84**, 152–168.

22. Jailani, N. S. J., Muhammad, Z., Rahiman, M. H. F. and Taib, M. N. (2021) Analysis of Kaffir Lime Oil Chemical Compounds by Gas Chromatography-Mass Spectrometry (GC-MS) and Z-Score Technique. In *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS), IEEE*.

23. Abas, M. A. H., Ismail, N., Ali, N., Tajuddin, S. and Tahir, N. M. (2020) Agarwood oil quality classification using support vector classifier and grid search cross validation hyperparameter tuning. *Int. J.*, **8**.

24. Mahabob, N. Z., Mohd Amidon, A. F., Mohd Yusoff, Z., Ismail, N., Tajuddin, S. N., Mohd Ali, N. and Taib, M. N. (2021) Comparison of ANN performance towards agarwood oil compounds pre-processing based on principal component analysis (PCA) and stepwise regression selection method. *Journal of Electrical and Electronic Systems Research (JEESR)*, **19**, 51–55.

25. Mohamad, M., Najib, M. S., Tajuddin, S. N., Daud, S. M., Majid, N. F. H., Zaib, S. and Zahari, M. F..(2022) kNN: Classification of Agarwood Types in Oil and Wooden Using E-nose. In *Proceedings of the 6th International Conference on Electrical, Control and Computer Engineering, Springer*.

26. Haron, M. H. (2020) Agarwood oil quality grading model using selforganizing map (SOM) from *Department of System, Universiti Teknologi MARA*: UiTMIR.

27. Haron, M. H., Taib, M. N., Ismail, N., Ali, N. A. M. and Tajuddin, S. N. (2018) Statistical analysis of agarwood oil compounds based on GC-MS data. in *2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC), IEEE.*

28. Ahmad, T. and Aziz, M. N. (2019) Data preprocessing and feature selection for machine learning intrusion detection systems. *ICIC Express Lett.*, **13(2)**, 93–101.