IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 4: August 2025, pp. 2889~2898
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i4.pp2889-2898

Application of self-organizing map for modeling the *Aquilaria malaccensis* oil using chemical compound

Mohammad Arif Fahmi Che Hassan, Zakiah Mohd Yusoff, Nurlaila Ismail, Mohd Nasir Taib
Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia

Article Info

Article history:

Received Apr 8, 2024
Revised Mar 25, 2025
Accepted Jun 8, 2025

Keywords:

Agarwood oil
Aquilaria species
Grading classification
Self-organizing map
System identification

ABSTRACT

Agarwood oil, known as 'black gold' or the 'wood of God,' is a globally prized essential oil derived naturally from the *Aquilaria* tree. Despite its significance, the current non-standardized grading system varies worldwide, relying on subjective assessments. This paper addresses the need for a consistent classification model by presenting an overview of *Aquilaria malaccensis* oil quality using the self-organizing map (SOM) algorithm. Derived from the Thymelaeaceae family, *Aquilaria malaccensis* is a primary source of agarwood trees in the Malay Archipelago. Agarwood oil extraction involves traditional methods like solvent extraction and hydro-distillation, yielding a complex mixture of chromone derivatives, oxygenated sesquiterpenes, and sesquiterpene hydrocarbons. This study categorizes agarwood oil into high and low grades based on chemical compounds, utilizing the SOM algorithm with inputs of three specific compounds: β-agarofuran, α-agarofuran, and 10-epi-φ-eudesmol. Findings demonstrate the efficacy of SOM-based quality grading in distinguishing agarwood oil grades, offering a significant contribution to the field. The non-standardized grading system's inefficiency and subjectivity underscore the necessity for a standardized model, making this research crucial for the agarwood industry's advancement.

Corresponding Author:

Zakiah Mohd Yusoff
Faculty of Electrical Engineering, Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
Email: zakiah9018@uitm.edu.my

1. INTRODUCTION

Agarwood oil, also known as "gaharu" oil in Malaysia and Indonesia, is extracted from agarwood trees of the genus *Aquilaria malaccensis* and the Thymelaeaceae family [1]–[4]. The formation of matured agarwood results from various factors, including animal grazing, insect attacks, microbial invasions, and lightning strikes [2], [5]. Consequently, agarwood is acknowledged as resin-impregnated heartwood, with every part of the plant serving a purpose, including the tree trunks, branches, and agarwood stems. The stem can undergo processing to yield essential oil, commonly referred to as agarwood oil [6]. In contemporary times, agarwood oil holds high regard for its applications in perfumery, as a symbol of luxury, medicinal uses, and religious rituals, leading to a steady increase in demand [1], [5], [7]. Some years ago, agarwood oil grading relied on traditional methods based on factors such as color and odor [8]. However, using human sensory panels, particularly the sense of smell, for grading agarwood oil was considered inefficient [7]. This approach presented more drawbacks than advantages, including a high level of subjectivity and causing fatigue among the sensory panel due to the repetitive and continuous nature of the method [9].

Over the course of technological advancements, the grading of agarwood oil has evolved, incorporating modern techniques aligned with current developments. Intelligent methods, such as Z-score

analysis, artificial neural networks (ANN), multilayer perceptron (MLP), support vector machine (SVM), k-nearest neighbors, and linear regression, have been proposed for grading [10]–[15]. These contemporary grading approaches rely on the chemical properties of agarwood oil, aiming to enhance the accuracy and reliability of the grading system [13], [14], [16].

Various extraction techniques are employed to obtain agarwood oil, including supercritical fluid extraction, solvent extraction, hydro-distillation, and others. Pre-treatment of agarwood samples, involving chemical treatment, soaking in water, and sonication, is deemed necessary before extraction [4], [17]. Some studies utilize gas chromatography-mass spectrometer (GC-MS) [12], [18], [19] and solid-phase microextraction (SPME) [18], [20] for further analysis of the extracted oil.

Several research studies incorporate statistical analysis techniques, such as the z-score, to determine the quality grades of agarwood oil. Additionally, machine learning algorithms, including ANN, support vector classifiers, and random forests, are employed to validate these grades [21]–[24]. The z-score method relies on detecting variations in the abundance patterns of individual compounds, and in the case of agarwood oil quality grading, seven specific compounds: β-agarofuran, α-agarofuran, 10-epi-ɤ-eudesmol, ɤ-eudesmol, longifolol, hexadecanol, and eudesmol significantly influence the assessment [25], [26]. These compounds play a pivotal role in determining the quality of agarwood oil.

The β-agarofuran, α-agarofuran, 10-epi-ɤ-eudesmol, and ɤ-eudesmol are instrumental in grading high-quality agarwood oil, while longifolol, hexadecanol, and eudesmol contribute to the grading of low-quality agarwood oil [25]. In specific agarwood oil samples, JBD and MA2 were identified as high quality, whereas CKE, HD, and R5 were classified as low quality [21]. When using the z-score method for grading agarwood oil quality, both ANN and random forest algorithms have proven effective in accurately categorizing agarwood oil as either high or low quality, with minimal prediction error [24], [27].

The self-organizing map (SOM), an ANN employing a clustering algorithm for high-dimensional visualization, is also known as the Kohonen network, a concept introduced by Teuvo Kohonen in 1981. The advantages of using SOM include the following [27]–[29]:

- Dimensional reduction: SOM facilitates the reduction of dimensions, simplifying the interpretation of clustering outcomes. By transforming a high-dimensional input space into a lower-dimensional output space, it retains the original topological relationships.
- Suitability for complex data: SOM is applicable in scenarios where a comprehensive understanding of the input data's characteristics is absent. It excels at identifying patterns and relationships even when the data is not thoroughly understood.
- Ease of use: the algorithm is uncomplicated and easy to compute, enhancing its practicality and usability across diverse applications.

SOM, also referred to as the Kohonen network, proves to be a valuable tool for both visualizing and clustering high-dimensional data. Its versatility, simplicity, and broad applicability make it an excellent choice for various data analysis tasks.

Figure 1 illustrates the neural network structure of the SOM, comprising two layers: the input layer and the output layer, often called the competition layer. The number of neurons in the input layer is determined by the quantity of vectors in the input network. These input layer neurons establish connections with neurons in the output layer through weights, denoted as W [30]. Each neuron in the output layer can be conceptualized as representing a class or cluster that characterizes the inputs [20]. The organization of neurons in the output layer forms a two-dimensional grid or node matrix, facilitating the visualization and organization of the clustering process within the SOM. This configuration enables the SOM to capture and represent intricate patterns and relationships present in the input data.

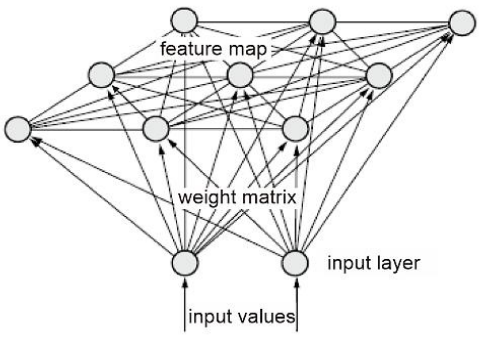


Figure 1. SOM architecture [20]

A SOM functions as a competitive neural network, following the principles of competitive learning. In the output layer, also known as the competition layer, neurons engage in competition to be selected as the 'winner,' determined by their proximity to input data vectors. Once the winning neuron and its neighboring neurons are identified, the weight vectors associated with them undergo modification. This adjustment process is strategically designed to enhance the responsiveness of the winning neuron and its neighbors to similar input patterns, fostering a self-organizing mechanism. These fundamental steps encapsulate Kohonen's SOM approach [28], [31].

Modeling the complex relationships between the significant compounds and oil quality is essential for unlocking the full potential of *Aquilaria malaccensis* in various applications. SOM have emerged as a powerful tool for modeling complex datasets and identifying patterns in multidimensional data [6]. SOM's ability to cluster and map high-dimensional data makes it a promising approach for understanding the intricate relationships within the sesquiterpenoid composition of *Aquilaria malaccensis* oil. It employs unsupervised learning, allowing it to reveal hidden patterns and structures within sesquiterpenoid datasets without the need for predefined categories. This characteristic is particularly advantageous in exploring the diverse chemical composition of *Aquilaria malaccensis* oil [32], [33]. SOM provides topological mapping, preserving the spatial relationships between different sesquiterpenoid compounds. This feature is crucial for understanding how subtle variations in the chemical structure impact oil quality [6], [33]. Unlike conventional techniques that may rely on subjective assessments, SOM provides an objective and data-driven approach to modeling sesquiterpenoids, offering a more nuanced understanding of their role in oil quality. SOM, a type of ANN, has proven effective in various fields, including cheminformatics and chemical pattern recognition. In the context of natural products, SOM has been successfully applied to model complex chemical data, providing insights into compound relationships and classifications [32], [33]. However, its application in grading the quality of *Aquilaria malaccensis* oil, particularly considering sesquiterpenoid profiles, remains an underexplored area.

2. THEORITICAL WORK

The outcomes derived from SOM learning offer valuable insights into the relationships among neighboring neurons, known as SOM neighbor distances, and the distribution of weight values, visualized as SOM weight planes. Typically, these results are presented using color maps [32], [34]. In SOM neighbor distances, hexagons and red lines depict neurons and their connections, respectively. The darkness of the color reflects the degree of distance, with darker shades indicating greater distances and lighter shades indicating smaller ones [34].

SOM weight planes visually demonstrate the link between color and the weight of the output neuron, as depicted in Figure 2. In this representation, lighter and darker colors correspond to larger and smaller weights, respectively. When the connection patterns of two inputs exhibit a high degree of similarity, meaning the shape and color of neurons are the same for both inputs and it suggests a strong correlation between those inputs [34]. The silhouette index (SI) functions as a valuable tool for cluster assessment, aiding in the identification of objects that are appropriately placed within their assigned clusters and those that might fall in between clusters. In Figure 3 [35] two clusters are labeled as A and C.

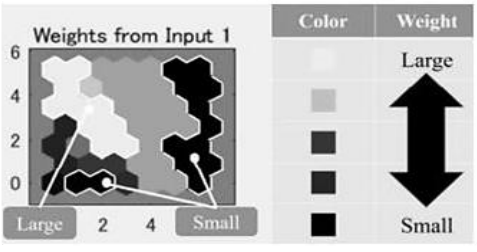


Figure 2. SOM weight planes of input [33]

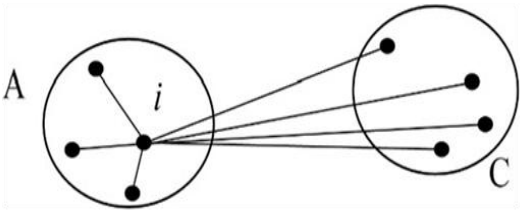


Figure 3. Computing silhouette index

To compute the SI, begin by selecting and labeling an object in cluster A as "i," then calculate a(i), representing the average dissimilarity of "i" to other objects within A. This provides insight into the average length of connections within cluster A. Subsequently, calculate c (i,C), denoting the average dissimilarity of "i" to all objects in cluster C, indicating the average length of connections from "i" (in cluster A) to cluster C. Proceed to calculate all c (i,C) values. Finally, identify the smallest number among these values [34], [35]. In

essence, the SI evaluates clustering quality by comparing how well an object within cluster A is internally connected (a(i)) versus its connections to objects in other clusters (c(i, C)), with a lower value indicating better clustering.

The SI ranges from -1 to 1, offering crucial information about clustering quality. An SI value close to 1 signifies that the object "i" is significantly closer to other objects within the same cluster than to those in the nearest neighboring cluster, indicating a robust and well-defined cluster. An SI close to 0 suggests uncertainty or ambiguity in the clustering of the focal object, implying it may not distinctly belong to any specific cluster. An SI value near -1 indicates misclustering, where the object seems more aligned with a different cluster than its assigned one.

To offer a practical interpretation: An SI falling between 0.71 and 1.00 is deemed an "excellent split," indicating a robust and clearly defined cluster separation. An SI ranging from 0.51 to 0.70 is labeled a "reasonable split," signifying a reasonably well-separated cluster. An SI between 0.26 and 0.50 is categorized as a "weak split," suggesting a less distinct cluster separation. An SI below 0.5 is termed a "bad split," indicating a poor separation of clusters [35]. Additionally, calculating the average SI values over a cluster can provide an assessment of the overall quality or "goodness" of that cluster. The advantages of SI are outlined as follows:

- SI validates clustering at the point level, providing the finest granularity.
- SI is independent of any specific algorithm.
- It relies solely on pairwise similarities-dissimilarities and the membership matrix as input.
- SI is applied for evaluating the clustering quality of a separation, fulfilling the objective of clustering by assessing both closeness and separation.

3. METHODOLOGY

3.1. Sample acquisition

The agarwood oil samples utilized in this study are exclusively derived from *Aquilaria* species, sourced from the Forest Research Institute Malaysia (FRIM) and Universiti Malaysia Pahang (UMP). A total of 660 samples, comprising 22 primary samples named as CKE, CM, EO2, EO3, EO4, HD, HG, JBD, KB, LA, LG, M, MA, MA1, MA2, MN, MNS, MPE, MS, R5, RG, and T, were employed in this research. Each sample consisted of 103 chemical compounds, which were meticulously extracted and analyzed using GC-MS. The GC-MS apparatus was configured with the following settings:

- The initial temperature of the apparatus was set at 60 ºC for 10 minutes.
- The temperature gradually increased, reaching 230 ºC with an increment of 3 ºC per minute.
- The flow rate of the helium gas carrier was maintained at 1 ml per minute.
- The temperature of the ion source was set at 280 ºC.

Identification of significant chemical compounds was accomplished by matching them to the mass spectral library (HPCH2205.L; Wiley7Nist05a.L; NIST05a.L), aided by a chemist.

3.2. Designation of agarwood oil grades

In this section, we employed the SOM clustering technique for categorizing agarwood oil grades. The input data for training and testing purposes was derived from principal component analysis (PCA) and Pearson's correlation. Initially, the data underwent per-row randomization, followed by division into an 80:20 ratio for the training and testing datasets. Subsequently, each dataset underwent transposition.

Prior to applying the SOM algorithm, a thorough assessment ensured the inclusion of all essential samples in both datasets. If this condition was met, the parameters of the SOM, including dimension, topology, and distance function, were set. In cases where the criteria were not satisfied, the randomization, division, and transpose processes were iterated. The SOM training and testing procedures were then executed.

Following the training and testing phases, silhouette values of the clusters were computed and scrutinized for negative values in both datasets. If negative values were detected, the entire clustering process was recalculated. Only clusters exhibiting positive silhouette values were acknowledged. Upon acknowledgment, the SOM validation procedure was initiated. The clustering rules applied in the SOM algorithm are outlined as follows:

a. Input: chemical compounds
b. Output: number of neurons that represent the grades of agarwood oil
c. Dimension:

- 1 by 2 grid for 2 grades (each neuron represents a cluster representing either high or low grade of agarwood oil)

– 1 by 3 grid for 3 grades (each neuron represents a cluster representing either high, medium, or low grade of agarwood oil)
– 2 by 2 grid for 4 grades (each neuron represents a cluster representing either high, medium-high, medium-low, or low grade of agarwood oil)

d. Topology function: Hextop (hexagonal pattern)
e. Distance function: Euclidean distances

Out of 103 chemical compounds, only three were selected as input variables for the SOM: β-agarofuran, α-agarofuran, and 10-epi-ɤ-eudesmol. The resulting output was indicative of the agarwood oil grade, categorized as either high or low. The SOM clustering process adhered to the following rules:

– The ratio of training and testing: 80 to 20
– Two neurons: each neuron represents a cluster that is either high or low quality
– Dimension: 1 by 2 grid
– Topology function: hextop (hexagonal pattern)
– Distance function: Euclidean distances
– CoverSteps: 100
– InitNeighbor: 1

These specified rules were incorporated into the SOM algorithm for implementation. The initial step involved randomizing the data on a per-row basis, followed by their division into training and testing datasets with an 80:20 ratio. Subsequently, each dataset underwent transposition. Prior to the SOM computation, the testing dataset underwent scrutiny to ensure the inclusion of all primary samples.

Following this, key SOM properties, including dimension, coverSteps, initNeighbor, topology, and distance function, were configured. The datasets for training, testing, and validation were then computed sequentially. Subsequently, silhouette values for each cluster were computed and scrutinized for negative values in both training and testing datasets. Only clusters exhibiting positive silhouette values were considered valid. The program concluded upon the fulfillment of this condition, as illustrated in Algorithm 1.

Algorithm 1. SOM algorithm for clustering

```
Input: data T, training data Tr, testing data Ts, validation data Tv, main samples M
Output: predicted cluster of training data Dtr, predicted cluster of testing data Dts,
predicted cluster of validation data Dtv
    load T, Tv
    while silhouette values of Dtr ≤ 0 or Dts ≤ 0 do
    While Ts ⊅ M do
    randomize T
    split T to Tr and Ts with 80 to 20 ratio
    Tr′ ← Tr; Ts′ ← Ts
    end while
    set SOM parameters: dimensions, topologyFcn, distanceFcn
    start training
    start testing
    return Dtr, Dts
    calculate and plot silhouette values of Dtr, Dts
    calculate average silhouette values of Dtr, Dts
    end while
    start validation
    return Dtv
```

4. RESULTS AND DISCUSSION

The outcomes encompassing weight distances between neurons, compound weights to neurons, silhouette values for both training and testing, and the assignment of neurons to agarwood oil samples will be presented and discussed within this subsection. Figure 4 illustrates the weight distance between neurons, with each neuron depicted as a blue hexagon. Neuron 1 is situated at the bottom, while neuron 2 is positioned at the top. The coloration of the region between the neurons serves as an indicator of their distance, with darker hues signifying greater distances and vice versa. In Figure 4, the region is colored red, indicating a moderate distance between the neurons. This visual representation aids in the interpretation of the neural relationships in the context of the study.

Figure 5 displays the compound weights assigned to neurons, with hexagons representing neurons labeled as Neurons 1 and 2 at the bottom and top, respectively. The colors of these neurons signify the respective compound weights, where lighter and darker shades denote larger and smaller contributions. Notably, the compounds β-agarofuran, α-agarofuran, and 10-epi-ɤ-eudesmol demonstrated a more substantial

contribution to neuron 1 compared to neuron 2. Literature sources have consistently identified these compounds as significant contributors to high-quality oil. Consequently, neuron 1 is indicative of a high-grade cluster, while neuron 2 represents a low-grade cluster. The connection pattern of these compounds remained consistent across all inputs, with neuron 1 being yellow and neuron 2 being black, highlighting a strong correlation between the compounds.

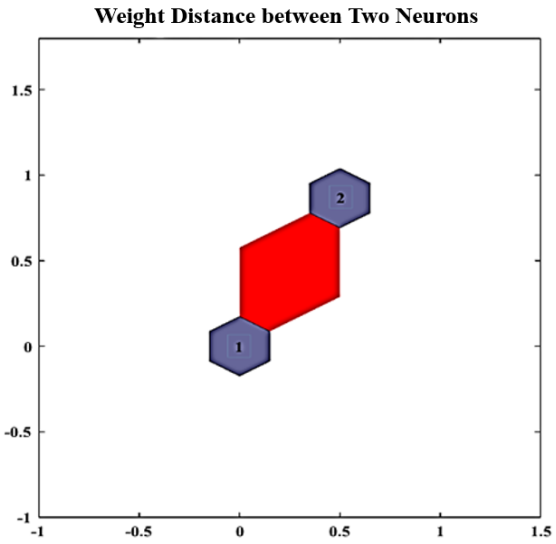


Figure 4. Weight distance between neurons

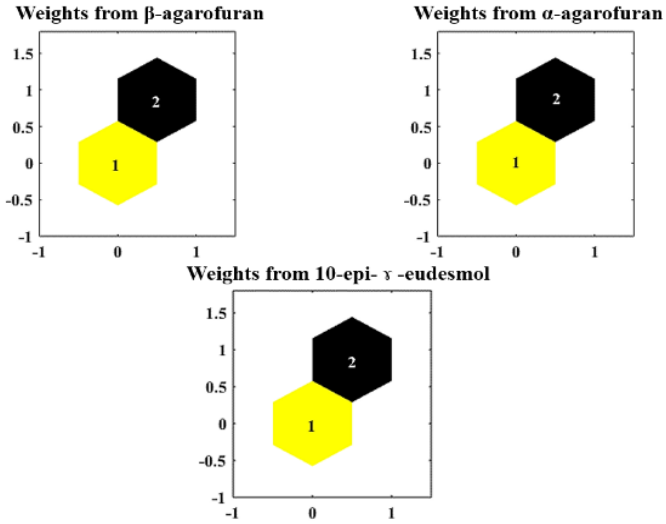
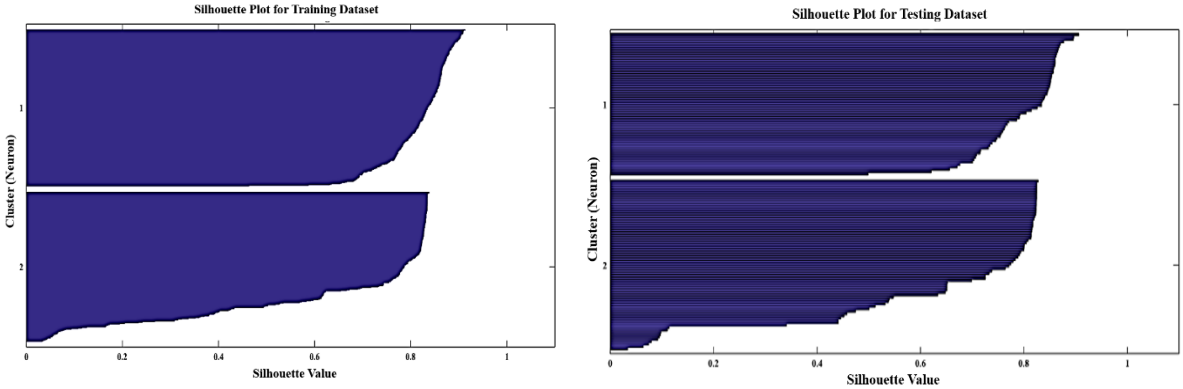


Figure 5. Weight of compounds to neurons 1 (high grade) and 2 (low grade)

Figure 6 presents silhouette plots for each neuron, the training plot shown in Figure 6(a) and the testing plot shown in Figure 6(b). The dataset comprised 528 training samples and 132 test samples, all of which exhibited positive silhouette values in both training and testing phases. Specifically, in the training set, the average silhouette values for neuron 1 (representing the high-grade cluster) and neuron 2 (representing the low-grade cluster) were 0.82 and 0.67, respectively. During testing, the average silhouette values for neuron 1 and neuron 2 were 0.79 and 0.58, respectively. This observation signifies that the high-grade cluster (neuron 1) consistently yielded superior average silhouette values compared to the low-grade cluster (neuron 2) in both the training and testing datasets. The silhouette plots suggest that the samples align well within their designated clusters, distinguishing between low and high grades, but exhibit poorer

alignment with their neighboring clusters. This alignment pattern provides valuable insights into the efficacy of the clustering process and the distinct separation of low and high-grade clusters.



(a) (b)

Figure 6. Silhouette plot for (a) training dataset and (b) testing dataset

The allocation of primary samples to distinct agarwood oil grades is meticulously detailed in Table 1, providing a comprehensive overview of the training and testing phases. Notably, the high grade category encompasses JBD, KB, LA, MA, MA1, MA2, MNS, MPE, RG, and T, collectively forming a cohesive unit within the high-grade cluster associated with neuron 1. Conversely, the low grade cluster, represented by neurons 2, comprises CKE, CM, EO2, EO3, EO4, HD, LG, M, MN, MS, and R5. The numeric values featured in the table denote the quantity of samples within each respective category.

This strategic assignment of primary samples underscores the precision of our approach, aligning with the neural network's ability to discern and classify agarwood oil grades. The explicit detailing of the sample distribution among neurons enhances the transparency and replicability of our methodology. The results manifest a clear demarcation between high and low-grade clusters, setting the stage for a robust evaluation of the proposed classification model.

Table 1. Training and testing data

Samples	Training		Testing	
	Neuron1 (high grade)	Neuron2 (low grade)	Neuron1 (high grade)	Neuron2 (low grade)
CKE	0	22	0	8
CM	0	24	0	6
EO2	0	23	0	7
EO3	0	21	0	9
EO4	0	24	0	6
HD	0	22	0	8
HG	27	0	3	0
JBD	25	0	5	0
KB	27	0	3	0
LA	23	0	7	0
LG	0	25	0	5
M	0	24	0	6
MA	26	0	4	0
MA1	24	0	6	0
MA2	23	0	7	0
MN	0	27	0	3
MNS	25	0	5	0
MPE	26	0	4	0
MS	0	21	0	9
R5	0	25	0	5
RG	21	0	9	0
T	23	0	7	0

5. CONCLUSION

The analysis demonstrates that a minimal yet targeted selection of three chemical compounds: β-agarofuran, α-agarofuran, and 10-epi-π-eudesmol can effectively classify agarwood oil into high or low grades. This finding reinforces the reliability and discriminative power of these specific markers in assessing agarwood oil quality, making them valuable indicators for both research and industry applications. Additionally, the application of SOM demonstrates notable proficiency in clustering agarwood oil into high and low grades, as evidenced by average silhouette values ranging from 0.58 to 0.82. This not only reinforces the efficacy of the chosen chemical compounds but also highlights the ability of SOM to provide a reliable and accurate classification of agarwood oil quality. In conclusion, the proposed method of quality grading for agarwood oil, relying on the chemical compounds β-agarofuran, α-agarofuran, and 10-epi-ɤ-eudesmol through SOM, has been validated as an effective and dependable approach. As a direction for future research, extending the quality classification into high, medium, and low grades could offer a more nuanced and refined understanding of agarwood oil variations. This expansion would contribute further to the advancement of agarwood industry standards and deepen our insight into the nuanced gradations within this valuable essential oil.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to all parties involved and to Universiti Teknologi MARA (UiTM) for their continuous support and contributions throughout the course of this work.

FUNDING INFORMATION

The authors wish to extend their appreciation to the Faculty of Electrical Engineering at UiTM Shah Alam, Selangor for their continuous financial support during this research under FRGS grant (600-RMC/FRGS 5/3 (154/2023).

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohammad Arif Fahmi Che Hassan	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			
Zakiah Mohd Yusoff		✓				✓	✓	✓	✓	✓	✓	✓	✓	
Nurlaila Ismail		✓			✓	✓	✓			✓	✓		✓	✓
Mohd Nasir Taib	✓	✓			✓			✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

Not applicable.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] P. Sarmah *et al.*, "An insight into the immunomodulatory potential of wood oil of Aquilaria malaccensis Lam. with an emphasis into related phytomedicine, biomarkers, pharmacology, and toxicity," *South African Journal of Botany*, vol. 151, pp. 695–712, 2022, doi: 10.1016/j.sajb.2022.10.020.

[2] N. Z. Mahabob, Z. M. Yusoff, A. F. M. Amidon, N. Ismail, and M. N. Taib, "A novel application of artificial neural network for classifying agarwood essential oil quality," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, pp. 6645–6652, 2022, doi: 10.11591/ijece.v12i6.pp6645-6652.

[3] R. Gogoi *et al.*, "Agarwood (Aquilaria malaccensis L.) a quality fragrant and medicinally significant plant based essential oil with pharmacological potentials and genotoxicity," *Industrial Crops and Products*, vol. 197, 2023, doi: 10.1016/j.indcrop.2023.116535.

[4] P. Wu *et al.*, "Extraction process, chemical profile, and biological activity of aromatic oil from agarwood leaf (Aquilaria sinensis) by supercritical carbon dioxide extraction," *Journal of CO2 Utilization*, vol. 77, 2023, doi: 10.1016/j.jcou.2023.102615.

[5] N. Zhang *et al.*, "Effects of various artificial agarwood-induction techniques on the metabolome of Aquilaria sinensis," *BMC Plant Biology*, vol. 21, no. 1, 2021, doi: 10.1186/s12870-021-03378-8.

[6] S. P. Munasinghe, S. Somaratne, S. R. Weerakoon, and C. Ranasinghe, "Ecological origin of the appearance of sesquiterpenes in Gyrinops walla Gaetner by wood anatomical and chemical analysis," *Journal of the Indian Academy of Wood Science*, vol. 18, no. 2, pp. 97–105, 2021, doi: 10.1007/s13196-021-00285-1.

[7] F. A. A. Kadir, K. A. Azizan, and R. Othman, "Datasets of essential oils from naturally formed and synthetically induced Aquilaria malaccensis agarwoods," *Data in Brief*, vol. 28, 2020, doi: 10.1016/j.dib.2019.104987.

[8] H. E. Shaian, C. M. Aizal, and S. N. Tajuddin, "Chemical composition of agarwood essential oil (Aquilaria malaccensis) upon exposure towards heat condition," *Malaysian Journal of Chemistry*, vol. 23, no. 2, pp. 199–204, 2021, doi: 10.55373/mjchem.v23i2.1013.

[9] M. H. Abdullah, S. A. C. Ghani, and S. R. Mohamed, "Experiment and numerical studies on heat loss in a hydro-distillation with various insulations," *MATEC Web of Conferences*, vol. 225, 2018, doi: 10.1051/matecconf/201822502014.

[10] S. M. H. M. Huzir *et al.*, "Stepwise regression of agarwood oil significant chemical compounds into four quality differentiation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 735–741, 2023, doi: 10.11591/ijeecs.v29.i2.pp735-741.

[11] J. Dutta, D. Lahon, P. K. Bora, M. Bhuyan, and S. Haldar, "A high performance thin layer chromatography (HPTLC) method for the quality assessment of agarwood (Aquilaria malaccensis) oil from Northeast India," *Natural Product Research*, vol. 38, no. 9, pp. 1634–1638, 2024, doi: 10.1080/14786419.2022.2162516.

[12] S. Lias, N. A. M. Ali, M. Jamil, M. H. Zainal, A. M. Jalil, and N. Ismail, "Classification of agarwood oils using K-NN K-fold," *Sensor Letters*, vol. 12, no. 6–7, pp. 1038–1040, 2014, doi: 10.1166/sl.2014.3187.

[13] M. Mohamad *et al.*, "kNN: classification of agarwood types in oil and wooden using e-nose," in *The 6th International Conference on Electrical, Control and Computer Engineering*, 2022, pp. 575–586. doi: 10.1007/978-981-16-8690-0_52.

[14] Y. Chen *et al.*, "Structure and chemical characteristics of lyuqinan agarwood from Huidong," *Chinese Journal of Wood Science and Technology*, vol. 36, no. 1, pp. 49–56, 2022, doi: 10.12326/j.2096-9694.2021059.

[15] N. H. M. Hussain, A. Anuar, S. Masrom, T. Mohd, N. A. Ahmad, and H. Byrd, "Machine learning in reverse migration classification," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 38, no. 2, pp. 45–55, 2024, doi: 10.37934/araset.38.2.4555.

[16] D. T. Ahmaed, A. M. Masaad, and S. N. Tajuddin, "Identification of agarwood (Aquilaria malaccensis) chips incense smoke and headspace volatile compounds by gc-MS.EI.Q.TOF, SPME," *Malaysian Journal of Chemistry*, vol. 20, no. 2, pp. 154–162, 2018.

[17] P. Mohan and K. Patil, "Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application," *International Journal of Intelligent Engineering and Systems*, vol. 11, no. 4, pp. 167–176, Aug. 2018, doi: 10.22266/ijies2018.0831.17.

[18] C. Y. Shu, H. L. Lai, P. Y. Lin, K. C. C. Chou, and M. Shelomi, "Preliminary studies on the effects of Taiwanese ritual smoke on Aedes aegypti (Linnaeus, 1762) (Diptera: Culicidae)," *Pan-Pacific Entomologist*, vol. 97, no. 2, pp. 67–74, 2021, doi: 10.3956/2021-97.2.67.

[19] H. H. Wang, Y. C. Wang, B. L. Wee, and S. W. Sim, "Novel feature extraction for oil palm bunches classification," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 34, no. 1, pp. 350–360, 2023, doi: 10.37934/araset.34.1.350360.

[20] A. N. M. Ramli, S. Yusof, P. Bhuyar, A. W. Aminan, S. N. Tajuddin, and H. A. Hamid, "Production of volatile compounds by a variety of fungi in artificially inoculated and naturally infected Aquilaria malaccensis," *Current Microbiology*, vol. 79, no. 5, 2022, doi: 10.1007/s00284-022-02840-6.

[21] 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, 2015, doi: 10.1007/s10600-015-1409-2.

[22] S. Lias, A. M. Jalil, M. Jamil, M. S. Y. Tolmanan, N. Ismail, and M. N. Taib, "The application of electronic nose coupled with 80:20 k-nearest neighbors classification technique for agarwood oil quality index establishment," *Malaysian Journal of Analytical Sciences*, vol. 26, no. 6, pp. 1240–1248, 2022.

[23] R. Kalra and N. Kaushik, "A review of chemistry, quality and analysis of infected agarwood tree (Aquilaria sp.)," *Phytochemistry Reviews*, vol. 16, no. 5, pp. 1045–1079, 2017, doi: 10.1007/s11101-017-9518-0.

[24] F. A. Aditama, L. Zulfikri, L. Mardiana, T. Mulyaningsih, N. Qomariyah, and R. Wirawan, "Electronic nose sensor development using ANN backpropagation for Lombok agarwood classification," *Research in Agricultural Engineering*, vol. 66, no. 3, pp. 97–103, 2020, doi: 10.17221/26/2020-RAE.

[25] E. Haslina Abd Latib, A. K. M. Moyeenul Huq, S. Zareen, and S. Nizam Tajuddin, "Gas chromatography mass spectrometry couple with quadrupole time-of-flight (GC-QTOF MS) as a powerful tool for profiling of oxygenated sesquiterpenes in agarwood oil," *Arabian Journal of Chemistry*, vol. 16, no. 9, 2023, doi: 10.1016/j.arabjc.2023.105025.

[26] Z. Wu and L. Yu, "Characteristic quality analysis for biologically induced agarwood columns in Aquilaria sinensis," *Environmental Research*, vol. 235, 2023, doi: 10.1016/j.envres.2023.116633.

[27] N. Z. Mahabob *et al.*, "A study on ANN performance towards three significant compounds of high quality agarwood oil," in *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, IEEE, 2022, pp. 116–120. doi: 10.1109/CSPA55076.2022.9782017.

[28] S. Rejeb, C. Duveau, and T. Rebafka, "Self-organizing maps for exploration of partially observed data and imputation of missing values," *Chemometrics and Intelligent Laboratory Systems*, vol. 231, 2022, doi: 10.1016/j.chemolab.2022.104653.

[29] K. Goel and W. Tabib, "Incremental multimodal surface mapping via self-organizing gaussian mixture models," *IEEE Robotics and Automation Letters*, vol. 8, no. 12, pp. 8358–8365, 2023, doi: 10.1109/LRA.2023.3327670.

[30] C. Li, C. Xiao, M. Li, L. Xu, and N. He, "The quality and quantity of SOM determines the mineralization of recently added labile C and priming of native SOM in grazed grasslands," *Geoderma*, vol. 432, Apr. 2023, doi: 10.1016/j.geoderma.2023.116385.
[31] A. H. Elsheikh *et al.*, "An artificial neural network based approach for prediction the thermal conductivity of nanofluids," *SN Applied Sciences*, vol. 2, no. 2, Feb. 2020, doi: 10.1007/s42452-019-1610-1.
[32] J. Polanski, "Unsupervised learning in drug design from self-organization to deep chemistry," *International Journal of Molecular Sciences*, vol. 23, no. 5, 2022, doi: 10.3390/ijms23052797.
[33] S. Licen, A. Astel, and S. Tsakovski, "Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review," *Science of the Total Environment*, vol. 878, 2023, doi: 10.1016/j.scitotenv.2023.163084.
[34] K. Saxena, S. Jaloree, R. S. Thakur, and S. Kamley, "Self organizing map (SOM) based modelling technique for student academic performance prediction," *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 3, no. 9, pp. 115–120, 2017.
[35] O. Murad and M. Malkawi, "An optimization method using clustering technique for the human emotions detection artificial neuro-fuzzy logic system," *International Journal of Computers & Technology*, vol. 15, no. 9, pp. 7090–7096, 2016, doi: 10.24297/ijct.v15i9.695.

BIOGRAPHIES OF AUTHORS

Mohammad Arif Fahmi Che Hassan was born in Malaysia, on May 1996. He received his Bachelor of Chemical Engineering Technology (Hons) (Biotechnology Industry) from Universiti Malaysia Perlis. He is currently engineer at Vision Industries and at the same time as full-time postgraduate students at School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, UiTM Shah Alam, Malaysia. He can be contacted at email: Ariffahmi533@gmail.com.

Assoc. Prof. Ts. Dr. Zakiah Mohd Yusoff is a senior lecturer who is currently working at UiTM Pasir Gudang. She received the B.Eng. in Electrical Engineering and Ph.D. in Electrical Engineering from UiTM Shah Alam, in 2009 and 2014, respectively. In May 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.

Assoc. Prof. Ir. Ts. Dr. Nurlaila Ismail received her Ph.D. in Electrical Engineering from Universiti Teknologi MARA, Malaysia. She is currently a senior lecturer at School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Malaysia. Her research interests include advanced signal processing and artificial intelligence. She can be contacted at email: nurlaila0583@uitm.edu.my.

Prof. Ir. Ts. Dr. Mohd Nasir Taib received his Ph.D. from UMIST, UK. He is a Senior Professor at Universiti Teknologi MARA (UiTM). He heads the Advanced Signal Processing Research Group at the School of Electrical Engineering, College of Engineering, UiTM. He has been a very active researcher and over the years had author and/or co-author many papers published in refereed journals and conferences. He can be contacted at email: dr.nasir@uitm.edu.my.