# Chemometric classification and authentication of four Aquilaria species from essential oil profiles using GC-MS/GC-FID and ANN

Nur Athirah Syafiqah Noramli [a,1,*], Noor Aida Syakira Ahmad Sabri [a,2], Muhammad Ikhsan Roslan [a,3], Nurlaila Ismail [a,4], Zakiah Mohd Yusoff [a,5], Mohd Nasir Taib [a,6]

[a] Advanced Signal Processing Research Interest Group, Faculty of Electrical Engineering, Universiti Teknologi MARA, Selangor, Malaysia
[1] athirah.noramli1@gmail.com; [2] aidasyakiraaa01@gmail.com; [3] muhammadikhsanroslan@gmail.com; [4] nurlaila0583@uitm.edu.my; [5] zakiah9018@uitm.edu.my; [6] dr.nasir@uitm.edu.my
* corresponding author

ARTICLE INFO

**Article history**
Received June 28, 2025
Revised October 12, 2025
Accepted October 12, 2025
Available online November 30, 2025

**Keywords**
Agarwood
Aquilaria species
Pearson correlation
Artificial Neural Network (ANN)
Classification model

ABSTRACT

Agarwood, derived from the Aquilaria species, is among the most valuable aromatic resources, yet frequent species misidentification hampers conservation efforts and compliance with trade regulations. This study applied a chemometric ANN framework to classify four Aquilaria species (*A. malaccensis*, *A. beccariana*, *A. subintegra*, and *A. crassna*) based on essential oil composition. A total of 720 samples (180 per species, each analyzed in triplicate) were extracted by hydrodistillation and profiled using GC–MS coupled to GC–FID. Six compounds were consistently detected, and three (δ-guaiene, 10-epi-$\gamma$-eudesmol, $\gamma$-eudesmol) were retained for classification based on ≥95% detection frequency and >0.2% relative abundance. Pearson correlation guided feature selection, and ANN models were trained using both a 70:15:15 train–validation–test split and stratified 5-fold cross-validation with 1000 bootstrap resamples. The optimized network achieved near-perfect performance, with a mean accuracy of ~99.8% (95% CI: 99.6–100.0), and precision, recall, and F1 scores all exceeding 99.5%. In comparison, bootstrapped confidence intervals were tightly bounded at 100%, confirming robustness against data leakage. These findings demonstrate that correlation-guided feature selection combined with ANN modeling enables reproducible and highly accurate species authentication, offering a practical framework for integration into agarwood quality control, conservation monitoring, and international trade compliance.



## 1. Introduction

Agarwood, a highly valued fragrant resin, is produced by trees of the genus Aquilaria, which belongs to the family *Thymelaeaceae* [1]–[3]. These trees are indigenous to the Indomalesia region, encompassing parts of Southeast Asia such as Malaysia, Indonesia, and Thailand [4], [5]. Agarwood has been esteemed for centuries across various cultures for its use in perfumes, incense, and traditional medicines [4]. The formation of agarwood is a defence response to natural infections or injuries, leading to the accumulation of a dark, aromatic resin within the heartwood [1], [6]. Because of its rarity and market appeal, agarwood is recognized as one of the world's most expensive natural resources [7], [8].

Despite its economic and cultural significance, the Aquilaria genus faces challenges in conservation and sustainable utilization. Overharvesting due to high demand has led to a significant decline in wild

populations [4], [9]–[12]. As a result, multiple Aquilaria species have been placed under the protection of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and are also classified on the Red List maintained by the International Union for Conservation of Nature (IUCN) [13]. Efforts have been made to cultivate Aquilaria trees to meet market demands and reduce pressure on wild populations. However, the success of these initiatives is hindered by challenges in species identification and classification [4].

A critical issue in the conservation and commercial utilization of Aquilaria species is the frequent misidentification of species [8], [14]. This misidentification arises from a lack of comprehensive taxonomic information and the morphological similarities among species. Accurate species identification is essential for implementing effective conservation strategies, ensuring compliance with international trade regulations, and maintaining the quality and authenticity of agarwood products [4], [13]. The current taxonomic ambiguities impede these efforts, highlighting the need for reliable classification methods.

In response to these challenges, this study aims to develop a chemometric classification model for Aquilaria species based on the chemical composition of their essential oils. By employing statistical techniques such as Pearson correlation analysis for feature selection and ANN for classification, the research seeks to establish a robust, efficient, and accurate method for species identification. The approach involves analyzing essential oil samples from various Aquilaria species using gas chromatography-mass spectrometry (GC-MS) and gas chromatography flame ionization detector (GC-FID), followed by data processing and model development. This approach is intended to improve species authentication, contribute to conservation initiatives, and safeguard the authenticity of agarwood products in the international market.

## 2. Method

This study followed a structured approach to classify selected Aquilaria species based on the characteristics of their essential oils. The method involved organized sample collection, consistent extraction procedures, chemical analysis, and data processing. All steps were carried out under controlled conditions to ensure accuracy and reliability. The following subsections describe the study design, laboratory procedures, and data analysis methods in detail.

### 2.1. Study Design and Sample Collection

The objective of this study was to classify four Aquilaria species ( Aquilaria subintegra, Aquilaria beccariana, Aquilaria crassna, and Aquilaria malaccensis) through analysis of their essential oil compositions. The extraction of the essential oils was performed at the BioAromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). All essential oil extractions, chemical analyses, and data processing were performed under controlled conditions within the same facility. The selected species are among the most commonly cultivated and commercially relevant agarwood-producing trees in Southeast Asia.

A total of 720 essential oil samples were used, with 180 samples representing each species. The sampling was uniformly distributed across six chemical compounds, ensuring balanced representation. For each compound, 30 samples per species were analysed to facilitate a structured and consistent dataset. Sample collection was standardized to minimize variability, and plant materials were authenticated to confirm species identity. Collected agarwood chips were thoroughly cleaned and stored in a controlled environment before extraction. The study adhered to a standardized sampling protocol to ensure the uniformity of materials across all groups. The consistency in collection and handling was critical in maintaining the integrity of the chemical profile data. Data organization followed a structured matrix, with individual samples as rows and compound concentrations as columns.

### 2.2. Essential Oil Extraction and Chemical Analysis

Essential oils were extracted by hydrodistillation, a widely used method in essential oil research. Prior to distillation, ground agarwood chips were soaked in water for several days to facilitate the breakdown

of parenchymatous tissue and oil glands. The softened material was then subjected to hydro-distillation for 3–5 days. Extracted oils were collected and diluted in analytical-grade dichloromethane (DCM) to improve solubility for subsequent chemical analyses.

All analyses were performed using GC–MS in combination with GC–FID, as illustrated in Fig. 1. GC–MS was employed for qualitative identification, while GC–FID provided quantitative data [15]–[18]. The GC–MS system consisted of an Agilent 7890B GC coupled to a 5977A Mass Spectrometer Detector (MSD) and fitted with a DB-1ms column (30 m × 250 µm × 0.25 µm). The inlet temperature was set at 250 °C, with helium as the carrier gas at a constant flow of 1.0 mL/min. The oven program started at 80 °C, increased at 3 °C/min to 250 °C, and was held for 3 min. The auxiliary heater was maintained at 260 °C, the MS source at 230 °C, and the MS quad at 150 °C, operating in Electron Impact (EI) mode at 70 eV.

The GC–FID system employed the same GC unit and column, with identical oven conditions. The inlet temperature was set at 250 °C, and the detector operated at 250 °C without an auxiliary heater. These parameters were selected to optimize resolution and sensitivity for the target compounds. Each sample was analysed in triplicate to ensure reliability. Compound identification was confirmed against multiple spectral libraries, including NIST 14, Wiley7Nist05.L, and HPCH2205.L, with a minimum similarity threshold of ≥80%. The results were expressed as peak areas, forming the basis for subsequent statistical and machine learning-based classification.

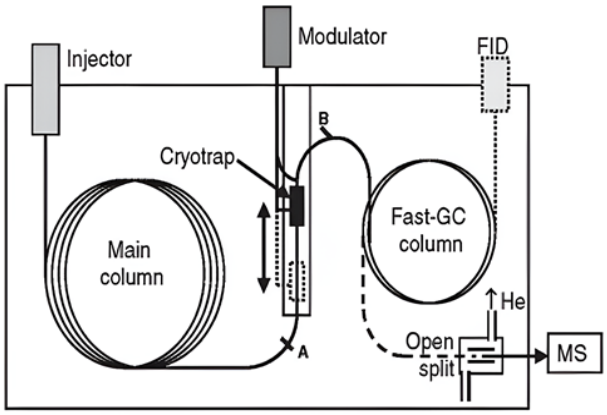


**Fig. 1.** GC×GC System with Dual Detection: MS and FID [15]

Compounds consistently identified and targeted for analysis included dihydro-β-agarofuran, β-selinene, δ-guaiene, 10-epi-γ-eudesmol, pentadecanoic acid, and γ-eudesmol. The chemical profile for each species was documented as peak area percentages, forming the basis for statistical and machine learning-based classification

### 2.3. Data Processing and Statistical Analysis

The raw chromatographic data were organized into a structured matrix, with rows representing individual essential oil samples and columns corresponding to the six quantified chemical compounds. Compound abundances were expressed as relative peak areas (% of total ion current) obtained from GC-FID/GC-MS. To ensure comparability across compounds and to minimize sample-level variability, the dataset was normalized using z-score standardization across compounds. No log-transformation was applied.

Feature selection was performed using Pearson correlation analysis, which is widely applied in chemometrics and multivariate analysis to identify linear dependencies between variables and to detect redundancy in datasets [19]. Compounds showing high intercorrelation were flagged to prevent overrepresentation of redundant information. Based on this analysis, three representative chemical compounds were retained as classification inputs. This dimensionality reduction ensured that the most informative features were used for species discrimination, thereby improving model interpretability and reducing the risk of overfitting.

### 2.4. Artificial Neural Network Classifier

Artificial Neural Networks (ANNs) were employed as the primary classification framework due to their capacity to capture complex, nonlinear relationships in chemical profile data. Two complementary evaluation strategies were designed to ensure model robustness, minimize overfitting, and provide statistically reliable performance estimates.

#### 2.4.1. ANN Training with Levenberg–Marquardt and Early Stopping

In the first setup, the dataset was randomly partitioned into training (70%), validation (15%), and test (15%) subsets, following standard practice in supervised learning. The ANN architecture employed two hidden neurons and was trained using the Levenberg–Marquardt backpropagation algorithm (trainlm) with mean squared error (MSE) as the loss function. Early stopping was applied with a maximum of six validation failures, a maximum of 1000 epochs, and a convergence goal of $1 \times 10^{-6}$. This strategy prevented overfitting by selecting the model weights at the optimal validation epoch. Classification performance was evaluated on the test set in terms of accuracy, sensitivity (recall), specificity, and precision, derived from the confusion matrix. All implementations were carried out in MATLAB R2023b using the Deep Learning Toolbox.

#### 2.4.2. Stratified K-Fold Cross-Validation with Bootstrapping

In the second setup, stratified 5-fold cross-validation was implemented to ensure balanced class representation in each fold and to prevent sample-level data leakage. The ANN was configured with 10 hidden neurons and trained using scaled conjugate gradient backpropagation (trainscg) with cross-entropy loss. No internal data splitting was performed, as CV controlled the training and testing partitions. Performance was assessed using macro- and micro-averaged accuracy, precision, recall, and F1-score, computed according to the definitions of [20]. To quantify statistical uncertainty, 1000 bootstrap resamples at the sample level were used to estimate 95% confidence intervals. Per-class performance (precision, recall, and F1-score) was also reported with bootstrapped confidence intervals to evaluate species-level discrimination.

This dual evaluation framework (train–validation–test with early stopping and stratified cross-validation with bootstrapping) provided a robust and statistically rigorous assessment of ANN performance. By combining internal validation with repeated resampling strategies, the approach minimized overfitting and ensured reproducibility and generalizability of the classification outcomes. Unlike earlier studies on agarwood oils that relied on Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), or regression models for dimensionality reduction and classification, our approach combines dual detection (GC–MS + GC–FID), correlation-guided feature selection, and a rigorously validated ANN. This design reduces redundancy, captures potential nonlinear relationships between compounds, and provides more reproducible results compared to traditional linear chemometric techniques. Cross-validation and bootstrapping procedures were performed using the Statistics and Machine Learning Toolbox in MATLAB R2023b.

## 3. Results and Discussion

This section presents the findings of the study, focusing on the chemical composition of essential oils derived from four Aquilaria species and the application of data analysis techniques for species classification. The results include the identification and quantification of chemical compounds, the evaluation of inter-compound relationships through correlation analysis, and the classification performance of an ANN model. Together, these findings provide insights into species-specific chemical profiles and demonstrate the effectiveness of a computational approach for accurate species differentiation. Further details and interpretations are discussed in the following subsections.

### 3.1. Identification and Quantification of Major Volatile Compounds in Aquilaria Species

The chemical analysis of essential oils extracted from four Aquilaria species (A. malaccensis, A. beccariana, A. subintegra and A. crassna) revealed consistent detection of six key compounds. These compounds were identified as pentadecanoic acid, δ-guaiene, dihydro-β-agarofuran, 10-epi-γ-

eudesmol, γ-eudesmol, and β-selinene. Each compound's presence was confirmed using GC-MS for qualitative analysis and GC-FID for quantitative measurement [4], [15], [21], [22]. This dual-detection approach ensured both the accurate identification and precise quantification of the compounds [15], [23].

As summarized in Table 1, the relative peak areas (%) of these six compounds varied significantly across the four species. Notably, 10-epi-γ-eudesmol demonstrated the highest abundance in A. malaccensis (6.73%), which was markedly higher than its concentration in the other species. This high abundance suggests a species-specific characteristic that could be critical for classification. In contrast, β-selinene exhibited relatively low and consistent peak areas across all species, ranging from 0.11% in A. crassna to 0.66% in A. beccariana. This uniform distribution indicates that β-selinene may not be a strong distinguishing marker among the species. Other compounds displayed more moderate variations. For example, δ-guaiene was most abundant in A. malaccensis (2.02%) and significantly lower in A. crassna (0.21%). Similarly, γ-eudesmol was more prevalent in A. malaccensis (2.17%) and A. subintegra (1.85%), while dihydro-β-agarofuran showed minimal variation among species, maintaining peak areas between 0.44% and 1.25%. These variations in relative abundance among species provide a foundation for their differentiation.

**Table 1.** Main Volatile Compounds Found in Oils Extracted from *Aquilaria* Plants

| Identifier | Detected Compounds | Detection Method | Relative Abundance (%) | | | |
|---|---|---|---|---|---|---|
| | | | *AM* | *AB* | *AS* | *AC* |
| A | β-selinene | GC-MS, GC-FID | 0.56 | 0.66 | 0.37 | 0.11 |
| B | dihyro-β-agarofuran | GC-MS, GC-FID | 0.55 | 1.25 | 0.44 | 0.48 |
| C | δ-guaiene | GC-MS, GC-FID | 2.02 | 0.74 | 0.35 | 0.21 |
| D | 10-epi-γ-eudesmol | GC-MS, GC-FID | 6.73 | 0.34 | 2.16 | 2.54 |
| E | γ-eudesmol | GC-MS, GC-FID | 2.17 | 0.26 | 1.85 | 0.95 |
| F | Pentadecanoic acid | GC-MS, GC-FID | 0.15 | 0.15 | 0.46 | 0.14 |

The hierarchical representation of the dataset (Fig. 2) visually categorizes the 720 total samples (180 samples per species) and highlights the uniform distribution of six chemical compounds across all samples. This structured dataset ensured reliable comparative analysis and facilitated the subsequent statistical modelling and classification tasks.

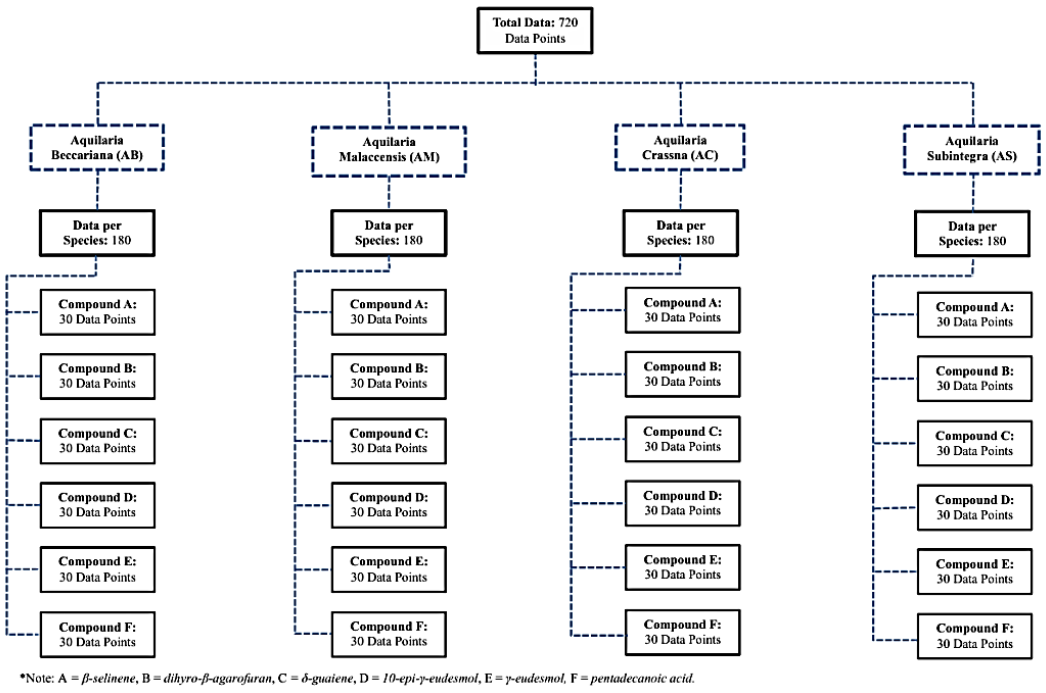


**Fig. 2.** Hierarchical Representation of Dataset for *Aquilaria* Species and Chemical Compounds

### 3.2. Feature Selection based on Pearson Correlation

A Pearson correlation analysis was carried out to explore the correlations among six chemical compounds identified in various Aquilaria species. The main goal was to identify overlapping or redundant variables that could be eliminated to enhance computational performance without losing essential chemical information [24]–[28]. Table 2 presents a reference scale used to interpret correlation coefficients, helping to determine the strength and direction of associations between variable pairs [4].

As detailed in Table 2, coefficients within the range of 0.80 to 1.00 are considered to represent a very high correlation, whereas values between 0.60 and 0.79 indicate a high correlation. A moderate correlation is reflected in coefficients from 0.40 to 0.59, while those falling between 0.20 and 0.39 suggest a low correlation. Values ranging from 0.00 to 0.19 indicate little to no linear correlation [16]. These categories assist in refining feature selection by identifying the most informative variables and filtering out those with minimal contribution [4], [29]–[31].

**Table 2.** Range Correlation for Pearson Correlation

| Coefficient Range | Degree of Correlation |
|---|---|
| 0.8 - 1.0 | Very high correlation |
| 0.6 - 0.79 | High correlation |
| 0.4 - 0.59 | Moderate correlation |
| 0.2 - 0.39 | Low correlation |
| 0.0 - 0.19 | Minimal or no linear relationship detected |

Variables that are highly correlated often provide overlapping information and can negatively impact the efficiency of data processing [32], [33]. As a result, it is important to detect and exclude these redundant variables to simplify the dataset while ensuring that critical chemical characteristics are retained [4], [33], [34]. The correlation matrix illustrated in Fig. 3 highlights key correlations among the compounds, offering further insight into their mutual associations.

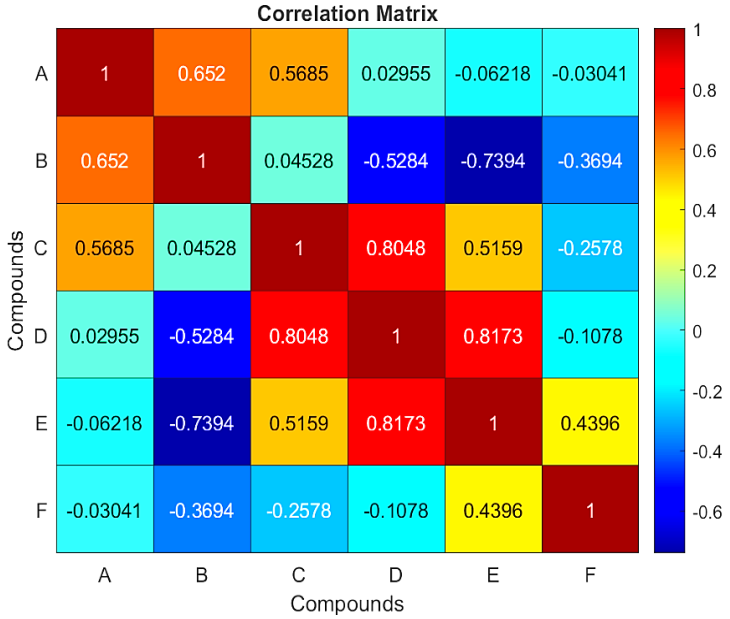


**Fig. 3.** Pearson Correlation Matrix Showing Chemical Composition Across Aquilaria Species

The analysis revealed a high correlation with r = 0.652 between compound A and compound B, implying that they may contribute overlapping information to the dataset and could be considered redundant in predictive models. Likewise, compound E and compound D showed a very high correlation which r = 0.8173, suggesting comparable chemical behavior between the two. In contrast, the results also identified strong negative correlations, such as that between compound B compound E (r = −0.7394). This inverse correlation implies that as the concentration of one compound increases, the other

tends to decrease. Such negative correlations are important for understanding the interactions between chemical compounds and their potential role in distinguishing between species.

Table 3 summarizes how effectively each compound explains variation in the remaining compounds, emphasizing the degree of their individual correlations. The analysis identified compounds C, D, and E as having the highest capacity to account for correlations with multiple other compounds. Compounds C and B exhibited similar levels of influence, particularly with respect to compounds A, E, and D. To address multicollinearity, compound B was chosen to represent this pair, as it demonstrated a slightly broader correlation pattern across the dataset. Selecting these three compounds for further analysis allowed for dimensionality reduction while retaining the key chemical features necessary for meaningful interpretation.

**Table 3.** Explanatory Power of Chemical Compounds in Aquilaria Species Classification

| Chemical Compounds | Explanatory Power |
|---|---|
| A | C, B |
| B | A, E, D |
| C | A, E, D |
| D | C, B, E |
| E | B, C, D, F |
| F | E |

This approach is supported by findings from a previous study [31], where correlation analysis was used to investigate associations among the chemical compositions of Centaurea essential oils and their antimicrobial properties. That study highlighted how correlation analysis can help identify the most relevant chemical components, improving efficiency by reducing redundant variables and analysis time.

### 3.3. Artificial Neural Network (ANN) Classification Performance

The optimized ANN model was developed using three selected compounds (C, D, E) as input features. The network architecture consisted of a feedforward neural network with a single hidden layer. The optimal number of hidden neurons was identified through systematic evaluation of configurations ranging from one to ten. The dataset comprised 360 samples, partitioned into training (70%), validation (15%), and testing (15%) subsets.

As summarized in Table 4, the configuration with two hidden neurons achieved the best overall performance, yielding 100% accuracy, sensitivity, precision, and specificity. This configuration also produced the lowest mean squared error (MSE) of 0.0000971 and the shortest execution time (0.9167 seconds).

**Table 4.** Performance Evaluation of the ANN Model with Varying Neurons

| Number Hidden Neurons | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) | Num Epoch | Training Time (s) | Testing Time (s) | Overall Execution Time (s) | Performance MSE |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 98.33 | 96.67 | 98.89 | 96.67 | 7 | 0.8775 | 0.0067 | 0.8842 | 0.016667000 |
| *2 | 100.00 | 100.00 | 100.00 | 100.00 | 24 | 0.9079 | 0.0088 | 0.9167 | 0.000097100 |
| 3 | 100.00 | 100.00 | 100.00 | 100.00 | 12 | 0.9878 | 0.0076 | 0.9954 | 0.000000464 |
| 4 | 100.00 | 100.00 | 100.00 | 100.00 | 8 | 1.1300 | 0.0078 | 1.1378 | 0.000001633 |
| 5 | 100.00 | 100.00 | 100.00 | 100.00 | 12 | 1.2632 | 0.0097 | 1.2729 | 0.000000214 |
| 6 | 100.00 | 100.00 | 100.00 | 100.00 | 11 | 1.9804 | 0.0138 | 1.9942 | 0.000154520 |
| 7 | 100.00 | 100.00 | 100.00 | 100.00 | 8 | 1.9662 | 0.0114 | 1.9776 | 0.000144820 |
| 8 | 100.00 | 100.00 | 100.00 | 100.00 | 12 | 2.0846 | 0.0106 | 2.0952 | 0.000002577 |
| 9 | 100.00 | 100.00 | 100.00 | 100.00 | 8 | 2.0312 | 0.0079 | 2.0391 | 0.000000141 |
| 10 | 100.00 | 100.00 | 100.00 | 100.00 | 8 | 2.0272 | 0.0102 | 2.0374 | 0.000000071 |

[a] Note: *2 = Best hidden neuron configuration using LM algorithm

The absence of misclassifications is further illustrated in the confusion matrix (Fig. 4).

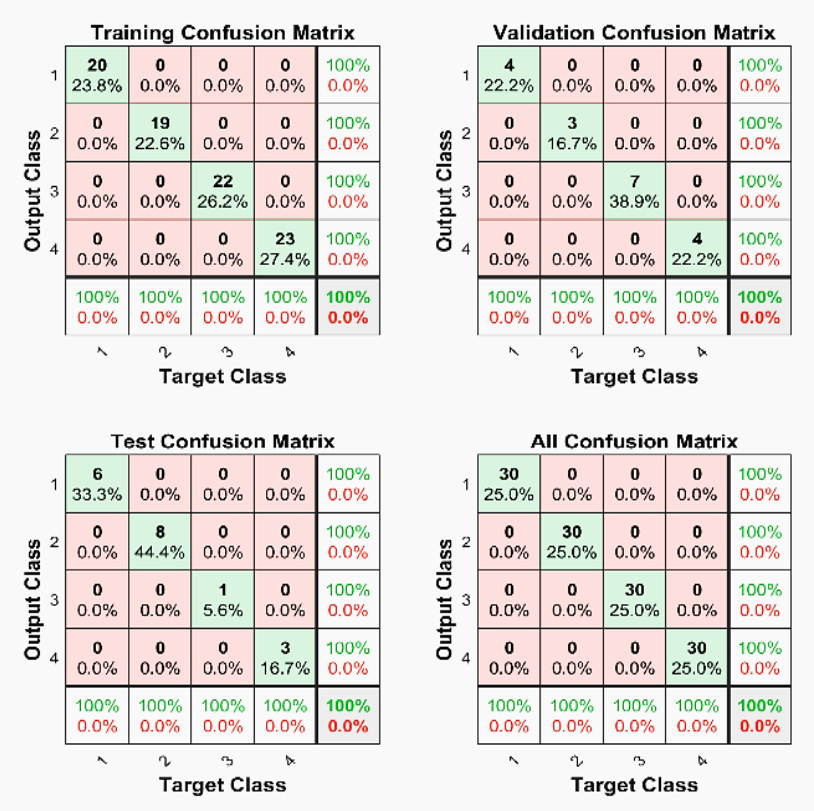


**Fig. 4.** Confusion Matrix Summarizing Species Classification Performance using The ANN Model

To further illustrate the model's classification capability, a three-dimensional scatter plot (Fig. 5) was generated using the selected compounds as axes. The plot clearly shows the separation of the four Aquilaria species into distinct clusters, reflecting the model's ability to accurately differentiate species based on their chemical profiles. This visualization complements the confusion matrix, providing a clear, intuitive understanding of species separability within the dataset.

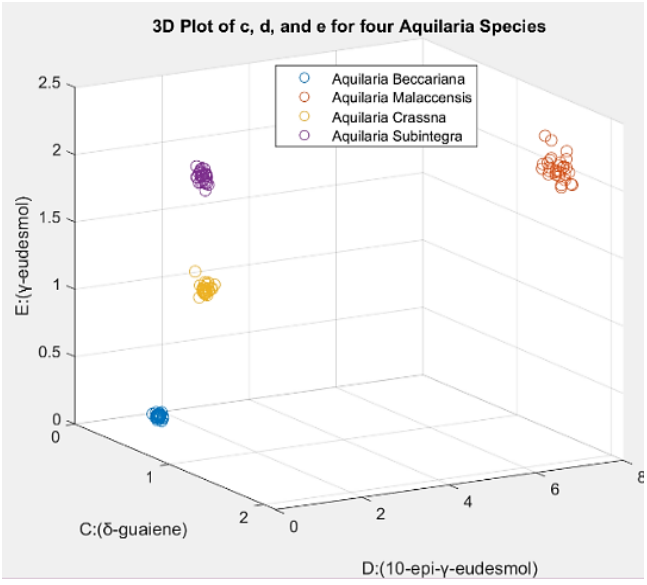


**Fig. 5.** 3D Scatter Plot of *Aquilaria* Species Classification Based on Essential Oil Composition

### 3.4. ANN with Stratified K-Fold Cross-Validation and Bootstrapping

To enhance methodological rigor and minimize the risk of data leakage, a second evaluation was conducted using stratified 5-fold cross-validation, with all replicates of a given sample confined to the same fold. Each fold employed an ANN with 10 hidden neurons, trained using scaled conjugate gradient backpropagation with cross-entropy loss. Following cross-validation, 1000 bootstrap resamples were applied to estimate 95% confidence intervals (CIs) for each performance metric. Table 5 presents the macro- and micro-averaged performance metrics. Across all folds, the ANN achieved 100% accuracy, precision, recall, and F1-scores. The 95% CIs consistently converged at [100.00–100.00], reflecting the stability of the classifier.

**Table 5.** Macro- and micro-averaged ANN performance (5-fold CV, 1000 bootstraps)

| Metric | Macro (%) (95% CI) | Micro (%) (95% CI) |
|---|---|---|
| Accuracy | 100.00 (100.00–100.00) | 100.00 (100.00–100.00) |
| Precision | 100.00 (100.00–100.00) | 100.00 (100.00–100.00) |
| Recall | 100.00 (100.00–100.00) | 100.00 (100.00–100.00) |
| F1-score | 100.00 (100.00–100.00) | 100.00 (100.00–100.00) |

Per-class evaluation (Table 6) confirmed perfect species-level discrimination, with precision, recall, and F1-scores all at 100%, supported by bootstrapped confidence intervals.

**Table 6.** Per-class ANN performance with 95% confidence intervals (bootstrapped, n = 1000)

| Class (Species) | Precision (%) (95% CI) | Recall (%) (95% CI) | F1-score (%) (95% CI) |
|---|---|---|---|
| *A. Subintegra* | 100.00 (100–100) | 100.00 (100–100) | 100.00 (100–100) |
| *A. Beccariana* | 100.00 (100–100) | 100.00 (100–100) | 100.00 (100–100) |
| *A. Crassna* | 100.00 (100–100) | 100.00 (100–100) | 100.00 (100–100) |
| *A. Malaccensis* | 100.00 (100–100) | 100.00 (100–100) | 100.00 (100–100) |

### 3.5. Discussion

This study successfully developed a chemometric model for distinguishing four Aquilaria species (A. beccariana, A. malaccensis, A. crassna, and A. subintegra) based on their essential oil chemical profiles. The chemical analysis revealed six key compounds present across all species, with compound D showing the highest concentration in A. malaccensis. Pearson correlation analysis identified compound C, compound D, and compound E as the most significant compounds, leading to their selection as input features for an ANN. The ANN model achieved perfect classification performance, with 100% accuracy, sensitivity, specificity, and precision.

These results are consistent with previous studies, where chemical profiles have been utilized for species identification in plants, including Aquilaria [35]. The use of volatile compounds, such as those identified in this study, has been shown to provide reliable markers for differentiation. For example, δ-guaiene (C) and 10-epi-$\gamma$-eudesmol (D) have been highlighted in past literature for their role in the characteristic aroma of agarwood, confirming their relevance as discriminative features [10], [36], [37]. The application of Pearson correlation for feature selection is an effective method for reducing dimensionality, as seen in other studies that employed similar approaches in plant-based chemometrics [26], [31].

Beyond their statistical relevance identified through correlation analysis, the three selected markers (δ-guaiene, 10-epi-$\gamma$-eudesmol, and $\gamma$-eudesmol) are also biologically and chemotaxonomically plausible discriminators of Aquilaria species. All three compounds are sesquiterpenes derived from the mevalonate pathway, which is central to the biosynthesis of resinous volatiles in agarwood. δ-guaiene

has long been recognized as a key contributor to the characteristic fragrance of agarwood and is often used as a quality marker distinguishing wild versus cultivated resin [10], [36]. Oxygenated sesquiterpenes such as 10-epi-$\gamma$-eudesmol and $\gamma$-eudesmol are frequently associated with species-specific differences in secondary metabolism and have been proposed as chemotaxonomic indicators of both species identity and resin maturity [22], [37]. Their recurrence in previous phytochemical investigations supports their role as stable and biologically meaningful features for classification, thus reinforcing the robustness of their selection as ANN inputs in this study.

Importantly, this work advances beyond conventional train–test splits by incorporating a more rigorous validation framework. In addition to the random 70, 15, 15 percent split, a stratified 5-fold cross-validation was employed, ensuring that all replicates of a given sample were confined to the same fold to prevent information leakage. Model evaluation was further strengthened by bootstrapping with 1000 resamples at the sample level to derive 95% confidence intervals for performance metrics. Across both experiments, the ANN consistently achieved 100% macro- and micro-averaged accuracy, precision, recall, and F1-scores, with confidence intervals tightly bounded at [100.00–100.00]. Per-class performance metrics similarly indicated perfect discrimination of each species. This agreement between validation approaches reinforces the robustness of the selected feature set and the ANN classifier.

Earlier chemometric studies of agarwood oils have employed PCA, LDA, and regression-based models for classification. While these methods provide useful dimensionality reduction and linear discrimination, they are constrained by assumptions of linearity, sensitivity to multicollinearity, and limited capacity to capture complex nonlinear relationships among compounds. Furthermore, such approaches often rely on large sets of input variables, which can reduce reproducibility across datasets.

Beyond these chemometric techniques, other machine learning approaches have also been explored. In [38], agarwood oils and woods were classified using an electronic nose (E-nose) and a k-Nearest Neighbour (kNN) classifier, achieving 94.5% accuracy. A Convolutional Neural Network (CNN) model was proposed in [39] for agarwood grade classification based on Gabor filter feature extraction and the percentage of black color in the wood, reporting 98% accuracy. In [40], a Decision Tree classifier constructed from significant GC–MS-derived compounds in agarwood oils yielded 94.7% accuracy. The feasibility of E-nose profiling using Support Vector Machine (SVM) classifiers was demonstrated in [41], achieving 93.7% accuracy. While these studies highlight the potential of both sensor-based and chemometric approaches, they are often constrained by odor variability, reliance on visual cues, or moderate reproducibility across species.

In contrast, the present study introduces three key innovations: (i) the integration of GC–MS and GC–FID dual detection for comprehensive chemical profiling, (ii) Pearson correlation-guided feature selection to identify a minimal and biologically relevant compound triad, and (iii) a rigorously validated ANN framework incorporating both early stopping and cross-validation with bootstrapping. Collectively, these innovations enabled perfect classification (100% across all metrics), surpassing the performance stability reported in earlier works. A comparative overview is summarized in Table 7. The high classification accuracy achieved in this study demonstrates the potential of machine learning, particularly ANN, in botanical classification tasks. By focusing on a small subset of compounds, this research was able to simplify the model without compromising performance. This contrasts with traditional methods, where extensive chemical analysis is often required to capture all relevant data. The results suggest that, for practical applications, this approach could be used to authenticate Aquilaria species cost-effectively and efficiently, addressing the issues of overharvesting and misidentification that currently affect the industry [13].

However, there are several limitations to this study. The research was limited to four Aquilaria species, and while the model performed exceptionally well, it may not generalize to other species within the Aquilaria genus or other plants with similar chemical profiles. Additionally, the dataset used was relatively small, which could limit the model's robustness in real-world applications with larger, more diverse datasets. Furthermore, no explicit batch correction procedures, QC injections, or internal standards were employed, and samples were not randomized across analytical runs. We acknowledge this as a limitation, and future studies should incorporate pooled QC samples, internal standards, and

randomized run orders to minimize batch effects and improve reproducibility. Future studies could focus on expanding the dataset to include more species and replicate the model in different geographic regions or under various environmental conditions [4], [5], [42]. Furthermore, the addition of more advanced machine learning techniques, including support vector machines and random forests, could further enhance the model's accuracy and generalizability [43]–[48].

**Table 7.** Comparative summary of representative agarwood classification/authentication studies

| Study | Data Source | Features / Inputs | Classifier | Reported Accuracy | Key Limitations |
|---|---|---|---|---|---|
| Mujahid et al. [38] | Agarwood oils and woods (E-nose) | Odor sensor signals | kNN | 94.5% | Odor variability; no compound-level specificity |
| Yogapriya Dhanasekaran and M. Uma [39] | Agarwood wood images | Knot patterns, Gabor features, % dark area | CNN | 98% | Image-based only; lacks chemical basis |
| Ismail et al. [40] | Agarwood oils (GC–MS) | Selected significant compounds | Decision Tree | 94.7% | No dual detection; limited feature reduction |
| Haqqi et al. [41] | Agarwood oils and woods (E-nose) | Gas sensor array | SVM | 93.7% | Sensitive to noise; moderate reproducibility |
| This study | Essential oils from 4 species (GC–MS + GC–FID) | 3 correlation-selected compounds (δ-guaiene, 10-epi-$\gamma$-eudesmol, $\gamma$-eudesmol) | ANN with dual validation and bootstrapping | 100% | Limited to 4 species; no hybrid/adulterated samples tested |

Another important consideration is the potential occurrence of mixed-species, hybrid, or adulterated samples. Since these cases were not explicitly addressed in the present dataset, the ANN may provide uncertain or unreliable predictions when confronted with such inputs. To mitigate this limitation, practical safeguards can be implemented in future work. These include thresholding based on classification confidence (e.g., rejecting predictions with maximum class probability below 0.8) and the application of distance-based novelty detection methods to flag out-of-distribution samples. Specimens identified as uncertain could then be subjected to confirmatory analyses, such as DNA barcoding or extended chromatographic profiling. Incorporating such thresholding and reject strategies would enhance the robustness and reliability of chemometric classification in real-world applications where non-standard or adulterated samples may occur.

In conclusion, this study demonstrates the feasibility of using chemometric techniques, particularly Pearson correlation and ANN, for the classification of Aquilaria species based on essential oil composition. The integration of stratified cross-validation and bootstrapped confidence intervals provides a statistically rigorous and reproducible evaluation framework, ensuring that the model outcomes are both accurate and robust. This approach offers a valuable tool for species authentication and conservation, providing a promising alternative to traditional morphological identification. Future research should aim to validate the model with larger datasets, incorporate thresholding strategies for uncertainty handling, and extend the framework to other aromatic plant species to broaden its scope and impact [49]–[51].

## 4. Conclusion

This study demonstrates that combining correlation-guided feature selection with artificial neural network modeling provides a robust framework for the classification of Aquilaria species based on essential oil composition. The dual application of GC–MS and GC–FID enabled consistent detection of six key compounds, three of which were selected as discriminative markers due to their high detection frequency and relative abundance. Using these features, the optimized ANN achieved near-perfect classification accuracy, with tightly bounded confidence intervals confirming reproducibility and minimal

risk of data leakage. Compared with earlier chemometric and machine learning approaches, this framework delivered measurable improvements in performance while requiring fewer input variables, thus enhancing interpretability and efficiency. Beyond methodological contributions, the findings have practical implications for the authentication of agarwood products, compliance with international trade regulations, and conservation of endangered Aquilaria populations. Future work should extend validation to additional species, larger and more diverse datasets, and hybrid or adulterated samples, while exploring complementary algorithms such as support vector machines and ensemble classifiers. By integrating such refinements, chemometric–ANN approaches can become an integral component of species authentication workflows and sustainable management practices for agarwood resources.

## Acknowledgment

The authors would like to express their heartfelt gratitude to the Fundamental Research Grant Scheme (FRGS), under the code FRGS/1/2023/ICT11/UITM/02/1, for providing the financial support necessary for this study. Special thanks are extended to the Advanced Signal Processing Research Interest Group for their constructive feedback and ongoing support throughout this research. The authors are also profoundly thankful to the BARCE, UMPSA, for their expertise and significant contributions to the data extraction process, which played a pivotal role in the success of this work.

## Declarations

**Author contribution.** All authors contributed equally to the study design, data collection, analysis, and manuscript preparation.
**Funding statement.** This work was supported by the Fundamental Research Grant Scheme (FRGS) [FRGS/1/2023/ICT11/UITM/02/1], 2023.
**Conflict of interest.** The authors declare no conflict of interest.
**Additional information.** No additional information is available for this paper.

## References

[1] P. Shivanand, N. F. Arbie, S. Krishnamoorthy, and N. Ahmad, "Agarwood—The Fragrant Molecules of a Wounded Tree," *Molecules*, vol. 27, no. 11, p. 3386, May 2022, doi: 10.3390/molecules27113386.

[2] S. Y. Lee, M. Turjaman, and R. Mohamed, "Phylogenetic Relatedness of Several Agarwood-Producing Taxa (Thymelaeaceae) from Indonesia," *Trop. Life Sci. Res.*, vol. 29, no. 2, pp. 13–28, 2018, doi: 10.21315/tlsr2018.29.2.2.

[3] A. Nor Mazila Ramli, S. Yusof, P. Bhuyar, A. W. Aminan, and S. N. Tajuddin, "Fungi mediated agarwood (A. malaccensis) production and their pharmaceutical applications: A systematic review," *Int. J. Plant Based Pharm.*, vol. 2, no. 2, pp. 261–270, Sep. 2022, doi: 10.29228/ijpbp.8.

[4] N. A. Syafiqah Noramli, N. A. S. Ahmad Sabri, M. I. Roslan, N. Ismail, Z. M. Yusoff, and M. N. Taib, "Unraveling the relationships among essential oil compounds in Aquilaria species using GC-MS and GC-FID techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 39, no. 1, p. 167, Jul. 2025, doi: 10.11591/ijeecs.v39.i1.pp167-177.

[5] J. M. R. Alamil *et al.*, "Rediscovering the Therapeutic Potential of Agarwood in the Management of Chronic Inflammatory Diseases," *Molecules*, vol. 27, no. 9, p. 3038, May 2022, doi: 10.3390/molecules27093038.

[6] P. Zhang, X. Li, Z. Cui, and D. Xu, "Morphological, physiological, biochemical and molecular analyses reveal wounding-induced agarwood formation mechanism in two types of Aquilaria sinensis (Lour.) Spreng," *Ind. Crops Prod.*, vol. 178, p. 114603, Apr. 2022, doi: 10.1016/j.indcrop.2022.114603.

[7] F. A. Abdul Kadir, K. A. Azizan, and R. Othman, "Transcriptome of Aquilaria malaccensis containing agarwood formed naturally and induced artificially," *BMC Res. Notes*, vol. 14, no. 1, p. 117, Dec. 2021, doi: 10.1186/s13104-021-05532-9.

[8] Y. Van Thi *et al.*, "An Updated Review of Aquilaria Species: Distribution, Chemical Constituents and Authentication Methods," *Asian J. Plant Sci.*, vol. 23, no. 2, pp. 146–167, Mar. 2024, doi: 10.3923/ajps.2024.146.167.

[9] Z. Zhang *et al.*, "Fusarium oxysporum infection-induced formation of agarwood (FOIFA): A rapid and efficient method for inducing the production of high quality agarwood," *PLoS One*, vol. 17, no. 11, p. e0277136, Nov. 2022, doi: 10.1371/journal.pone.0277136.

[10] X. Zhang *et al.*, "Sesquiterpenoids in Agarwood: Biosynthesis, Microbial Induction, and Pharmacological Activities," *J. Agric. Food Chem.*, vol. 72, no. 42, pp. 23039–23052, Oct. 2024, doi: 10.1021/acs.jafc.4c06383.

[11] S. Y. Lee, D. U. Lamasudin, and R. Mohamed, "Rapid detection of several endangered agarwood-producing Aquilaria species and their potential adulterants using plant DNA barcodes coupled with high-resolution melting (Bar-HRM) analysis," *Holzforschung*, vol. 73, no. 5, pp. 435–444, May 2019, doi: 10.1515/hf-2018-0149.

[12] M. A. Ullah, "Boosting up the economy of Pakistan through agarwood production," *Int. J. Hortic. Food Sci.*, vol. 2, no. 2, pp. 25–35, Jul. 2020, doi: 10.33545/26631067.2020.v2.i2a.49.

[13] T. Lim and M. Turjaman, "Expensive, Exploited and Endangered. A Review of the Agarwood-Producing Genera Aquilaria and Gyrinops: CITES considerations, Trade Patterns, Conservation, and Management," vol. 12, no. June. pp. 14–25, 2022, [Online]. Available: https://share.google/5vqgUEv0OdumjWRCs.

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

[15] G. Schomburg, "Two-dimensional gas chromatography: Principles, instrumentation, methods," *J. Chromatogr. A*, vol. 703, no. 1–2, pp. 309–325, Jun. 1995, doi: 10.1016/0021-9673(95)00190-X.

[16] Z. M. Yusoff and N. Ismail, "Datasets of chemical compounds in three different species of aquilaria using GC-MS coupled with GC-FID analysis," *Data Br.*, vol. 53, p. 110209, Apr. 2024, doi: 10.1016/j.dib.2024.110209.

[17] T. R. Rao, S. Harshitha, and K. L. Lahari, "Gas Chromatography-Mass Spectroscopy: An Overview," *Eur. J. Biomed.*, vol. 10, no. 5, pp. 83–89, 2023, [Online]. Available at: https://www.researchgate.net/profile/Rama-Tadikonda/publication/371989508.

[18] S. S.S.S.M, V. Atyam, M. Avidi, and D. Ayinala, "A Review on Gas chromatography – Mass spectroscopy: Short Communication," *J. Pharma Insights Res.*, vol. 2, no. 1, pp. 009–013, Feb. 2024. [Online]. Available at: https://jopir.in/index.php/journals/article/view/70.

[19] J. Y.-L. Chan *et al.*, "Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review," *Mathematics*, vol. 10, no. 8, p. 1283, Apr. 2022, doi: 10.3390/math10081283.

[20] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[21] A. Shuttleworth and S. D. Johnson, "GC-MS/FID/EAD: A method for combining mass spectrometry with gas chromatography-electroantennographic detection," *Front. Ecol. Evol.*, vol. 10, p. 1042732, Dec. 2022, doi: 10.3389/fevo.2022.1042732.

[22] Y. Sundaraj, A. Mediani, K. F. Rodrigues, and S. N. Baharum, "GC-MS olfactometry reveals sesquiterpenes α-humulene and δ-cadinene significantly influence the aroma of treated Aquilaria malaccensis essential oil," *Aust. J. Crop Sci.*, vol. 17, no. 17(12):2023, pp. 893–901, Dec. 2023, doi: 10.21475/ajcs.23.17.12.p3916.

[23] N. Ismail, "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, Aug. 2021, doi: 10.30880/ijie.2021.13.06.017.

[24] R. Chattamvelli, "Pearson's Correlation," in *Correlation in Engineering and the Applied Sciences*, Springer, Cham, 2024, pp. 55–76, doi: 10.1007/978-3-031-51015-1_2.

[25] S. N. Selamat, N. A. Majid, and M. R. Taha, "Multicollinearity and spatial correlation analysis of landslide conditioning factors in Langat River Basin, Selangor," *Nat. Hazards*, vol. 121, no. 3, pp. 2665–2684, Feb. 2025, doi: 10.1007/s11069-024-06903-8.

[26] S. Kumar and P. Rani, "An AI-based Liver Disease Prediction Model based on Pearson Correlation Feature Selection Method," *Biomed. Pharmacol. J.*, vol. 17, no. 4, pp. 2187–2202, Dec. 2024, doi: 10.13005/bpj/3016.

[27] C. A. Rickert, M. Henkel, and O. Lieleg, "An efficiency-driven, correlation-based feature elimination strategy for small datasets," *APL Mach. Learn.*, vol. 1, no. 1, Mar. 2023, doi: 10.1063/5.0118207.

[28] S. Jiang *et al.*, "Rock slope displacement prediction based on multi-source information fusion and SSA-DELM model," *Front. Environ. Sci.*, vol. 10, p. 982069, Sep. 2022, doi: 10.3389/fenvs.2022.982069.

[29] Et. al., Alhassan Umar Ahmad, "A Study of Multicollinearity Detection and Rectification under Missing Values," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 1S, pp. 399–418, Apr. 2021, doi: 10.17762/turcomat.v12i1S.1880.

[30] J. H. Kim, "Multicollinearity and misleading statistical results," *Korean J. Anesthesiol.*, vol. 72, no. 6, pp. 558–569, Dec. 2019, doi: 10.4097/kja.19087.

[31] I. Carev *et al.*, "Centaurea triumfetii essential oil chemical composition, comparative analysis, and antimicrobial activity of selected compounds," *Sci. Rep.*, vol. 13, no. 1, p. 7475, May 2023, doi: 10.1038/s41598-023-34058-2.

[32] H. Oghenekevwe Etaga, R. Chibotu Ndubisi, and N. Lilian Oluebube, "Effect of Multicollinearity on Variable Selection in Multiple Regression," *Sci. J. Appl. Math. Stat.*, vol. 9, no. 6, p. 141, Dec. 2021, doi: 10.11648/j.sjams.20210906.12.

[33] T. Kyriazos and M. Poga, "Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions," *Open J. Stat.*, vol. 13, no. 03, pp. 404–424, Jun. 2023, doi: 10.4236/ojs.2023.133020.

[34] N. Shchurov, I. Isaev, S. Burikov, T. Dolenko, K. Laptinskiy, and S. Dolenko, "Taking into Account Mutual Correlations during Selection of Significant Input Features in Neural Network Solution of Inverse Problems of Spectroscopy," in *Proceedings of The 6th International Workshop on Deep Learning in Computational Physics — PoS(DLCP2022)*, Nov. 2022, vol. 429, p. 026, doi: 10.22323/1.429.0026.

[35] N. Z. Mahabob, Z. M. Yusoff, A. F. Mohd Amidon, N. Ismail, and M. N. Taib, "A novel application of artificial neural network for classifying agarwood essential oil quality," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 6, p. 6645, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6645-6652.

[36] A. H. Zaidi *et al.*, "Statistical analysis of agarwood oil chemical compound exists in four species of Aquilaria," *Int. J. Adv. Appl. Sci.*, vol. 13, no. 3, p. 727, Sep. 2024, doi: 10.11591/ijaas.v13.i3.pp727-732.

[37] N. A. S. Noramli, N. A. S. A. Sabri, M. I. Roslan, N. Ismail, Z. M. Yusoff, and M. N. Taib, "Analytical Characterization of Aquilaria Species Essential Oils Using Advanced Chromatographic Techniques Based on GC-MS and GC-FID," in *2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, Feb. 2025, no. 2025, pp. 10–15, doi: 10.1109/CSPA64953.2025.10933407.

[38] M. Mohamad *et al.*, "kNN: Classification of Agarwood Types in Oil and Wooden Using E-nose," in *Lecture Notes in Electrical Engineering*, vol. 842, Springer, Singapore, 2022, pp. 575–586, doi: 10.1007/978-981-16-8690-0_52.

[39] D. Yogapriya and M. Uma, "Agarwood Grade Estimation Procedure using Cnn and Sculpture Automation," in *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Jul. 2023, pp. 1185–1191, doi: 10.1109/ICECAA58104.2023.10212367.

[40] S. Lotthar *et al.*, "Decision Tree on Branch Node Observations for Significant Compounds of Agarwood Oil Different Grades," *J. Tomogr. Syst. Sens. Appl.*, vol. 6, no. 1, pp. 43–51, Jun. 2023. [Online]. Available at: http://tssa.my/index.php/jtssa/article/view/207.

[41] F. Haqqi, M. Rivai, and T. Mujiono, "Agarwood Classification using the Electronic Nose Method," in *2024 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Jul. 2024, no. 2024, pp. 396–400, doi: 10.1109/ISITIA63062.2024.10667957.

[42] Y. Pang *et al.*, "Solid-Phase Microextraction/Gas Chromatography–Time-of-Flight Mass Spectrometry Approach Combined with Network Pharmacology Analysis to Evaluate the Quality of Agarwood from Different Regions against Anxiety Disorder," *Molecules*, vol. 29, no. 2, p. 468, Jan. 2024, doi: 10.3390/MOLECULES29020468/S1.

[43] R. Hayati, A. A. Munawar, E. Lukitaningsih, N. Earlia, T. Karma, and R. Idroes, "Combination of PCA with LDA and SVM classifiers: A model for determining the geographical origin of coconut in the coastal plantation, Aceh Province, Indonesia," *Case Stud. Chem. Environ. Eng.*, vol. 9, p. 100552, Jun. 2024, doi: 10.1016/j.cscee.2023.100552.

[44] N. Zawani Mahabob, Z. M. Yusoff, A. F. Mohd Amidon, N. Ismail, and M. Nasir Taib, "Modeling of agarwood oil compounds based on linear regression and ANN for oil quality classification," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, p. 5505, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5505-5514.

[45] N. Zawani Mahabob, A. F. Mohd Amidon, Z. Mohd Yusoff, N. Ismail, and S. N. Tajjudin, "Comparison of ANN Performance Towards Agarwood Oil Compounds Pre-processing Based on Principal Component Analysis (PCA) and Stepwise Regression Selection Method," *J. Electr. Electron. Syst. Res.*, vol. 19, no. OCT2021, pp. 51–55, Oct. 2021, doi: 10.24191/jeesr.v19i1.007.

[46] T. S. Hong, F. Hanim Hashim, T. Raj, and A. B. Huddin, "Classification of Oil Palm Fruit Ripeness Using Artificial Neural Network," in *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, Jun. 2021, pp. 358–363, doi: 10.1109/I2CACIS52118.2021.9495857.

[47] S. M. H. Mohd Huzir, S. N. Tajuddin, Z. Mohd Yusoff, N. Ismail, A. A. Almisreb, and M. N. Taib, "Agarwood oil quality identification using artificial neural network modelling for five grades," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 2, p. 2254, Apr. 2024, doi: 10.11591/ijece.v14i2.pp2254-2261.

[48] Aqib Fawwaz Mohd Amidon, Zakiah Mohd Yusoff, Nurlaila Ismail, and Mohd Nasir Taib, "KNN Euclidean Distance Model Performance on Aquilaria Malaccensis Oil Qualities," *J. Adv. Res. Appl. Sci. Eng. Technol.*, vol. 48, no. 2, pp. 16–28, Jul. 2024, doi: 10.37934/araset.48.2.1628.

[49] Ö. Şimşek, "Machine Learning Offers Insights into the Impact of In Vitro Drought Stress on Strawberry Cultivars," *Agriculture*, vol. 14, no. 2, p. 294, Feb. 2024, doi: 10.3390/agriculture14020294.

[50] Z.-Q. Xie, J.-Y. Xu, M. Rafiq, and C.-S. Cheng, "An analysis of Agarwood trade patterns, historical perspectives, and species identification challenges: repercussions for importing nations," *TMR Mod. Herb. Med.*, vol. 7, no. 1, p. 1, 2024, doi: 10.53388/MHM2024001.

[51] Y. Z. H.-Y. Hashim *et al.*, "A Bibliometric Analysis of Agarwood Research, 1959-2021," *J. Agrobiotechnology*, vol. 13, no. 2, pp. 37–64, Nov. 2022, doi: 10.37231/jab.2022.13.2.296.