Enhancing Aquilaria Species Classification using Agarwood Oil Analysis and K-Nearest Neighbors (KNN) as Machine Learning

Amir Hussairi Zaidi, Anis Hazirah 'Izzati Hasnu Al-Hadi, Noor Aida Syakira Ahmad Sabri, Nik Fasha Edora Nik Kamaruzaman, Nurlaila Ismail*, Zakiah Mohd Yusoff, Mohd Nasir Taib

***Abstract- Aquilaria* species is renowned for the aromatic resinous wood also known as agarwood and 17 accepted species are known can produce it. This tree has been long prized in various cultural and commercial context for its aromatic resinous wood, esteemed for its rich fragrance and versatile applications. The accurate classification of this agarwood-producing plant from *Aquilaria* species is essential for purposes such as conservation, sustainable resource management, trade regulation, research, and the preservation of cultural and economic traditions. This study employs machine learning, specifically the k-Nearest Neighbors (kNN) algorithm, to classify *Aquilaria* species based on chemical compounds features extracted from agarwood oil. Agarwood from four *Aquilaria* (A.) species: A. *beccariana*, A. *crassna*, A. *malaccensis*, and A. *subintegra*, is being used and the chemical compound is being analysed using the Gas Chromatography-Flame Ionization Detector (GC-FID). Subsequently, classifier performance is assessed using a confusion matrix to measure accuracy. The study not only demonstrates the effectiveness of this technique but also highlights its potential for future research related to *Aquilaria* and agarwood, reinforcing its relevance and applicability.**

***Index terms*—*Aquilaria*, agarwood, GC-FID, kNN, machine learning**

I. INTRODUCTION

Aquilaria is renowned for its abundance of gaharu, commonly referred to as agarwood. This *Aquilaria* is a tree belonging to the Thymelaeaceae family and consists of a total of 21 species accepted names [1], [2] where 17 species are recorded to produce agarwood [3], [4]. The part of the *Aquilaria* tree that is primarily used for mass production is the heartwood. The heartwood of certain species of *Aquilaria* is valued for the resin it produces, which is commonly known as gaharu or agarwod. This resin is formed in response to an injury or infection in the tree [5], and it contributes to the distinctive fragrance and properties associated with agarwood. Agarwood is highly sought after for its use in perfumes, incense, traditional medicine, and various cultural and religious practices. The mass production often involves the cultivation and stimulation of agarwood formation in *Aquilaria* trees.

The continuous imbalance between the demand and supply of agarwood has positioned it as one of the world's most costly and desirable aromatic materials. This heightened demand, alongside unsustainable harvesting methods and habitat destruction, has endangered numerous *Aquilaria* species [6]–[8]. Consequently, the preservation and sustainable stewardship of these species have become crucial to safeguard their existence.

Historically, the categorization of *Aquilaria* species has depended on botanical traits like leaf shape, bark characteristics, and where they are found geographically [9]–[11]. Yet, these approaches have shortcomings, especially with closely related species or hybrids that display nuanced differences in their physical traits. Such limitations in precisely identifying *Aquilaria* species can impact the grading of agarwood sourced from these trees, given that the quality of agarwood is intricately tied to the particular *Aquilaria* species it comes from. Hence, there's a demand for more precise and dependable methods in classifying these species.

The chemical compounds of agarwood oil offers significant vision into the taxonomy and phytochemistry of *Aquilaria* species. It consists mainly of volatile and semi-volatile compounds [12], [13], such as sesquiterpenes, sesquiterpene derivatives, and phenolic compounds [7], [14], [15]. Intriguingly, agarwood oil's compound can vary significantly not only between different *Aquilaria* species but also within individuals of the same species [16]. This diversity poses a challenge for traditional classification methods but also opens avenues to utilize chemical compounds for species differentiation.

Gas chromatography has appeared as a promising solution in addressing the shortcomings of traditional methods for classifying *Aquilaria* species. This method grant more precise identification of *Aquilaria* species by analyzing the volatile or semi-volatile compounds present in their wood, bark, or leaves. By aiding in the identification of visually similar *Aquilaria* species, this method also facilitates accurate grading of agarwood based on its chemical composition. Thus, using this method as the modern approach, it overcomes the limitations of traditional methods, in enhancing more understanding of various species of *Aquilaria* while ease the assessment of agarwood quality.

Nowadays, machine learning approaches, such as k-Nearest Neigbors (kNN) classifier, has been used widely in learning the chemical compounds of essential oil [17]–[19]. This supervised

This manuscript is submitted on 22 March 2024, revised on 26 March 2024, accepted on 29 March 2024 and published on 30 April 2024. Amir Hussairi Zaidi, Anis Hazirah 'Izzati Hasnu Al-Hadi, Noor Aida Syakira Ahmad Sabri, Nik Fasha Edora Nik Kamaruzaman, Nurlaila Ismail* and Mohd Nasir Taib are from School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia, while Zakiah Mohd Yusoff is from School of Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA Cawangan Johor, Kampus Pasir Gudang, Johor Malaysia.

*Corresponding author
Email address: nurlaila0583@uitm.edu.my

learning algorithm categorizes data points by their resemblance to to known data points. Previous study of *Aquilaria* and agarwood, this machine learning has been used to grade their quality based on their chemical compounds [20]–[22]. Thus, this study employs machine learning, specifically the k-Nearest Neighbors (kNN) algorithm, to classify Aquilaria species based on chemical compounds features extracted from agarwood oil. By compiling a comprehensive dataset of agarwood oil chemical profiles from four *Aquilaria* species, the research seeks to classify those species based on similar chemical compounds that exist in all species. Employing the KNN classifier on this dataset automates and enhances the classification process, ensuring swift and reliable species identification.

II. LITERATURE REVIEW

A. *Aquilaria and agarwood*

Aquilaria species are renowned for their production of agarwood, a prized aromatic resin with cultural and commercial significance [5], [23]. These trees exhibit a remarkable chemical diversity, containing volatile [24] and semi-volatile [13] compounds such as sesquiterpenes, sesquiterpene derivatives, and phenolic compounds in their wood, bark, and leaves. Fig. 1 shows some of the chemical structure of sesquiterpene that has been found in agarwood. Agarwood formation, triggered by injury or infection [25], leads to the accumulation of resinous compounds in the heartwood, each contributing to agarwood's unique fragrance and therapeutic properties. Advanced analytical techniques like gas chromatography-mass spectrometry (GC-MS) have enabled the identification and characterization of these complex chemical profiles [26], [27]. However, *Aquilaria* species face threats such as habitat loss and unsustainable harvesting, highlighting the need for conservation efforts to preserve these valuable trees and their habitats while exploring sustainable cultivation methods and biotechnological approaches for agarwood production [4], [7].

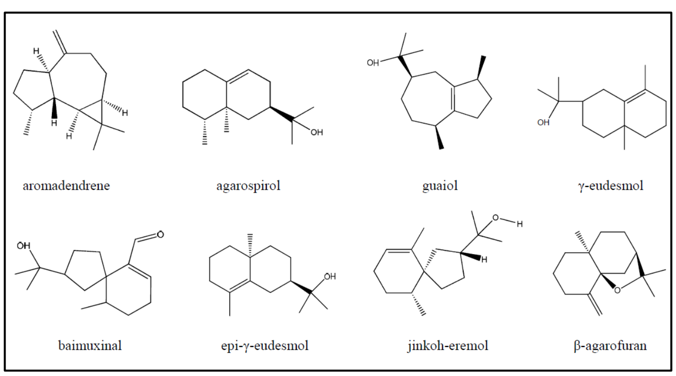

Fig. 1. Chemical structure of sesquiterpene compound found in agarwood.

B. *Chemical compound in agarwood oil*

Previous studies has been conducted in the research of chemical compounds of agarwood particularly in terpenes, terpenoids, acid and chromones [28]–[30]. Based on the studies, sesquiterpenes and chromones are the major producer for this agarwood oil. For instance, sesquiterpenes had been analyse to produce a lot of resin in high quality of agarwood oil. These include '(-)-guaia-1(10),11-dien-15-al', '(-)-selina-3,11-dien,9-one' and '(+)-selina-3,11-dien,9-ol' [28], [31], [32]. Guaiene compounds also can be found in high quality agarwood where the condition muss less than 0.05%. If no presence of this Guaiene compounds, it considered as low quality of agarwood oil [33]. Sesquiterpenoids such as '9-11-eremophiladien-8-one' [34] and 'oxo-Agarospirol' [35] also can be found in low quality of agarwood oil. Table I tabulates the chemical compounds of agarwood that contribute to the high quality or low quality of the oil.

TABLE I. LIST OF COMPOUNDS THAT CONTRIBUTE TO AGARWOOD OIL QUALITY

COMPONENT	LOW QUALITY	HIGH QUALITY	REF
Sesquiterpenes	Undefined	Major	[28]
Resin content	Low	High	[32]
(-)-guaia-1(10),11-dien-15-al, (-)-selina-3,11-dien,9-one and (+)-selina-3,11-dien,9-ol	No	Yes	[31]
Guaiene	No	Yes	[33]
9-11-eremophiladien-8-one	Yes	No	[34]
oxo-Agarospirol	Yes	No	[35]

C. *Gas Chromatography*

Gas Chromatography (GC) has been used widely by researcher in analysis of chemical compound, to identify the individual compounds in their analysis [18], [27], [36], [37]. One of it is Gas Chromatography with Flame Ionization Detection (GC-FID). This GC is used to separate the compounds while the FID is used to identify the volatile compounds in it. In the analysis of essential oil, GC-FID has emerged as an effective tools in separating and quantifying complex compound mixtures since it provides detailed chemical profiles, notably highlighting sesquiterpenes [29], [38] which are an aromatic compounds. This technique's precision in resolving components and its sensitivity to substances with carbon-hydrogen bonds make it particularly well-suited for the precise analysis of volatile compounds [27], [39]. Fig. 2 illustrates the schematic diagram of the set-up for the GC-FID [40].

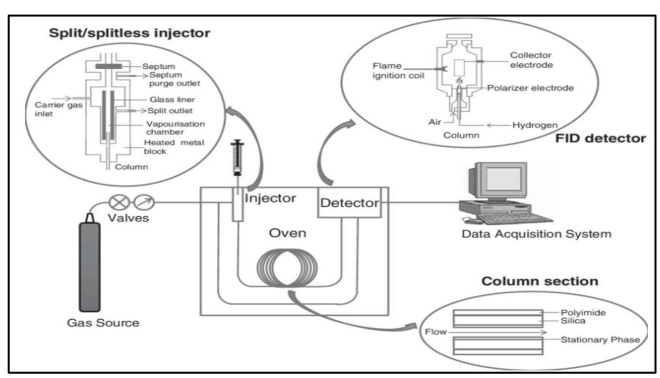

Fig. 2. Schematic Diagram of the GC-FID set up [40].

D. *k-Nearest Neighbors*

The k-Nearest Neighbors (kNN) classifier has become a valuable tool in classifying chemical compounds [41], [42], offering a straightforward yet effective approach. By leveraging similarity measures between data points, kNN assigns a class label based on the majority vote of its nearest neighbors (k-value). Here, the nearest neigbors and distance metric is set by the user's preference [43], [44] such as k-value of 10 with Euclidian distance as the distance metrics. These preference will then be the contribution to the decision making process.

In this study, the chosen distance metrics is Euclidian distance. This metrics, as described in [43], calculates the distance between two objects by taking the square root of the sum of the squares of their corresponding coordinates, as defined in (1). Based on the equation, d_{st} is the distance between the point of *sj* and *tj*. Then, x_{sj} is an object that located at coordinate *sj*, while y_{tj} is another object that located at coordinate *tj*.

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

III. METHODOLOGY

The experimental set-up for this classification of *Aquilaria* species based on agarwood oil is unfolded into several stages. Each stage explain the process from collecting data, feature engineering, classifying, and evaluating the model of the data. The first stage involves the sample collection of the data from physical ground into transformation of digital data (raw data) by Bioaromatic Research Centre of Excellence (BARCE) , Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). Next, using the pattern of similar compounds that exist across every *Aquilaria* species, the process of feature engineering will be employed to determine the input and output feature for secondary dataset that can be used in classification process. Here, generation of synthetic data will be executed to produce a dataset that suitable for classification process. Then, classification will be performed using k-Nearest Neighbors (kNN) classifier. When the classification has been execute, the model classification will be evaluated using confusion matrix to illustrate the model accuracy. The employment of the process from feature engineering to model evaluation will be implimented using MATLAB software and the stage mention is illustrated in Fig. 3.

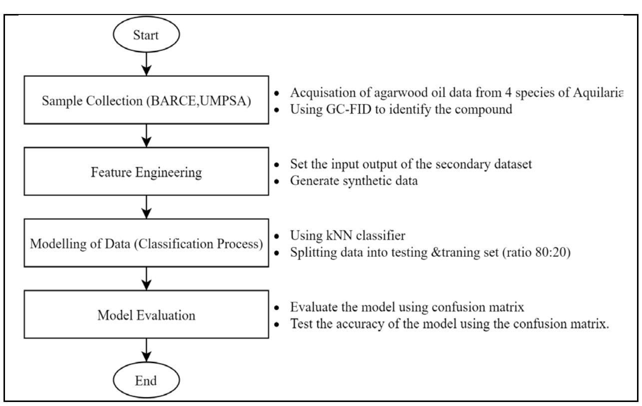

Fig. 3.Flowchart of the experimental set-up.

A. *Sample Preparation*

The sample preparation has been executed by Bioaromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) [45]–[48]. The physical sample ground is collected from four different species of *Aquilaria*: *Aquilaria beccariana*, *Aquilaria crassna*, *Aquilaria malaccencis*, *Aquilaria subintegra*, originates from Malaysia. After the physical sample ground has been acquired, the sample then will be processed to produce the agarwood oil. To do that, firstly the samples will be soaked in alcohol and oils for a week. Then the hydrodistillation process will be carried out to extract the oil for three to five days. When the oil has been acquired, the oil then will be placed inside dark glass bottle to prevent degradation of the oil and a label has been marked for each oil based on their type of species for futher used.

The agarwood oils will then be extracted using Gas Chromatography – Flame Ionization Detector (GC-FID) to identify the chemical compounds for every species of *Aquilaria*. This GC-FID uses Agilent 7890A network system gas. For this GC-FID instrument set up, the initial temperature of it is set for 60°C for about 10 minutes. Next the temperature ramped up from 3°C per minute to 250°C for about 10 minutes. The oil sample will then be injected for about 1 µL at a split ratio of 1:5. During the process, the chemical compound has been identified based on their retention time in the generated chromatogram. This process of GC-FID analysis is repeated for all agarwood oils from different species of *Aquilaria*.

When all the samples from each species has been required, the data from each samples will be combine together to form a primary dataset that can be used for classification process.

B. *Feature Engineering*

Using the primary dataset, the selection of the significant compound will be made based on the similar compound that can possibly exist across every *Aquilaria* species. This significant compound will then be the features for secondary dataset that will be used for the classification process. The input and output set up for the secondary dataset will be as follows:

- The input data will be the selected significant compound.
- The output data will be the species of *Aquilaria*.

To improve the performance of the secondary dataset for classification, the generation of synthetic data will be made for the selected compound. The improvement has been made to increase the sample size since it is less than 30 (n<30) and no algorithm work well with limited sample size [49]. The generation of synthetic data will be using the method of adding the noise with signal-to-noise ratio (SNR) as recommended by [50] leveraging the capability of MATLAB software to generate random Gaussian noise using the function 'randn'. The formulae of this generation of synthetic data can be express as (2), where $Y_{synthetic}$ is the synthetic data, $Y_{original}$ is the selected secondary data input-output and Y_{SNR} is the random signal with a desired SNR.

$$Y_{synthetic} = Y_{original} + Y_{SNR} \quad (2)$$

C. Modal Classification

The model development of classification process for *Aquilaria* species based on the agarwood oil is executed by using the kNN classifier. The input output of the data from the secondary dataset will be used as the feature for this model. Before the classification is executed using the kNN classifier, the data needs to be split into training and testing set. The ratio of the split will be 80:20 so a portion of 80% of data will be used for the training set and the remaining 20% will be used for the testing set. Next, the utilization of this kNN classifier will use the MATLAB with several setting as follows:

- The k-value is set in range of 1 until 10.
- The distance metrics will use Euclidian distance.
- The training set will be used to run the model.
- The testing set will be used to predict the class of species.

After the classification had been completed for both training and testing set, the performance of the model needs to be evaluated. The confusion matrix will be used to illustrate the amount of actual and predicted class of species for the testing set. Then using the confusion matrix value, the parameter of accuracy will be measured to showcase the proportion of the correctly classified class of the species.

IV. RESULTS AND DISCUSSIONS

The results are divided into three parts:

A. Sample Preparation
B. Feature Engineering
C. Model Classification

A. Sample Preparation

The sample ground prepation is consist of four agarwood oils from four different species of *Aquilaria*, namely- *Aquilaria beccariana*, *Aquilaria crassna*, *Aquilaria malaccencis*, *Aquilaria subintegra*. This sample is originated from Malaysia and the sample has been prepared by BARCE, UMP using their standard operating procedure (SOP). After the agarwood oils has been acquired using the hydrodistillation process, the GC-FID analysis is used to identify the chemical compounds from each species of *Aquilaria*. From there, 40 samples have been identified from agarwood oil of *Aquilaria beccariana* species, 47 samples from *Aquilaria crassna*, 61 samples from *Aquilaria malaccencis* and 55 samples from *Aquilaria subintegra*. Combining the samples makes the total sample to be 203 samples with 82 identified chemical compounds for all *Aquilaria* species.

Then the all sample has been combine together as one dataset to be used as primary dataset. Based on the primary dataset, 14 compounds has been highlighted to be similar compound that exist across every *Aquilaria* species. This 14 chemical compounds is tabulated in Table II.

B. Feature Engineering

It was noted that not all chemical compounds have the same value for each *Aquilaria* species. So, the selection of significant compounds will be based on similar chemical compounds that exist across every species of *Aquilaria*. There are 14 chemical compounds that possibly exist across every *Aquilaria* species as tabulated at Table II. Using these chemical compounds, a new dimension of data had been set to be 4 by 14 with a total sample of and the input and output has been set as follows:

- Input = 14 similar chemical compounds.
- Output = 4 species of *Aquilaria* (*Aquilaria beccariana*, *Aquilaria crassna*, *Aquilaria malaccencis*, *Aquilaria subintegra*).

For the input, the numbering of each selected compound will be used as follows:

- allo-Aromadendrene compound will be number ‘1’.
- β-Selinene compound will be number ‘2’.
- Dihyro-β-Agarofuran compound will be number ‘3’.
- δ-Guaiene compound will be number ‘4’.
- α-Calacorene compound will be number ‘5’.
- Caryophyllene oxide compound will be number ‘6’.
- Tetradecanal compound will be number ‘7’.
- 10-epi-γ-Eudesmol compound will be number ‘8’.
- inkoh-eremol compound will be number ‘9’.
- Selina-3, 11-dien-9-ol compound will be number ‘10’.
- 9, 11-Eremophiladien-8-one compound will be number ‘11’.
- Selina-3, 11-dien-14-oic acid compound will be number ‘12’.
- Pentadecanoic acid compound will be number ‘13’.
- 2-hydroxyguaia-1(10), 11-dien-15-oic acid compound will be number ‘14’.

For the output, the index numbering of the species will be used as follows:

- *Aquilaria beccariana* is ‘1’.
- *Aquilaria crassna* is ‘2’.
- *Aquilaria malaccencis* is ‘3’.
- *Aquilaria subintegra* is ‘4’.

This setting of input output has been executed to make a secondary data for the classification process. Table III illustrates this secondary data after the setting of their input and ouput. A generation of synthetic data has been applied to mimic the secondary data to increase the sample size. The synthetic data used the formulae as mentioned in subsection III.B. From there, 9 synthetic data have been generated for each species of *Aquilaria*. Table IV tabulates the descriptive statistic of this secondary data against synthetic data, for all *Aquilaria* species, to examine the consistency of the pattern to be mimicking the secondary data. The tables indicates that the synthetic data closely matches the original data for each species. This confirms the reliability of the synthetic data in faithfully mimicking the attributes of the original data. After the synthetic data has been generated, the synthetic data has been combined with the secondary data to be used in the classification process.

C. Model Classification

The model classification of *Aquilaria* species using agarwood oil data has been implemented using kNN classifier. Before the modeling, the data has been split first into training and testing set using the combination of synthetic data and secondary data achieve from subsection IV.B. The dimensions of the new secondary data are 40 by 14 with a class of four species making the total sample to be 560 sample. Then, the data has been split into a ratio of 80:20, where 80 percent of the secondary data is used for training set, while remaining 20 percent of the secondary data is used for testing set. Here, a total of 448 sample is used for training set and 112 sample is used for testing set.

After the execution of the model, the performance of it has been evaluated using confusion matrix based on the testing set. Fig. 4 illustrates the confusion matrix of the kNN model for the testing set. Based on the confusion matrix, the 12.5% of the data has been classified as *Aquilaria beccariana*, 25.0% of the data has been classified as *Aquilaria crassna* and *Aquilaria malaccencis* and 37.5% of the data has been classified as *Aquilaria subintegra*.

TABLE II. LIST OF SIMILAR CHEMICAL COMPOUNDS THAT EXIST ACROSS EVERY *AQUILARIA* SPECIES

INDEX	COMPOUND NAME	*AQUILARIA* SPECIES ABUNDANCES			
		AQUILARIA BECCARIANA	*AQUILARIA CRASSNA*	*AQUILARIA MALACCENCIS*	*AQUILARIA SUBINTEGRA*
1	allo-Aromadendrene	1.89	14.48	1.32	13
2	β-Selinene	0.66	0.11	0.54	0.37
3	Dihyro-β-Agarofuran	1.25	0.49	0.55	0.43
4	δ-Guaiene	0.71	0.21	2.03	0.34
5	α-Calacorene	0.13	0.25	0.31	0.32
6	Caryophyllene oxide	0.39	2.24	1.28	1.65
7	Tetradecanal	0.13	0.33	1.33	0.39
8	10-epi-γ-Eudesmol	0.33	2.52	6.58	2.13
9	Jinkoh-eremol	0.16	8.3	1.25	0.37
10	Selina-3, 11-dien-9-ol	0.35	0.3	0.52	2.23
11	9, 11-Eremophiladien-8-one	0.29	1.95	1.87	0.78
12	Selina-3, 11-dien-14-oic acid	1.6	7.06	0.97	4.09
13	Pentadecanoic acid	0.2	0.14	0.18	0.35
14	2-hydroxyguaia-1(10), 11-dien-15-oic acid	0.59	0.55	3.72	0.93

TABLE III. LIST OF SECONDARY DATA FOR SELECTED CHEMICAL COMPOUNDS AGAINST THEIR SPECIES OF *AQUILARIA*

CHEMICAL COMPOUNDS														SPECIES OF *AQUILARIA*
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1.89	0.66	1.25	0.71	0.13	0.39	0.13	0.33	0.16	0.35	0.29	1.60	0.20	0.59	1
14.48	0.11	0.49	0.21	0.25	2.24	0.33	2.52	8.30	0.30	1.95	7.06	0.14	0.55	2
1.32	0.54	0.55	2.03	0.31	1.28	1.33	6.58	1.25	0.52	1.87	0.97	0.18	3.72	3
13.00	0.37	0.43	0.34	0.32	1.65	0.39	2.13	0.37	2.23	0.78	4.09	0.35	0.93	4

*Note: Column 1 = allo-Aromadendrene compound, Column 2 = β-Selinene compound, Column 3 = Dihyro-β-Agarofuran compound, Column 4 = δ-Guaiene compound, Column 5 = α-Calacorene compound, Column 6 = Caryophyllene oxide compound, Column 7 = Tetradecanal compound, Column 8 = 10-epi-γ-Eudesmol compound, Column 9 = Jinkoh-eremol compound, Column 10 = Selina-3, 11-dien-9-ol compound, Column 11 = 9, 11-Eremophiladien-8-one compound, Column 12 = Selina-3, 11-dien-14-oic acid compound, Column 13 = Pentadecanoic acid compound, Column 14 = 2-hydroxyguaia-1(10), 11-dien-15-oic acid compound; Species of *Aquilaria*: 1 = *Aquilaria beccariana*, 2 = *Aquilaria crassna*, 3 = *Aquilaria malaccencis*, 4 = *Aquilaria subintegra*.

TABLE IV. DESCRIPTIVE STATISTICS OF SECONDARY DATA FOR COMPARISON OF ORIGINAL DATA (OD) AGAINST SYNTHETIC DATA (SD)

SPECIES OF *AQUILARIA*	DESCRIPTIVE STATISTICS (VALUE)	OD	SD1	SD2	SD3	SD4	SD5	SD6	SD7	SD8	SD9
AQUILARIA BECCARIANA	MIN	0.13	0.15	0.09	0.09	0.12	0.16	0.11	0.16	0.09	0.13
	MAX	1.89	1.87	1.92	1.86	1.88	1.91	1.90	1.90	1.85	1.90
	MEDIAN	0.37	0.39	0.37	0.37	0.34	0.38	0.39	0.38	0.37	0.33
AQUILARIA CRASSNA	MIN	0.11	0.00	0.05	0.06	0.01	0.04	0.02	0.11	0.08	0.00
	MAX	14.48	14.43	14.50	14.39	14.65	14.70	14.39	14.49	14.28	14.26
	MEDIAN	0.52	0.64	0.69	0.60	0.52	0.56	0.53	0.53	0.54	0.53
AQUILARIA MALACCENCIS	MIN	0.18	0.06	0.15	0.19	0.22	0.27	0.10	0.16	0.24	0.10
	MAX	6.58	6.54	6.46	6.54	6.57	6.47	6.59	6.71	6.55	6.53
	MEDIAN	1.27	1.26	1.27	1.29	1.17	1.31	1.22	1.21	1.28	1.35
AQUILARIA SUBINTEGRA	MIN	0.32	0.24	0.06	0.24	0.26	0.24	0.28	0.23	0.21	0.12
	MAX	13.00	12.95	13.07	13.02	12.80	13.01	13.01	12.97	12.88	13.16
	MEDIAN	0.61	0.69	0.68	0.73	0.70	0.73	0.74	0.65	0.59	0.66

The actual and predicted species of the *Aquilaria* for the testing set has been recorded in Table V, based on the confusion matrix from Fig. 4. From Table V, the total number of samples for the class of actual and predicted species of *Aquilaria* is 8. Here, the error recorded to be 0.0% for every samples due to no missinterpretations of actual and predicted species of *Aquilaria* during the testing period. Based on these, it can be concluded that the accuracy of the model to be 100% (accuracy = 100%) since the model does not have any single mistake during the classification process.

Fig. 4. Confusion matrix for the testing phase.

TABLE V. LIST OF ACTUAL AND PREDICTED SPECIES OF *AQUILARIA* DURING TRAINING PHASE FOR THE FIRST DATASET

NO.OF SAMPLE	ACTUAL	PREDICTED	ERROR (%)
I	2	2	0.0
II	4	4	0.0
III	3	3	0.0
IV	4	4	0.0
V	1	1	0.0
VI	4	4	0.0
VII	2	2	0.0
VIII	3	3	0.0

*Note: Label '1' is *Aquilaria beccariana*, Label '2' is *Aquilaria crassna*, Label '3' is *Aquilaria malaccencis*, Label '4' is *Aquilaria subintegra*

V. CONCLUSION

The study demonstrated the succces in classifying *Aquilaria* species using kNN classifier without any misinterpretation with an accuracy of 100% (accuracy = 100%). Based on the 14 similar compound, those compounds of *Aquilaria* can be used as a new marker for futher research in classification of *Aquilaria* species, *Aquilaria* industry and its related research area.

ACKNOWLEDGMENT

The authors would like to express their gratitude and acknowledgement to the Institute of Research Management and Innovation (IRMI), Universiti Teknologi MARA (UiTM) campus Shah Alam for the financial support given through the 'Geran Insentif Penyeliaan' (GIP). Special recognition is extended to the College of Engineering, School of Electrical Engineering, UiTM for the opportunity to conduct and complete this research.

References

[1] "Plants of the World Online | Kew Science." https://powo.science.kew.org/ (accessed Aug. 09, 2023).

[2] S. Y. Lee and R. Mohamed, "The Origin and Domestication of Aquilaria, an Important Agarwood-Producing Genus BT - Agarwood: Science Behind the Fragrance," R. MOHAMED, Ed. Singapore: Springer Singapore, 2016, pp. 1–20. doi: 10.1007/978-981-10-0833-7_1.

[3] M. F. Elias, H. Ibrahim, and W. R. W. Mahamod, "A Review on the Malaysian Aquilaria species in Karas Plantation and Agarwood Production," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 7, no. 4, May 2017, doi: 10.6007/ijarbss/v7-i4/2911.

[4] S. Y. Lee and R. Mohamed, "The Origin and Domestication of Aquilaria, an Important Agarwood-Producing Genus," pp. 1–20, 2016, doi: 10.1007/978-981-10-0833-7_1.

[5] A. López-Sampson and T. Page, "History of Use and Trade of Agarwood," *Econ. Bot.*, vol. 72, no. 1, pp. 107–129, 2018, doi: 10.1007/s12231-018-9408-4.

[6] R. Mohamed and S. Y. Lee, "Keeping Up Appearances: Agarwood Grades and Quality," R. MOHAMED, Ed. Singapore: Springer Singapore, 2016, pp. 149–167. doi: 10.1007/978-981-10-0833-7_10.

[7] H. Chhipa, K. Chowdhary, and N. Kaushik, "Artificial production of agarwood oil in Aquilaria sp. by fungi: a review," *Phytochem. Rev.*, vol. 16, no. 5, pp. 835–860, 2017, doi: 10.1007/s11101-017-9492-6.

[8] C. Andary, D. Longepierre, K. Le Cong, S. Hul, A. Zaremski, and G. Michaloud, "Study of a chemotaxonomic marker able to identify the genus aquilaria (Thymelaeaceae)," *Bois Forets des Trop.*, vol. 341, no. 3, pp. 29–38, 2019, doi: 10.19182/bft2019.341.a31744.

[9] S. A. Razak and N. W. Haron, "Phytosociology of aquilaria malaccensis lamk. And its communities from a tropical forest reserve in peninsular Malaysia," *Pakistan J. Bot.*, vol. 47, no. 6, pp. 2143–2150, 2015.

[10] W. D. Nugroho, R. Pujiarti, and N. P. Tiyasa, "Wood anatomical characteristics of agarwood-producing species (Aquilaria sp.)," *Wood Res.*, vol. 64, no. 5, pp. 759–768, 2019.

[11] Lukman, D. Dinarti, U. J. Siregar, M. Turjaman, and Sudarsono, "Characterization and identification of agarwood-producing plants (Aquilaria spp.) from North Aceh, Indonesia, based on morphological and molecular markers," *Biodiversitas*, vol. 23, no. 9, pp. 4861–4871, 2022, doi: 10.13057/biodiv/d230955.

[12] 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)," *2015 Int. Conf. Smart Sensors Appl. ICSSA 2015*, pp. 7–10, 2015, doi: 10.1109/ICSSA.2015.7322500.

[13] R. Naef, "The volatile and semi-volatile constituents of agarwood, the infected heartwood of Aquilaria species: A review," *Flavour and Fragrance Journal*, vol. 26, no. 2. pp. 73–87, Mar. 2011. doi: 10.1002/ffj.2034.

[14] Y. Liu, J. Wei, Z. Gao, Z. Zhang, and J. Lyu, "A Review of Quality Assessment and Grading for Agarwood," *Chinese Herb. Med.*, vol. 9, no. 1, pp. 22–30, 2017, doi: 10.1016/s1674-6384(17)60072-8.

[15] N. S. Ismail, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, and S. N. Tajuddin, "Polynomial tuned kernel parameter in SVM of agarwood oil for quality classification," *Proc. - 2018 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2018*, no. October, pp. 77–82, 2019, doi: 10.1109/I2CACIS.2018.8603686.

[16] N. Nazir, A. R. Mohd-hairul, and T. S. Hian, "Representational Difference Analysis (RDA) for Identification of DNA Markers from Aquilaria hirta Abstract :," no. December, 2021.

[17] M. Artini *et al.*, "Antimicrobial and antibiofilm activity and machine learning classification analysis of essential oils from different mediterranean plants against pseudomonas aeruginosa," *Molecules*, vol. 23, no. 2, 2018, doi: 10.3390/molecules23020482.

[18] K. A. Athirah, N. Ismail, M. N. Taib, N. A. M. Ali, M. Jamil, and S. Lias, "Modelling of cymbopogon oils species using k-Nearest neighbours (k-NN)," *Proceeding - 2019 IEEE 7th Conf. Syst. Process Control. ICSPC 2019*, no. December, pp. 5–9, 2019, doi: 10.1109/ICSPC47137.2019.9068086.

[19] M. Artini *et al.*, "Essential Oils Biofilm Modulation Activity and Machine Learning Analysis on Pseudomonas aeruginosa Isolates from Cystic Fibrosis Patients," *Microorganisms*, vol. 10, no. 5, pp. 1–20, 2022, doi: 10.3390/microorganisms10050887.

[20] M. A. Aiman Ngadilan, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. Mohd Ali, and S. N. Tajuddin, "Radial basis function (RBF) tuned kernel parameter of agarwood oil compound for quality classification using support vector machine (SVM)," in *2018 9th IEEE Control and System Graduate Research Colloquium, ICSGRC 2018 - Proceeding*, Mar. 2019, pp. 64–68. doi: 10.1109/ICSGRC.2018.8657524.

[21] N. Ismail, M. A. Nor Azah, M. Jamil, M. H. F. Rahiman, S. N. Tajuddin, and M. N. Taib, "Analysis of high quality agarwood oil chemical compounds by means of SPME /GC-MS and Z-score technique," *Malaysian J. Anal. Sci.*, vol. 17, no. 3, pp. 403–413, 2013, [Online]. Available: http://umpir.ump.edu.my/6304/1/fist-2013-saiful-analysis_of_high_quality.pdf

[22] M. H. Haron, "Grading Of Agarwood Oil Quality Based On Its Chemical Compounds Using Self Organizing Map (SOM)," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 7, pp. 3728–3736, 2020, doi: 10.30534/ijeter/2020/135872020.

[23] X. Ding *et al.*, "Genome sequence of the agarwood tree Aquilaria sinensis (Lour.) Spreng: the first chromosome-level draft genome in the Thymelaeceae family," *Gigascience*, vol. 9, no. 3, pp. 1–10, 2020, doi: 10.1093/gigascience/giaa013.

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

[25] J. M. R. Alamil *et al.*, "Rediscovering the Therapeutic Potential of Agarwood in the Management of Chronic Inflammatory Diseases," *Molecules*, vol. 27, no. 9. 2022. doi: 10.3390/molecules27093038.

[26] N. C. Radzi and F. A. Kasim, "Effect of microwave pretreatment on gaharu essential oil using hydrodistillation method," *Indones. J. Chem.*, vol. 20, no. 4, pp. 960–966, 2020, doi: 10.22146/ijc.43191.

[27] T. B. Yücel *et al.*, "Analysis of Volatile Constituent by Hydrodistillation and Solid-Phase Microextraction Techniques and Antimicrobial and Scolicidal Activities of Essential Oil and Soxhlet Extracts of Ulva rigida grown in Turkey," *Chem. Biodivers.*, vol. 20, no. 9, p. e202300965, Aug. 2023, doi: 10.1002/cbdv.202300965.

[28] N. Ismail, N. A. M. Ali, M. Jamil, M. H. F. Rahiman, S. N. Tajuddin, and M. N. Taib, "A review study of agarwood oil and its quality analysis," 2014. doi: 10.11113/jt.v68.2419.

[29] S. N. Tajuddin, C. M. Aizal, and M. M. Yusoff, "Resolution of Complex Sesquiterpene Hydrocarbons in Aquilaria malaccensis Volatile Oils Using Gas Chromatography Technique," pp. 103–124, 2016, doi: 10.1007/978-981-10-0833-7_7.

[30] Q. Li *et al.*, "Molecular identification of three aquilaria (thymelaeaceae) species through DNA barcoding," *Biol. Pharm. Bull.*, vol. 41, no. 6, pp. 967–971, 2018, doi: 10.1248/bpb.b18-00050.

[31] M. Ishihara, T. Tsuneya, and K. Uneyama, "Fragrant sesquiterpenes from agarwood," *Phytochemistry*, vol. 33, no. 5, pp. 1147–1155, 1993.

[32] W.-L. Mei *et al.*, "Characterization and determination of 2-(2-phenylethyl) chromones in agarwood by GC-MS," *Molecules*, vol. 18, no. 10, pp. 12324–12345, 2013.

[33] M. Ishihara, T. Tsuneya, and K. Uneyama, "Guaiane sesquiterpenes from agarwood," *Phytochemistry*, vol. 30, no. 10, pp. 3343–3347, 1991.

[34] 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," *J. Essent. Oil Res.*, vol. 23, no. 4, pp. 46–53, 2011.

[35] M. Ishihara, T. Tsuneya, and K. Uneyama, "Components of the volatile concentrate of agarwood," *J. Essent. Oil Res.*, vol. 5, no. 3, pp. 283–289, 1993.

[36] H. I. Al-Jaber *et al.*, "Patterns in Volatile Emission of Different Organs of Inula viscosa growing wild in Jordan," *J. Essent. Oil-Bearing Plants*, vol. 20, no. 1, pp. 24–35, 2017, doi: 10.1080/0972060X.2016.1277167.

[37] R. Costa *et al.*, "GC–MS, GC–O and enantio–GC investigation of the essential oil of Tarchonanthus camphoratus L.," *Flavour Fragr. J.*, vol. 23, no. 1, pp. 40–48, 2008.

[38] M. A. Rather *et al.*, "Comparative GC–FID and GC–MS analysis of the mono and sesquiterpene secondary metabolites produced by the field grown and micropropagated plants of Artemisia amygdalina Decne," *Acta Physiol. Plant.*, vol. 34, pp. 885–890, 2012.

[39] R. Aparicio-Ruiz, D. L. García-González, M. T. Morales, A. Lobo-Prieto, and I. Romero, "Comparison of two analytical methods validated for the determination of volatile compounds in virgin olive oil: GC-FID vs GC-MS," *Talanta*, vol. 187, pp. 133–141, 2018, doi:

10.1016/j.talanta.2018.05.008.
[40] N. M. Sarih, N. Ismail, N. Akhtar, and S. N. Tajuddin, "Analysis of GC-FID and GC-MS Microwave-Assisted Hydrodistillation Extraction (MAHD) of Agarwood Chips," *Int. J. Integr. Eng.*, vol. 13, no. 6, pp. 180–189, 2021, doi: 10.30880/ijie.2021.13.06.017.
[41] Y. Kusuma Arbawa and C. Dewi, "Soil Nutrient Content Classification for Essential Oil Plants using kNN," no. Iceo 2019, pp. 96–100, 2020, doi: 10.5220/0009957400960100.
[42] A. Agarwal, P. Sharma, M. Alshehri, A. A. Mohamed, and O. Alfarraj, "Classification model for accuracy and intrusion detection using machine learning approach," *PeerJ Comput. Sci.*, vol. 7, pp. 1–22, 2021, doi: 10.7717/PEERJ-CS.437.
[43] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," *J. Adv. Res.*, vol. 4, no. 4, pp. 331–344, 2013, doi: 10.1016/j.jare.2012.05.007.
[44] M. E. M. Samad, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, and S. N. Tajuddin, "Analysis of distance metric variations in KNN for agarwood oil compounds differentiation," *Proc. - 2017 IEEE Conf. Syst. Process Control. ICSPC 2017*, vol. 2018-Janua, no. December, pp. 151–156, 2017, doi: 10.1109/SPC.2017.8313038.
[45] BioAromatic Research Centre, "Consultation For Agarwood Oil Quality Assessment (Aquilaria Beccariana)," 2022.
[46] BioAromatic Research Centre, "Consultation For Agarwood Oil Quality Assessment (Aquilaria Crassna)," 2022.
[47] BioAromatic Research Centre, "Consultation For Agarwood Oil Quality Assessment (Aquilaria Malaccensis)," Mar. 2022.
[48] BioAromatic Research Centre, "Consultation For Agarwood Oil Quality Assessment (Aquilaria Subintegra)," 2022.
[49] M. S. Wisz *et al.*, "Effects of sample size on the performance of species distribution models," *Divers. Distrib.*, vol. 14, no. 5, pp. 763–773, 2008.
[50] L. Sevgi, "Synthetic radar-signal environment: Computer generation of signal, noise, and clutter," *IEEE Antennas Propag. Mag.*, vol. 49, no. 5, pp. 192–198, 2007.