The significance of artificial intelligent technique in classifying various grades of agarwood oil

Aqib Fawwaz Mohd Amidon¹, Siti Mariatul Hazwah Mohd Huzir², Zakiah Mohd Yusoff², Nurlaila Ismail¹, Mohd Nasir Taib¹,³

¹School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia
²School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Pasir Gudang, Malaysia
³Malaysia Institute of Transport (MITRANS), Universiti Teknologi MARA, Shah Alam, Malaysia

ABSTRACT

Agarwood oil quality is often separated into two or three categories. This makes classifying agarwood oil quality using current methods difficult. Current approaches rely solely on human perception to determine the quality of agarwood, whether in raw material or oil. This technique has other undesirable implications. It can affect the human sensory system, particularly the eyes and nose. Categorization takes time, which is a considerable expense to succeed in this method. As a result, a new classification system should be devised. The chemical components in agarwood oil are used to classify it in this study. In this study, samples with preprocessing data from two to five quality levels were used. The purpose is to categorize this data based on its qualities and analyze whether this new quality group is acceptable. The K-nearest neighbours (KNN) approach was used to classify all samples and their properties for this dataset. All samples may be correctly classified by grade without any errors. This shows the chemical compound-based classification of agarwood oil can be retained. With these findings, future agarwood oil research may focus on building a new classification.

Keywords: Agarwood oil, Chemical compound, Classification, KNN technique, Quality levels

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

It is generally known that agarwood oil is one of the most in-demand oils on the market today. Despite the fact that each agarwood oil is fairly expensive, especially those of higher grade, this product continues to attract orders from perfume connoisseurs almost all the time. Agarwood is frequently used in the creation of medicine, fragrances, and as a burner in religious rituals, ethnocultural festivities, and others.

However, the process of grading agarwood oil is only done by those who are experienced in the search for agarwood trees. They will evaluate the physical terms of agarwood oil, which is, by looking at the concentration of its colour and also by evaluating the level of odour emitted by agarwood oil. These are the two main methods they use to determine whether the agarwood oil is of high quality or low quality. Unfortunately, this implementation will have some adverse effects, including weakening the ability of the eyes and nose involved, requiring a long period of time to run the process, and also requiring a high cost to carry out such a long classification process.

Thus, transformation in the agarwood oil classification process needs to be implemented. With the advancement of technology today, scientific methods should be implemented in order to conduct the
classification of agarwood more accurately and in detail. The chemical compounds that exist in agarwood oil can be used to determine its quality. Even from just two levels of grading (high and low quality), it can also be expanded to four or five quality levels. It has been proven before. Previous researchers, as in [1], have successfully established up to five classifications for the quality of agarwood oil.

In order to continue the efforts in the study on the transformation of the agarwood oil quality classification process, this study has been conducted based on these previous findings. The classification process for the classification of the five qualities of agarwood oil using artificial intelligence has been implemented. The goal of this implementation is to find out if the classification based on these agarwood oil components is acceptable or not.

In previous studies, several techniques have been used in validating the classification of agarwood oil, especially for two qualities of agarwood oil. Among the techniques are support vector machines (SVM), artificial neural networks (ANN), radial basis functions (RBF), multilayer perceptron, and more as mentioned in [2], [3]. However, for this study, the K-nearest neighbours (KNN) technique was chosen and has been assigned as artificial technique to validate the classification of five qualities of agarwood oil. The MATLAB software platform version r2020a was used for the entire study.

2. THEORETICAL WORK

This section is divided into three subsections. The first focuses on the history, trading and benefits of agarwood oil. The second point of emphasis is on the KNN technique. The third section describes the standard performance measures that are commonly used to evaluate the performance of classification models.

2.1. Agarwood oil

Agarwood is not the only name used for it. Among the many other names for agarwood are: gaharuwood, aloeswood, eaglewood, and many more [4]-[7]. Agarwood has been used in the manufacture of medicine, as a major ingredient in the production of perfumes, and as a source of burning for religious ceremonies or festivals since time immemorial [7], [8]. This shows that this agarwood is very close to human life. It is hilly when, to this day, the demand for it remains high in the market [9]. Demand for agarwood usually comes from countries such as China, Japan, India, the United Arab Emirates, and Saudi Arabia [10].

In short, most agarwood can be found around the archipelago, such as Malaysia, Indonesia, Thailand, Cambodia, Vietnam, Laos, and others [11]. Usually, the agarwood that exists in the archipelago is from the species Aquilaria, which comes from the family of Thymelaeaceae [1], [4]. The process of discovering agarwood is not simple. It can only be found by those who are well-versed in the throughs and breakaways of the forest.

In terms of quality classification, the process uses human senses such as the nose and eyes [12]. Those who are experts will assess the level of color concentration of agarwood oil as well as evaluate the level of odor emitted by agarwood oil. These two evaluations will determine whether the agarwood oil evaluated is high quality or not. This method of classification is still used to this day. However, this method cannot be continued as it will continue to adversely affect those who make assessments on an ongoing basis. Certainly, the use of the eyes and nose at their maximum level at all times will affect their function and body health. In addition, making an assessment one by one takes a long time to complete the classification process and certainly requires a high cost to complete each batch of classification [12].

Thus, an in-depth study has been conducted in an effort to establish a new method to make this agarwood oil classification. Previous studies have found an idea to solve this problem by using chemical compounds to determine the quality level of agarwood oil [13]. As a result, the classification of agarwood oil based on the proportion of chemical components in agarwood samples will be varied using the KNN technique in this study.

2.2. K-nearest neighbors (KNN) technique

The KNN technique is a supervised machine learning approach that can be applied to classification and regression issues [14]-[17]. However, it is mostly employed in the industry to solve categorization issues. The KNN method is based on the Supervised Learning approach and is one of the most basic machine learning algorithms [18]-[21]. The KNN method assumes that the new case/data and existing cases are comparable, and it places the new case in the category that is most similar to the existing categories. The KNN algorithm saves all existing data and classifies fresh data points according to their similarity. This implies that new data present can be quickly sorted into a well-suited category using the KNN algorithm as it arises [22]-[23]. The KNN algorithm is a non-parametric algorithm, which means it makes no assumptions about the data [24], [25]. It's also known as a "lazy learner" algorithm since it doesn't learn from the training set right away; instead, it saves the dataset and performs an action on it when it comes time to classify it [26].
The KNN method simply saves the information during the training phase, and when it receives new data, it classifies it into a category that is similar to the new data.

Assume that there are two categories, Category A and Category B, and that we have a new data point (X1), which will fall into one of these categories. A KNN method was required to address this situation. The new data point may readily identify the category or class of a dataset with the use of KNN as shown in Figure 1.

Figure 1. The KNN technique concept [27]

2.3. Standard performance measure

The purpose of performance measurements in a categorisation is to investigate the categorization system's behaviour and abilities. The categorization system is based on a number of factors. The examples of performance measure used such as confusion matrix, accuracy, specificity, sensitivity, precision, error rate, mean square error (MSE), error rate, receiver operating characteristic (ROC), curve analysis, and regression [28]-[30].

A confusion matrix is a two-dimensional matrix with two parameters: the matrix row and matrix column. The matrix column represents the desired class or classification for that particular datum, whereas the matrix row represents the actual class or datum under test. The matrix is investigated in order to assess the performance of many classifiers in vision systems. This matrix has the benefit that previous knowledge about the classifier is a type of confusion matrix, which is useful for generating response vectors and rankings [28].

The level of uncertainty in a measurement with regard to an absolute standard is known as accuracy [30], [31]. The influence of mistakes due to gain and offset settings is frequently included in accuracy criteria. Offset errors can be expressed in volts or ohms, and they are independent of the concentration of the original signal being analysed. The formula of accuracy as shown in (1).

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{1}
\]

Sensitivity is a numerical value. The tiniest absolute amount of change that a measurement can detect [32]. As shown in (2) is the formula of sensitivity.

\[
\text{Sensitivity} = \frac{TP}{(TP+FN)} \tag{2}
\]

The proportion of genuine negatives that are accurately recognised as such, as well as the percentage of healthy persons who are correctly diagnosed as not having the ailment, is measured by specificity, also known as the true negative rate [33]. Calculation of specificity as (3).

\[
\text{Specificity} = \frac{TN}{TN+FP} \tag{3}
\]

The measurement's precision refers to its repeatability. Measure a steady state signal several times, for example. If the values are close together in this scenario, the precision or repeatability is high [28]. The values don't have to be genuine values; they can just be grouped together. The difference between the average of the measurements and the actual outcome called precision. The formula of precision as shown in (4).
\[ \text{Precision} = \frac{TP}{(TP+FP)} \] (4)

3. METHOD

The process of verifying the quality of each agarwood sample is done in an orderly and neat manner as shown in Figure 2. Basically, this process was divided into three stages. First stage was the data pre-processing. Second stage was the development of classification model using KNN algorithm. Third stage was covered by the evaluation of model performance.

![Process flow in the implementation of agarwood oil quality verification using the KNN classification model](image)

600 samples of data consists of five different qualities of agarwood oil: Supreme, High, Medium High, Medium Low, Low quality.

1. Data normalization.
2. Data randomization.
3. Data division (80%:20%) using ‘Holdout’ for training dataset and testing dataset.

- Varying the training dataset using Euclidean (Nearest neighbors search method)
- Performance criteria:
  1. Confusion matrix
  2. Accuracy
  3. Sensitivity
  4. Specificity
  5. Precision

- 1. Each quality is clearly classified.
- 2. Characteristics that are similar for each data set from the same quality group
- 3. Distinctive characteristics for each data set from various quality groups

Initially, this data only grouped by three different quality groups. It includes high, medium, and low quality. After that, all the sample data was weighed through the process of statistical analysis, and the identification of new characteristics of agarwood oil according to quality was also made. Thus, a new data set of agarwood oil samples was produced with five quality differences after visual validation was done on them. These five new quality groups include supreme, high, medium high, medium low-, and low-quality groups.
In total, the sample data included 600 samples, with each recording the percentage abundance of eleven chemical compounds.

All sample data is pre-processed before being used to develop a classification model to classify the quality of agarwood oil. All the sample data was normalized before the arrangement was randomized in this method. Finally, the sample data was separated into two groups: training and testing datasets, with an average ratio of 80 to 20%. Holdout's validation is used in this isolation.

After the data pre-processing is carried out, the model for the classification is designed. Training datasets (480 sample of data) were used to construct this classification model. As previously discussed, the KNN algorithm was used as the classification technique for this model. This KNN algorithm requires the nearest neighbor search method to allow it to make a classification for each sample. Therefore, the "Euclidean" nearest neighbor search method has been adapted.

The testing dataset (120 sample of data) is used to evaluate the capabilities of the KNN classification model once it has been developed. It is at this step that the re-grading performed on the sample data at an earlier level can be applied or not. Evaluation of performance measure will cover the confusion matrix, accuracy, specificity, sensitivity, and also precision.

4. RESULTS AND DISCUSSION

Figure 3 shows that the position of the supreme cluster is quite accurate and precise, making the classification between the supreme cluster and other quality clusters easy. This is because it can be clearly seen that there is a gap between the supreme group and the high-quality group. Certainly, this behaviour facilitates the classification of the model to identify the differences in the features that exist for each sample.

Figure 3. 3-Dimension graph of testing dataset for supreme and high-quality agarwood oil samples

There is also a gap between the low, medium-low, and medium-high quality clusters, as seen in Figure 4. This raises optimism about the classification's ability to reliably determine the quality of each sample data set. As shown in Figure. 5, it has been proved that the initial result shown in the 3-demension graph in Figures 3 and 4 were accurate. Because, based on the confusion matrix table in Figure 5, no classification error occurs when the testing process is conducted against the KNN classification model.

As a result, the percentage value for accuracy, sensitivity, specificity, and precision were calculated to be 100% as shown in Table 1. The ability to accurately categorize agarwood oil sample data based on its quality has led to the great outcomes for this performance measure. Furthermore, as previously said, the graph's results reveal that each quality groups were clearly classified with a gap between one the others. Also, with the specific distinctive characteristics (the abundance of significant chemical compounds) for each dataset from various quality groups was affects the result. As conclusion of this discussion, the KNN technique was capable to identify the qualities of each of the agarwood oil samples with excellent performance.
5. CONCLUSION

In general, relating to the proportional abundance of major chemical components can provide a more complete and accurate categorization of agarwood oil qualities. As a consequence, the prior study was able to create a data set of agarwood oil samples with five different quality clusters. Based on that, this study was undertaken to classify the data gathered using an artificial intelligence technique (K-nearest neighbors technique) in order to examine if the characteristics of each group could scientifically contribute to the categorization process. Finally, the scientific categorization of agarwood oil data sets can be successfully accomplished. The KNN technique's capacity to accurately identify each set of data according to its quality is particularly impressive. These results were affected by several factors, as mentioned in the results and discussion sections before. Next, regarding the performance measure, when no categorization mistakes occur, this will undoubtedly add to the measurement assessment outcomes. As a result, accuracy, sensitivity,
specificity, and precision are indeed very high. As a result, the development of this KNN classification in generating a classification of up to five different qualities of agarwood oil data sets in huge quantities gives fresh hope to the interest in learning about developing a standard classification method for agarwood oil.

ACKNOWLEDGEMENTS
The Institute of Research Management and Innovation (IRMI) at Universiti Teknologi MARA Shah Alam, Selangor, and Universiti Teknologi MARA, Cawangan Johor, Kampus Pasir Gudang, financed this study under Grant No. 600-RMC 5/3/GPM (030/2022). The authors would like to express their gratitude to all of the people involved, particularly those from School of Electrical Engineering, College of Engineering at UiTM Shah Alam, Bioaromatic Research Centre of Excellence (BARCE) at Universiti Malaysia Pahang (UMP) and Forest Research Institute Malaysia (FRIM) for their contributions to this study at various stages.

REFERENCES
[1] M. H. Haron, “Agarwood oil quality grading model using self-organizing map (SOM),” Universiti Teknologi MARA (UiTM) Shah Alam, 2020.
[2] H. Jantan, I. M. Yassin, A. Zabidi, N. Ismail, and M. S. A. M. Ali, “Differentiation of agarwood oil quality using support vector machine (SVM),” Journal of Engineering and Applied Sciences, vol. 12, no. 15, pp. 3810–3812, 2017, doi: 10.3923/jeas.2017.3810.3812.
[3] M. A. Aiman Ngadilan, N. Ismail, M. H. F. Rahman, 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),” 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC) 2018 - Proceeding, pp. 64–68, 2019, doi: 10.1109/ICSGRC.2018.8657524.
[4] N. Ismail, M. H. F. Rahman, M. N. Taib, M. Ibrahim, S. Zareen, and S. N. Tajuddin, “Analysis on agarwood vapour using headspace volatile DVB-CAR-PDMS SPME with different sampling time,” in 2016 7th IEEE Control and System Graduate Research Colloquium (ICSGRC), Aug. 2016, pp. 170–174, doi: 10.1109/ICSGRC.2016.7813322.
[5] K. A. A. Kamarulzaini, N. Ismail, M. H. F. Rahman, M. N. Taib, N. A. M. Ali, and S. N. 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), Mar. 2018, pp. 250–254, doi: 10.1109/CSPA.2018.8368721.
[6] M. H. Haron, M. N. Taib, N. Ismail, N. A. Mohd Ali, and S. N. Tajuddin, “Statistical analysis of agarwood oil compounds based on GC-MS Data,” in 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC), Aug. 2018, pp. 27–30, doi: 10.1109/ICSGRC.2018.8657571.
[7] Z. Xiao, S. Jia, H. Bao, Y. Niu, Q. Ke, and X. Kou, “Protection of agarwood essential oil aroma by nanocellulose-graft-polyactic acid,” International Journal of Biological Macromolecules, vol. 183, pp. 743–752, Jul. 2021, doi: 10.1016/j.ijbiomac.2021.04.097.
[8] W. Li, H. Q. Chen, H. Wang, W. L. Mei, and H. F. Dai, “Natural products in agarwood and Aquilaria plants: Chemistry, biological activities and biosynthesis,” Natural Product Reports, vol. 38, no. 3, pp. 528–565, 2021, doi: 10.1039/d0np00042f.
[9] S. Ma, Y. Fu, Y. Li, P. Wei, and Z. Liu, “The formation and quality evaluation of agarwood induced by the fungi in Aquilaria sinensis,” Industrial Crops and Products, vol. 173, p. 114129, Dec. 2021, doi: 10.1016/j.indcrop.2021.114129.
[10] A. H. M. Fadzli, K. H. K. Hamid, M. N. M. Rodhi, and L. M. M. Kamarudin, “Extraction of essential oil from biologically inoculated agarwood,” in 2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC), Apr. 2013, pp. 889–892, doi: 10.1109/BEIAC.2013.6550264.
[11] N. Ismail, M. H. F. Rahman, M. N. Taib, N. A. M. Ali, M. Jamil, and S. N. Tajuddin, “Analysis of chemical compounds of agarwood oil based on headspace-solid phase microextraction combined with gas chromatography mass-spectrometry,” in 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, Mar. 2013, pp. 215–218, doi: 10.1109/CSPA.2013.6530044.
[12] M. S. Najib, M. N. Taib, N. A. M. Ali, M. N. M. Arif, and A. M. Jalil, “Classification of Agarwood grades using ANN,” in International Conference on Electrical, Control and Computer Engineering 2011 (InECCE), Jun. 2011, pp. 367–372, doi: 10.1109/INECCE.2011.5953908.
[13] 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 International Conference on Smart Sensors and Applications, ICSSA 2015, pp. 7–10, 2015, doi: 10.1109/ICSSA.2015.7322500.
[14] Y. Sari, M. Maulida, E. Gunawan, and J. Wahyuadi, “Artificial intelligence approach for BAZNAS website using K-Nearest neighbor (KNN),” in 2021 Sixth International Conference on Informatics and Computing (ICIC), Nov. 2021, pp. 1–4, doi: 10.1109/ICIC54025.2021.9632954.
[15] A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve, and N. Lal, “Crime prediction using k-nearest neighboring algorithm,” in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETTE), Feb. 2020, pp. 1–4, doi: 10.1109/ic-ETTE47903.2020.155.
[16] C. Ren, L. Sun, and Q. Wu, “Study on density peaks clustering based on hierarchical K-nearest neighbors,” in 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nov. 2019, pp. 664–668, doi: 10.1109/ISKE74853.2019.9170434.
[17] M. S. Kadmim, H. Ayad, and M. J. Mohammed, “Palmprint recognition system based on proposed features extraction and (c5.0) decision tree, k-nearest neighbor (KNN) classification approaches,” Journal of Engineering Science and Technology, vol. 16, no. 1, pp. 816–831, 2021.
[18] S. A. Dadami, “The distance-weighted k-nearest-neighbor rule,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976, doi: 10.1109/TSMC.1976.5408784.
[19] I. Saini, D. Singh, and A. Khosla, “QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases,” Journal of Advanced Research, vol. 4, no. 4, pp. 331–344, 2013, doi: 10.1016/j.jare.2012.05.007.
[20] T. Jayalakshmi and A. S. Kumanar, “Statistical normalization and back propagation,” International Journal of Computer Theory and Engineering, vol. 3, p. 1793, 2011, doi: 10.7763/IJCTE.2011.V3.288.
BIOGRAPHIES OF AUTHORS

Aqib Fawwaz Mohd Amidon was born in Malaysia, in September 1996. He received his B. Eng. (Hons) of Electronic Engineering from Universiti Teknologi MARA (UiTM). He is currently a Software Engineer at Greatech Technology Berhad 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: aqibfawwaz.academic@gmail.com.

Siti Mariatul Hazwa Mohd Huzir was born in Malaysia who is currently persue her studies as a postgraduate student majoring in Electrical Engineering at UiTM Cawangan Johor, Kampus Pasir Gudang. She received her B. Eng (Hons) of Electronics Engineering from Universiti Teknologi MARA (UiTM) Shah Alam Malaysia. She can be contacted at email: mariatulhazwa97@gmail.com.
Zakiah Mohd Yusoff is a senior lecturer who is currently working at UiTM Pasir Gudang. She received the B. ENG in Electrical Engineering and PhD in Electrical Engineering from UiTM Shah Alam, in 2009 and 2014, respectively. In 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.

Nurlaila Ismail received her PhD 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.

Mohd Nasir Taib received his PhD 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.