SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

KNN Euclidean Distance Model Performance on *Aquilaria Malaccensis* Oil Qualities

Aqib Fawwaz Mohd Amidon[1], Zakiah Mohd Yusoff[2,*], Nurlaila Ismail[1], Ali Abd Almisreb[3], Mohd Nasir Taib[1]

[1] School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
[2] Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA Kampus Pasir Gudang, Johor, Malaysia
[3] Faculty Computer Science and Engineering, International University of Sarajevo, Sarajevo, Bosnia and Herzegovina

ARTICLE INFO

Article history:
Received 25 November 2023
Received in revised form 7 June 2024
Accepted 15 June 2024
Available online 15 July 2024

Keywords:
Agarwood grading; machine learning, artificial intelligence; k-Nearest Neighbor (KNN) technique

ABSTRACT

Agarwood is a highly prized and useful forest product. In Southeast Asia, *Aquilaria Malaccensis* species are typically the most prevalent. This agarwood is usually used in the manufacture of medicine, the production of high-quality perfumes, and is also used in religious and ethnic ceremonies. According to the study, the agarwood grading process entirely relies on human senses. The graders will evaluate the agarwood oil's color concentration with their unaided eyes and evaluate the amount of scent emitted with their noses. This approach has been proven to have several limitations, including that the graders' health will suffer, the grading procedure will take a very long time, and will consume high operating expenses. Therefore, an established standard grading model that is faster, easier, and more accurate needs to be introduced. Previous researchers found that chemical compounds contained in agarwood oil can be used to grade the quality of agarwood oil. Therefore, this study has used the data obtained that contains significant chemical compounds as input to develop a grading model with the support of machine learning and artificial intelligence, which is the k-Nearest Neighbors (KNN) technique. The output of this grading model is the classification of agarwood oil according to its quality, which includes four different grades. The results of the implementation of KNN grading of this model found that this model has very excellent performance by obtaining 100% for each measurement for the performance evaluator of the classifier.

1. Introduction

The resinous heartwood of diseased trees of the evergreen *Aquilaria Malaccensis* (Thymelaeaceae) is known as agarwood [1-4]. Normally this species of Agarwood can be found in Southeast Asia [1,3,4]. Agarwood has various benefits for humans. This is proven when there are discoveries related to its use in old manuscripts such as Sahih Muslim, Charaka Samhita, Torah, and Bhagavad Gita [5]. Before performing worship, every Muslim will usually apply agarwood oil, also

* *Corresponding author.*
E-mail address: zakiah9018@uitm.edu.my

https://doi.org/10.37934/araset.48.2.1628

known as "Minyak Attar", to their clothes. This is because wearing fragrance is one of the Sunnah of Prophet Muhammad S.A.W. For Buddha believers, the scent of Nirvana is created by the aroma of burning agarwood [5].

Next, in Japan's history, before going to war, Samurai warriors would scent their amour with agarwood smoke to get good luck [5]. These are three examples of uses involving agarwood. In general, this agarwood is used to produce high-quality perfume oil, medicinal production, and use in worship events of a religion or ethnicity [1,3,6-11]. Also, in the past, it was considered a symbol of wealth for a kingdom or empire [1,5].

Although it has been used for a long time, the quality grading process of agarwood still maintains the old method [1,3,4,8-10]. Those who are experts at agarwood trees are indispensable in grading the quality of agarwood. First, they will measure the concentration level of the resulting agarwood colour [1]. Secondly, they will evaluate the strength of the smell of the agarwood either by burning it or smelling it directly [3]. This process only uses their senses, such as their eyes and nose. Usually, manufacturers of agarwood products such as agarwood oil and so on will obtain agarwood raw materials in large quantities for each delivery batch. Certainly, the process of evaluating one by one is very difficult because the specialist workforce is limited [1,3,5].

With this kind of situation, there can be some shortcomings in maintaining this grading method. Firstly, performing grading directly in large quantities will affect the health of those involved [1,3]. Second, the grading process requires a relatively long period of time since there are not many skilled workers [6-8]. Third, the long grading process will incur high operating costs. Fourth, training new employees in the intricacies of agarwood takes a long time and costs a lot [5]. Fifth, the grading through human sense is a little doubtful and cannot give a high guarantee regarding the accuracy of the grading made [1].

Because of that, a new grading method needs to be introduced. Previous research found a discovery that can achieve this goal. The discovery is related to the chemical compounds contained in agarwood oil [1,11]. For this research, the sample used is from the type of *Aquilaria Malaccensis* species. So, the chemical compounds contained in agarwood oil can be used to determine the level of quality of agarwood. The idea is to use artificial intelligence methods that involve machine learning. Some of the suggested techniques are Support Vector Machine (SVM), Multilayer Perception (MLP), Self-Organizing Map (SOM), Artificial Neural Network (ANN), k-Nearest Neighbors(KNN), and more [1,6-11]. All the techniques that are suggested are suitable because all these techniques are developed to make classification either for linear data or non-linear data or both. Each of these techniques also has its own advantages. But for this study, the k-Nearest Neighbors (KNN) technique was chosen. This is because the KNN technique is easier to understand its concept and easier to implement in developing the grading model compared to other techniques [12-16]. Moreover, the objective of this study is to create an easy, fast, more efficient, and accurate grading model of agarwood oil qualities.

In summary, the dataset used involves 660 samples of agarwood oil that includes eleven significant chemical compounds and has four different qualities (high, medium high, medium low, and low quality). This dataset will be used in developing the KKN grading method of agarwood oil qualities.

2. Theoretical Work

2.1 k-Nearest Neighbors (KNN) Technique

One of the recommended methods for pattern classifiers is k-Nearest Neighbors (KNN) [17,18-25]. It was discovered that the KNN offers a trustworthy result in categorization, classification, and

recognition, for examples in ECG databases [26], English and Arabic articles [18], textual information [21], network traffic [20], transformer windings [27], and EEG signals [22,28]. Therefore, this technique is both straightforward and effective [21]. If it functions as a classifier, this is because its training phase stores all the training data. Because the decision-making is based on a generalization technique outside of the training data immediately before a new instance is classified, it is also known as a lazy learner [21].

In 1991, Dasarathy developed the first theory behind the KNN algorithm [29]. In a complex dimensional space, it is described as the categorization of an item in a test data set based on the class of its k-Nearest Neighbors as shown in Figure 1 [30]. It implies that the classification is carried out utilizing the assignment of an example to the class related to its k-Nearest Neighbors [31]. The training and testing phases are considered vital aspects of implementing the KNN algorithm, encompassing tasks such as data preprocessing, parameter optimization, and model evaluation to ensure robust and accurate classification results [18,26,29]. The class labels and attribute vectors of training samples in a multidimensional space are kept during the training phase. Assigning a label that is frequently repeated among the k (number of Nearest Neighbors selected by the user) training examples closest to the test point is how a test example is categorized in the testing stage [26]. All training test points are separated from this test point by a computed distance, and the closest distance is chosen [18]. Several distance measures must be followed when dealing with KNN. Some of them are Euclidean, city block, and correlation [26]. Figure 2 shows the hierarchy of KNN techniques followed by how KNN works in term of closest neighbors' rule in Figure 3.

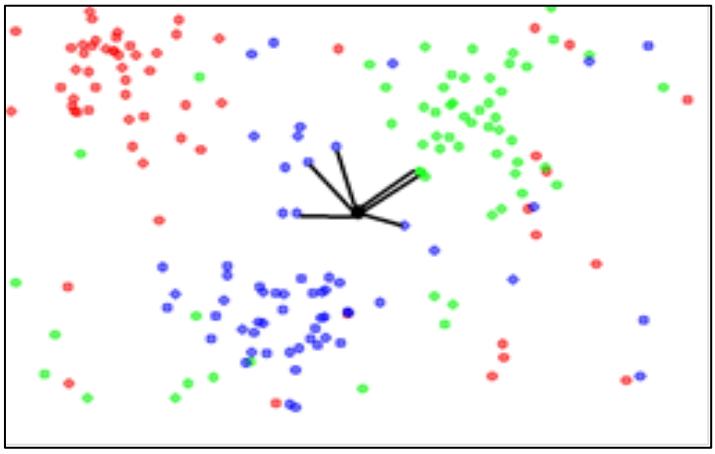

Fig. 1. k-Nearest Neighbors (KNN) technique works [18]

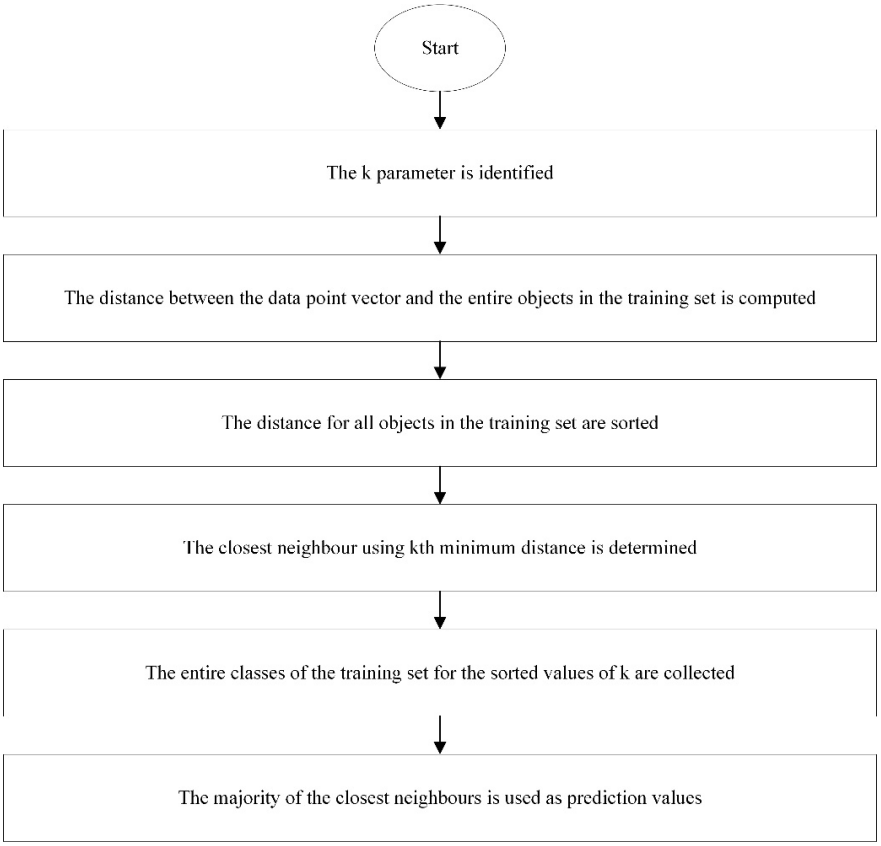

Fig. 2. The hierarchy of the KNN technique

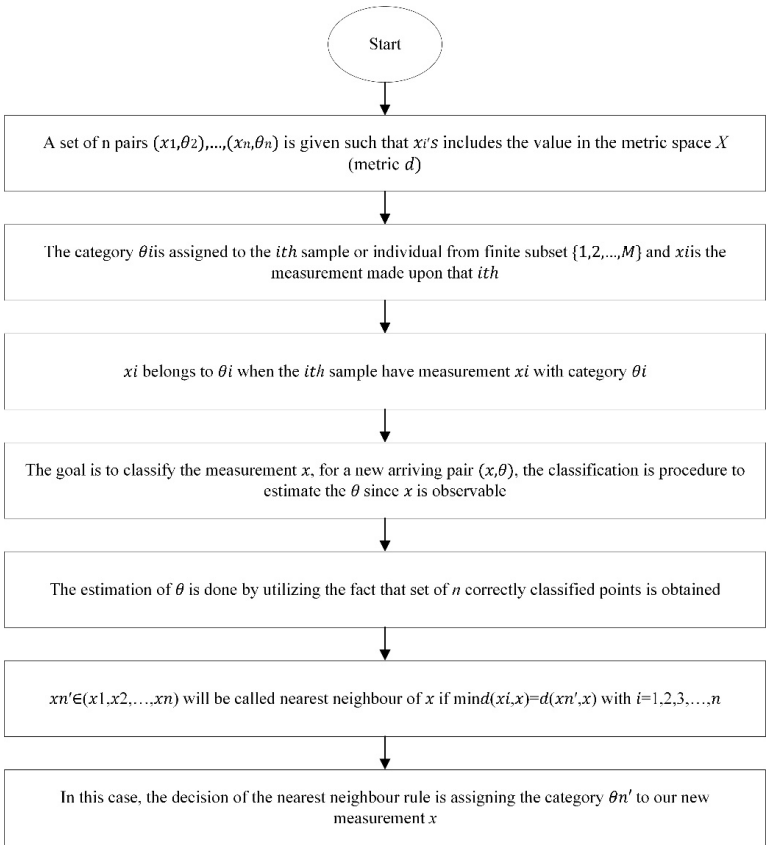

Fig. 3. KNN works in term of closest neighbors' concept

It is necessary to modify the distance measure before KNN execution. A variety of distance metrics, including city block, correlation, and Euclidean, can be applied. The distance metric used in this investigation was the Euclidean. It is determined by using the square root discrepancies between two objects coordinates as shown in Figure 4.

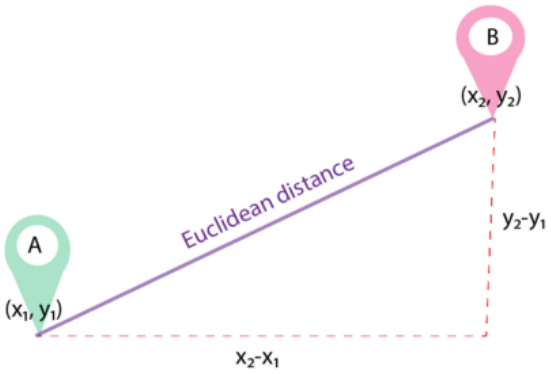

Fig. 4. Euclidean distance based on Pythagorean theorem [26]

The formula for Euclidean distance is displayed in equation (1) [26].

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

where x_{sj} is an object at coordinate s_j and y_{tj} is another object at coordinate t_j and d_{st} is a distance between them.

KNN has been employed as a classifier in numerous investigations on agarwood oil [32-34]. Both the sensor intensities of chemical compounds received from the E-Nose [34] and the abundances of chemical compounds derived from the GC-MS analysis [33,35] were used as inputs in these experiments. Euclidean distance was utilized in all experiments [33-35] as a distance metric for KNN parameters, and the chosen number of nearest neighbors was 5 [34] and 20 [33]. Training and testing data sets were split 80:20 [34] and 70:30 [32] respectively. These research [33-35] have successfully classified agarwood oil into high and low grade with accuracy ranging from 80% to 100%.

2.2 *Confusion Matrix*

A confusion matrix is made up of the categorization of that specific class stated as a matrix column and the actual class expressed as a matrix row. The frequency of a class i object assigned to a class j is provided by the element M*[i][j]*. True classifications can be shown in the diagonal elements. The total row i in an unnormalized matrix equals the total number of elements from class i that are shown in the data set. [35]. Table 1 presents a confusion matrix for binary forecast.

Table 1
4x4 Confusion matrix classification

	Predicted				
Actual	class I	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*
	class II	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*
	class III	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*	False negative *(fn)*
	class IV	true negative *(tn)*	true negative *(tn)*	true negative *(tn)*	true negative *(tn)*

**Examples calculate for class IV*

By counting the number of class samples that were correctly identified, true positive (*tp*) is evaluated. The number of accurately categorized samples that are not included in the class is known as the true negative (*tn*). False positives are samples that were misclassified into the category. Samples that were not designated as class samples are known as false negatives (*fn*) [36,37].

Accuracy, precision, sensitivity, and specificity are four measurements that can be assessed using the confusion matrix. The ultimate success of a classifier is accuracy (acc). It can be computed using equation (2) [37].

$$acc = \frac{tp + tn}{tp + fn + fp + tn} \tag{2}$$

The class repeatability of the data labels with the positive labels is known as precision (prec). The equation is displayed in equation (3) [37]

$$prec = \frac{tp}{tp + fp} \tag{3}$$

The ability of a classifier to identify positive labels is known as sensitivity (sens). The equation is shown in equation (4) [37].

$$sens = \frac{tp}{tp + fn} \tag{4}$$

The ability of a classifier to identify false positives is known as specificity (spec). The calculation is a representation of the equation (5) [37].

$$spec = \frac{tn}{fp + tn} \tag{5}$$

3. Methodology

The process of developing the KNN grading model of agarwood oil qualities in stages is explained in Figure 5. Essentially, this procedure is divided into three steps. Data preprocessing is used to prepare the data set in the first stage. The second stage concentrates on the KNN grading model development. The evaluation of the model grade that has been developed is the third and last stage.

3.1 Data Preprocessing

Data normalization, data randomization, and data division have all been part of this data preprocessing. To determine the quantity of chemical components, the raw sample (agarwood oil) was first subjected to a GC-MS analysis. The content data was then transformed into digital data for further purpose. The 22 samples used contained 103 different chemical compounds, according to the results of the GC-MS analysis. The 22 samples data information was insufficient. Therefore, significant chemical compounds of agarwood oil quality were obtained using Principal Component Analysis (PCA).

Out of 103 chemical compounds, only 11 were determined to be significant using PCA analysis. Due to the greater cost of agarwood oil samples and the limited sample size of the genuine data, which can result in a failure in the suitability test for factor analysis, artificial data was manufactured using a technique known as MUNGE. MUNGE was approached to generate the artificial data to overcome the higher cost of agarwood oil samples and the limitation of actual data samples issues. As a result, 660 samples in total were generated as artificial data consisting of eleven chemical compounds and classified into four qualities (high, medium high, medium low, and low quality) referred to the real data.

The data was split into training and testing datasets using a "Holdout" partition with an 80:20 ratio as recommended by previous researcher, before moving on to the next phase. The reason is that the training dataset (80 percent of total data) is used to develop the model, and the testing dataset (20 percent of total data) is used to test the developed model.

3.2 Model Development

The classification method employed was known as k-Nearest Neighbors (KNN). The inputs for training and testing were obtained from data preparation as mentioned before. Next, the KNN settings were established according to the classification rules listed below, including the number of nearest neighbors, distance metric, and tie-breaking algorithm:

- Input: eleven significant chemical compounds.
- Output: grade of Agarwood oil (high, medium high, medium low, and low quality).
- Number of nearest neighbors (k): 5 [38-40].
- Distance metric: Euclidean distances [38-40].
- Tie breaking algorithm: random [39,40].

For each nearest neighbor’s number, KNN data were trained, tested, and validated.

3.3 Confusion Matrices

After that, the accuracy, precision, sensitivity, and specificity of the KNN training, testing, and validation results were computed using confusion matrices. Starting with the production of a 4x4 confusion matrix table, followed by the calculation of accuracy, specificity, and sensitivity, and ending with calculation of precision. The evaluation of the developed model will determine if it is acceptable or needs to be improved based on the value of the outcome of this computation as shown in Figure 5.

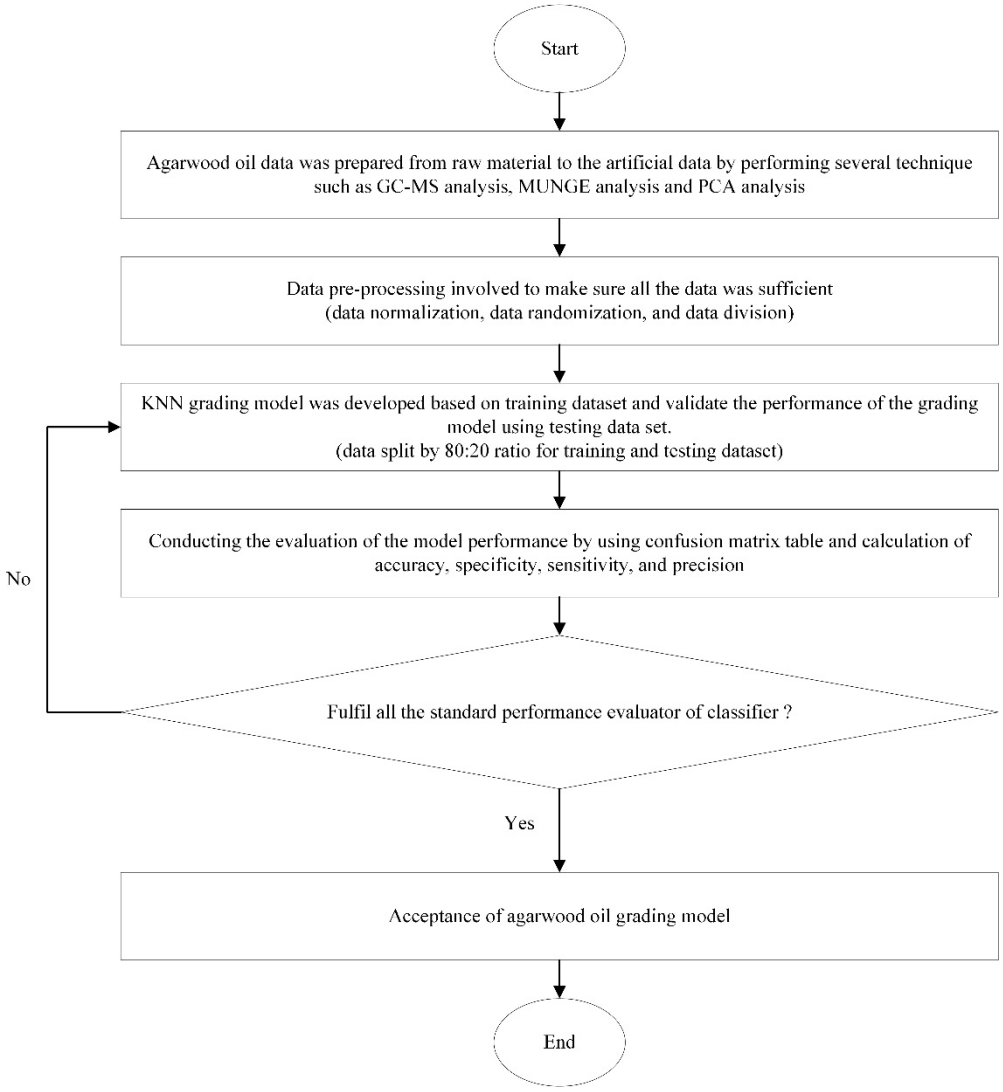

Fig. 5. The process of developing the KNN grading model of agarwood oil qualities

4. Results and Discussion

Based on Agarwood samples and chemical substances that are represented by percentages of abundances, the GC-MS results were tallied and analyzed. The dimensions of the GC-MS data were 22 by 103. There were 22 samples of agarwood oil and 11 significant chemical compounds after data preprocessing process. The agarwood oil samples were given the following designations: CKE, CM,

EO2, EO3, EO4, HD, HG, JBD, KB, LA, LG, M, MA, MA1, MA2, MN, MNS, MPE, MS, R5, RG, and T. Based on some past published works, Table 2 shows the Agarwood oil grades for the samples indicated.

Table 2
Agarwood oil grades for 22 samples according to literatures

References	Grade	Agarwood Oil Samples
[41,42]	High	MA2, JBD
[43]	High	CM, EO2, E03, EO4, HG, JBD, KB, LA, LG, M, MA, MA1, MA2, MN, MNS, MPE, MS, RG, T
	Low	CKE, HD, R5
[44]	A Supreme, Deluxe	JBD, KB, MA, MA1, MA2, T

Next, the 11 significant chemical compounds that were selected are stated in Table 3. All these 11 significant chemical compounds were obtained through the implementation of Principal Component Analysis (PCA) as stated in the Methodology section.

Table 3
Names and retention indices of all eleven significant chemical compounds after the PCA analysis

No.	Name of Significant Chemical Compounds	Retention Index
1	10-epi-γ-eudesmol	1621
2	α-agarofuran	1547
3	β-agarofuran	1474
4	γ-eudesmol	1635
5	dihydrocollumellarin	1962
6	valerianol	1656
7	ar-curcumene	1479
8	β-dihydro agarofuran	1502
9	α-guaiene	1437
10	allo aromadendrene epoxide	1639
11	γ-cadinene	1513

There were 660 samples that were used as inputs to develop this grading model. Table 4 lists the number of each quality implemented throughout this process. Out of 660 samples, 330 samples were grouped as high quality (50%), 30 samples represented the medium high-quality group (4.55%), 90 samples were grouped as medium low quality (13.64%), and the rest, with 210 samples, were from the low quality group (31.83%).

Table 4
Data quantity based on the qualities

Qualities	Quantity	Percentages (%)
High	330	50.00
Medium High	30	4.55
Medium Low	90	13.64
Low	210	31.82

Using the 80:20 ratio as described, a total of 528 random samples were taken as training data to develop the KNN grading model. The model has been successfully developed. Then 132 samples were used to test this model. The testing process is conducted by making a grading comparison between actual data and predicted data. Based on the results of the comparison between actual data and predicted data during this testing process, a 4x4 confusion matrix table was created. Figure 6 shows the confusion matrix table that compares the grading process between actual and predicted data. Through the confusion matrix table, there is no grading error for predicted data comparing to actual data used.

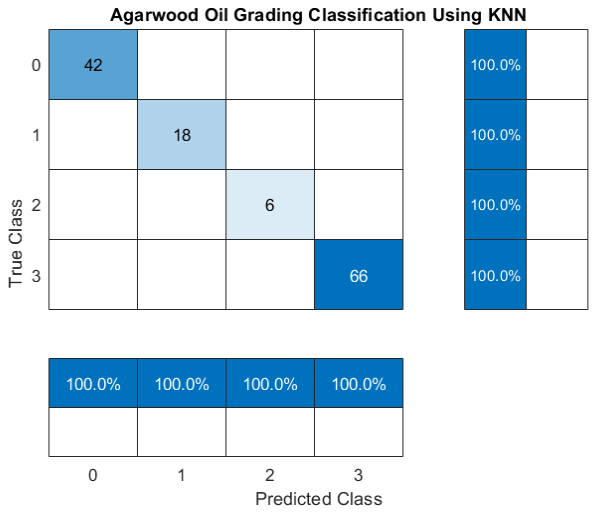

Fig. 6. 4x4 confusion matrix table classifier of agarwood oil qualities

Based on the results of the confusion matrix table, accuracy, specificity, sensitivity, and precision were calculated. All four measurements are at 100%. This excellent result was obtained because there was no grading error, and this made the values zero for both "false negative" and "false positive".

5. Conclusions

An algorithm with multiclass classification ability was needed to classify agarwood oil into more than two grades. If the algorithm that was created to classify binary classification was used in this study, it needs to be improved first. That is why the k-Nearest Neighbors (KNN) algorithm was chosen for this study, because it has already been designed to solve the multiclass classification problem with no need for improvement. Plus, this KNN model was using Euclidean distance to search for the nearest neighbors before classifying it. The Euclidean distance metric is one of the good choices in handling such huge data. Based on the evaluation result, it is proven that the KNN model is one of the suitable algorithms to be implemented. Furthermore, with a classification accuracy of 100 percent for such large nonlinear data samples as the agarwood oil data sample used, there is no doubt that the model should be proposed as the main algorithm in future nonlinear data classification studies.

The findings of this study will aid other researchers in determining the unique properties of agarwood oil. The k-Nearest Neighbors (KNN) technique was used to create the intelligent model, which is important for the forthcoming investigation, especially in the agarwood oil categorization system. It is suggested that the agarwood oil model be designed using different technique to categorize more than two grades of agarwood oil quality in future studies. Their outcomes may then be compared and evaluated.

Acknowledgement
Under Grant No. 600-RMC 5/3/GPM, the Research Management Centre (RMC), Universiti UiTM Shah Alam, Selangor, and UiTM Johor Branch, Kampus Pasir Gudang, financed this research. The authors would like to express their gratitude to all of the staff at the Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, Natural Product Division, Forest Research

Institute Malaysia (FRIM), and Bioaromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang (UMP) for their assistance throughout the study.

References

[1] Haron, Mohamad Hushnie. "Agarwood oil quality grading model using selforganizing map (SOM)." PhD diss., Universiti Teknologi MARA, 2020.

[2] Wang, Shuai, Zhangxin Yu, Canhong Wang, Chongming Wu, Peng Guo, and Jianhe Wei. "Chemical constituents and pharmacological activity of agarwood and Aquilaria plants." *Molecules* 23, no. 2 (2018): 342. https://doi.org/10.3390/molecules23020342

[3] Ismail, N. "ANN modeling for agarwood oil significant compounds for oil quality discrimination." *PhD Dissertation, Faculty of Electrical Engineering, Universiti Teknologi Mara (UiTM)* (2015).

[4] Ismail, Nurlaila, NA Mohd Ali, Mailina Jamil, Mohd Hezri Fazalul Rahiman, Saiful Nizam Tajuddin, and Mohd Nasir Taib. "A review study of agarwood oil and its quality analysis." *J. Teknol* 68, no. 1 (2014). https://doi.org/10.11113/jt.v68.2419

[5] López-Sampson, Arlene, and Tony Page. "History of use and trade of agarwood." *Economic botany* 72, no. 1 (2018): 107-129. https://doi.org/10.1007/s12231-018-9408-4

[6] Kamarulzaini, Khairul Anis Athirah, Nurlaila Ismail, Mohd Hezri Fazalul Rahiman, Mohd Nasir Taib, Nor Azah Mohd Ali, and Saiful Nizam Tajuddin. "Evaluation of RBF and MLP in SVM kernel tuned parameters for agarwood oil quality classification." In *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 250-254. IEEE, 2018. https://doi.org/10.1109/CSPA.2018.8368721

[7] Haron, Mohamad Hushnie, Mohd Nasir Taib, Nurlaila Ismail, Nor Azah Mohd Ali, and Saiful Nizam Tajuddin. "Statistical analysis of agarwood oil compounds based on GC-MS data." In *2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, pp. 27-30. IEEE, 2018. https://doi.org/10.1109/ICSGRC.2018.8657571

[8] Zubir, N. S. A., M. A. Abas, Nurlaila Ismail, Nor Azah M. Ali, M. H. F. Rahiman, N. K. Mun, M. N. Taib, and N. T. Saiful. "Pattern classifier of chemical compounds in different qualities of agarwood oil parameter using scale conjugate gradient algorithm in MLP." In *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, pp. 18-22. IEEE, 2017. https://doi.org/10.1109/CSPA.2017.8064917

[9] Abas, M. A. H., N. S. A. Zubir, N. Ismail, N. A. M. Ali, M. H. F. Rahiman, A. I. M. Yassin, S. N. Tajuddin, and M. N. Taib. "Development of web application for agarwood oil quality discriminator in virtualization platform." In *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, pp. 13-17. IEEE, 2017. https://doi.org/10.1109/CSPA.2017.8064916

[10] Ismail, Nur Shazareen, Nurlaila Ismail, Mohd Hezri Fazalul Rahiman, Mohd Nasir Taib, Nor Azah Mohd Ali, and Saiful Nizam Tajuddin. "Polynomial tuned kernel parameter in SVM of agarwood oil for quality classification." In *2018 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 77-82. IEEE, 2018. https://doi.org/10.1109/I2CACIS.2018.8603686

[11] Ngadilan, Mohamad Amirul Aiman, Nurlaila Ismail, Mohd Hezri Fazalul Rahiman, Mohd Nasir Taib, Nor Azah Mohd Ali, and Saiful Nizam 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)*, pp. 64-68. IEEE, 2018.

[12] Taunk, Kashvi, Sanjukta De, Srishti Verma, and Aleena Swetapadma. "A brief review of nearest neighbor algorithm for learning and classification." In *2019 international conference on intelligent computing and control systems (ICCS)*, pp. 1255-1260. IEEE, 2019. https://doi.org/10.1109/ICCS45141.2019.9065747

[13] Deng, Shaobo, Lei Wang, Sujie Guan, Min Li, and Lei Wang. "Non-parametric Nearest Neighbor Classification Based on Global Variance Difference." *International Journal of Computational Intelligence Systems* 16, no. 1 (2023): 26. https://doi.org/10.1007/s44196-023-00200-1

[14] Kansal, Liza, Anoushka Pandey, Sanidhya Madhav Shukla, and Parneeta Dhaliwal. "Review of Machine Learning Techniques for Crop Recommendation." In *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*, pp. 443-449. 2023. https://doi.org/10.1145/3607947.3608045

[15] Mathew, George, Amritanshu Agrawal, and Tim Menzies. "Finding trends in software research." *IEEE Transactions on Software Engineering* (2018).

[16] Kim, Jaeyoung, Sion Jang, Eunjeong Park, and Sungchul Choi. "Text classification using capsules." *Neurocomputing* 376 (2020): 214-221. https://doi.org/10.1016/j.neucom.2019.10.033

[17] Mohamad, Mujahid, Muhammad Sharfi Najib, Saiful Nizam Tajuddin, Suhaimi Mohd Daud, Nur Farina Hamidon Majid, Suziyanti Zaib, and Muhamad Faruqi Zahari. "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:*

InECCE2021, Kuantan, Pahang, Malaysia, 23rd August, pp. 575-586. Singapore: Springer Singapore, 2022. https://doi.org/10.1007/978-981-16-8690-0_52

[18] Hmeidi, Ismail, Bilal Hawashin, and Eyas El-Qawasmeh. "Performance of KNN and SVM classifiers on full word Arabic articles." *Advanced Engineering Informatics* 22, no. 1 (2008): 106-111. https://doi.org/10.1016/j.aei.2007.12.001

[19] Li, Leping, Clarice R. Weinberg, Thomas A. Darden, and Lee G. Pedersen. "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method." *Bioinformatics* 17, no. 12 (2001): 1131-1142. https://doi.org/10.1093/bioinformatics/17.12.1131

[20] Su, Ming-Yang. "Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification." *Journal of Network and Computer Applications* 34, no. 2 (2011): 722-730. https://doi.org/10.1016/j.jnca.2010.10.009

[21] Tan, Songbo. "An effective refinement strategy for KNN text classifier." *Expert Systems with Applications* 30, no. 2 (2006): 290-298. https://doi.org/10.1016/j.eswa.2005.07.019

[22] Ventouras, Errikos M., Pantelis Asvestas, Irene Karanasiou, and George K. Matsopoulos. "Classification of error-related negativity (ERN) and positivity (Pe) potentials using kNN and support vector machines." *Computers in biology and medicine* 41, no. 2 (2011): 98-109. https://doi.org/10.1016/j.compbiomed.2010.12.004

[23] Koutroumbas, K., and N. Kalouptsidis. "Nearest neighbor pattern classification neural networks." In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, vol. 5, pp. 2911-2915. IEEE, 1994.

[24] Denoeux, Thierry. "A k-nearest neighbor classification rule based on Dempster-Shafer theory." *IEEE transactions on systems, man, and cybernetics* 25, no. 5 (1995): 804-813. https://doi.org/10.1109/21.376493

[25] Houben, Isabelle, Louis Wehenkel, and Mania Pavella. "Coupling of K-NN with decision trees for power system transient stability assessment." In *Proceedings of International Conference on Control Applications*, pp. 825-832. IEEE, 1995.

[26] Saini, Indu, Dilbag Singh, and Arun Khosla. "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases." *Journal of advanced research* 4, no. 4 (2013): 331-344. https://doi.org/10.1016/j.jare.2012.05.007

[27] Rahmatian, M., B. Vahidi, A. J. Ghanizadeh, G. B. Gharehpetian, and H. A. Alehosseini. "Insulation failure detection in transformer winding using cross-correlation technique with ANN and k-NN regression method during impulse test." *International Journal of Electrical Power & Energy Systems* 53 (2013): 209-218. https://doi.org/10.1016/j.ijepes.2013.04.020

[28] Yazdani, Ashkan, Touradj Ebrahimi, and Ulrich Hoffmann. "Classification of EEG signals using Dempster Shafer theory and a k-nearest neighbor classifier." In *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pp. 327-330. IEEE, 2009. https://doi.org/10.1109/NER.2009.5109299

[29] Beckmann, Marcelo, Nelson FF Ebecken, and Beatriz SL Pires de Lima. "A KNN undersampling approach for data balancing." *Journal of Intelligent Learning Systems and Applications* 7, no. 04 (2015): 104-116. https://doi.org/10.4236/jilsa.2015.74010

[30] Albano, Christer, William Dunn III, Ulf Edlund, Erik Johansson, Bo Nordén, Michael Sjöström, and Svante Wold. "Four levels of pattern recognition." *Analytica Chimica Acta* 103, no. 4 (1978): 429-443. https://doi.org/10.1016/S0003-2670(01)83107-X

[31] Dudani, Sahibsingh A. "The distance-weighted k-nearest-neighbor rule." *IEEE Transactions on Systems, Man, and Cybernetics* 4 (1976): 325-327. https://doi.org/10.1109/TSMC.1976.5408784

[32] Abas, M. A., N. S. A. Zubir, N. Ismail, I. M. Yassin, N. A. M. Ali, M. H. F. Rahiman, N. T. Saiful, and M. N. Taib. "Agarwood oil quality classifier using machine learning." *Journal of Fundamental and Applied Sciences* 9, no. 4S (2017): 62-76. https://doi.org/10.4314/jfas.v9i4S.4

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

[34] Wang, Hui Hui, Yin Chai Wang, Bui Lin Wee, and Shiang Wei Sim. "Novel Feature Extraction for Oil Palm Bunches Classification." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 34, no. 1 (2024): 350-360. https://doi.org/10.37934/araset.34.1.350360

[35] Samad, Muhamad Ezeywan Mohd, Nurlaila Ismail, Mohd Hezri Fazalul Rahiman, Mohd Nasir Taib, Nor Azah Mohd Ali, and Saiful Nizam Tajuddin. "Analysis of distance metric variations in KNN for agarwood oil compounds differentiation." In *2017 IEEE conference on systems, process and control (ICSPC)*, pp. 151-156. IEEE, 2017. https://doi.org/10.1109/SPC.2017.8313038

[36] Parker, J. R. "Rank and response combination from confusion matrix data." *Information fusion* 2, no. 2 (2001): 113-120. https://doi.org/10.1016/S1566-2535(01)00030-6

[37] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks." *Information processing & management* 45, no. 4 (2009): 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

[38] Salvador–Meneses, Jaime, Zoila Ruiz–Chavez, and Jose Garcia–Rodriguez. "Compressed k NN: K-nearest neighbors with data compression." *Entropy* 21, no. 3 (2019): 234. https://doi.org/10.3390/e21030234

[39] Pulungan, Annisa Fadhillah, Muhammad Zarlis, and Saib Suwilo. "Analysis of braycurtis, canberra and euclidean distance in knn algorithm." *Sinkron: jurnal dan penelitian teknik informatika* 4, no. 1 (2019): 74-77. https://doi.org/10.33395/sinkron.v4i1.10207

[40] De Souza, Cristiano Antonio, Carlos Becker Westphall, Renato Bobsin Machado, João Bosco Mangueira Sobral, and Gustavo dos Santos Vieira. "Hybrid approach to intrusion detection in fog-based IoT environments." *Computer Networks* 180 (2020): 107417. https://doi.org/10.1016/j.comnet.2020.107417

[41] Ismail, Nurlaila, Nor Azah Mohd Ali, Mailina Jamil, Mohd Hezri Fazalul Rahiman, Mohd Nasir Taib, and Saiful Nizam Tajuddin. "Major volatile chemical compounds of agarwood oils from Malaysia based on z-score technique." *Chemistry of Natural Compounds* 51 (2015): 776-779. https://doi.org/10.1007/s10600-015-1409-2

[42] Ismail, Nurlaila, Mohd Ali Nor Azah, Mailina Jamil, Mohd Hezri Fazalul Rahiman, Saiful Nizam Tajuddin, and Mohd Nasir Taib. "Analysis of high quality agarwood oil chemical compounds by means of SPME/GC-MS and Z-score technique." *Malaysian Journal of Analytical Sciences* 17, no. 3 (2013): 403-413.

[43] Ismail, Nurlaila, Nor Azah Mohd Ali, Mailina Jamil, Mohd Hezri Fazalul Rahiman, Saiful Nizam Tajuddin, and Mohd Nasir Taib. "Differentiating agarwood oil quality using artificial neural network." *Malaysian Journal of Analytical Sciences* 17, no. 3 (2013): 490-498.

[44] Mailina, J., Nor Azah, and Mohamad Nasir. "Profiles of selected supreme agarwood oils from Malaysia." In *Proceedings of the Seminar on Medicinal and Aromatic Plants (MAPS 2008)*. 2009.