Jurnal Teknologi

Full Paper

INTEGRATING COUPLED GC–MS/GC–FID AND DATA ANALYTICS FOR PREDICTIVE AUTHENTICATION OF *AQUILARIA* SPECIES

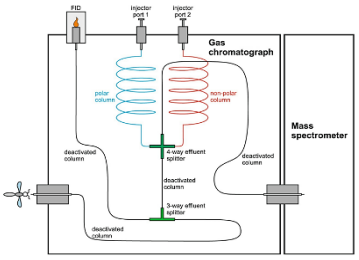
Nur Athirah Syafiqah Noramli, Noor Aida Syakira Ahmad Sabri, Muhammad Ikhsan Roslan, Nurlaila Ismail*, Zakiah Mohd Yusoff, Mohd Nasir Taib

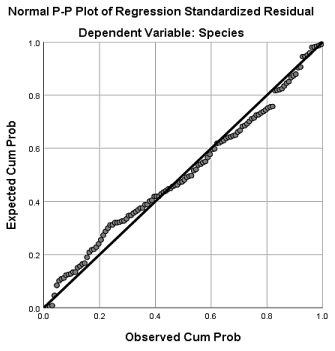
Advanced Signal Processing Research Interest Group, Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

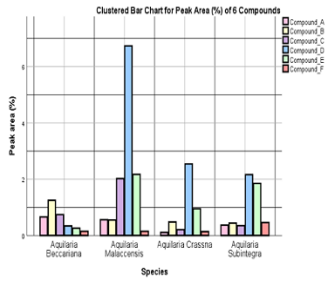
Article history
Received
29th November 2025
Received in revised form
16th December 2025
Accepted
11th February 2026
Published
1st November 2026

*Corresponding author
nurlaila0583@uitm.edu.my

Graphical abstract







Abstract

Agarwood (oud) is a highly valuable resinous wood derived from *Aquilaria* species, widely used in perfumery, incense, and traditional medicine. Traditional sensory-based classification methods are often subjective and susceptible to fraud and misidentification. This study applies gas chromatography–mass spectrometry (GC-MS) coupled with gas chromatography–flame ionization detection (GC-FID) integrated with statistical modelling to establish a predictive framework for authenticating *Aquilaria* species. Stepwise regression identified three chemical markers (γ-eudesmol, δ-guaiene, and 10-epi-γ-eudesmol) as significant differentiators among species. The optimized model achieved high predictive accuracy (R^2 = *0.990*), demonstrating the reliability of a chemometric approach for species classification. The proposed framework provides a scientific basis for standardizing agarwood authentication, reducing market adulteration, and supporting regulatory enforcement and conservation of endangered *Aquilaria* resources. Future work will expand chemical datasets and incorporate machine-learning algorithms to further enhance predictive precision and industrial scalability.

***Keywords*:** *Aquilaria* species, agarwood authentication, GC-MS coupled with GC-FID, stepwise regression, chemical compounds

Abstrak

Gaharu atau oud ialah kayu berresin bernilai tinggi daripada spesies *Aquilaria* yang digunakan dalam industri minyak wangi, kemenyan dan perubatan tradisional. Kaedah penilaian tradisional berasaskan deria bersifat subjektif serta mudah terdedah kepada ralat dan penipuan pasaran. Kajian ini menggunakan teknik GC-MS dan GC-FID yang digabungkan dengan analisis statistik bagi membangunkan model ramalan untuk pengesahan spesies *Aquilaria*. Analisis regresi berperingkat mengenal pasti tiga sebatian utama γ-eudesmol, δ-guaiena dan 10-epi-γ-eudesmol sebagai penanda penting dalam pembezaan spesies. Model ini mencapai ketepatan tinggi (R^2 = 0.990), membuktikan keberkesanan pendekatan kemometrik dalam klasifikasi gaharu. Pendekatan ini menyediakan asas saintifik bagi penyeragaman autentikasi, mengurangkan penipuan pasaran serta menyokong pemuliharaan sumber *Aquilaria* yang terancam.

***Kata kunci*:** Spesies *Aquilaria*, pengesahan gaharu, GC-MS bersama GC-FID, regresi berperingkat, sebatian kimia.

©2026 Penerbit UTM Press. All rights reserved

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

1.0 INTRODUCTION

Agarwood, also known as oud, is a highly valuable resinous wood derived from *Aquilaria* and *Gyrinops* species, primarily found in Southeast Asia, South Asia, and parts of China [1]. The resin is produced when the tree undergoes natural or artificial stress, such as fungal infection, insect attacks, or physical injury [2]. Over time, the resin saturates the heartwood, creating agarwood's distinct aroma, making it one of the most expensive natural products in the world [3], [4]. Due to its high demand and limited supply, illegal harvesting of agarwood has led to several species being classified as vulnerable or endangered under the Convention on International Trade in Endangered Species (CITES) [5].

The chemical composition of agarwood plays a crucial role in determining its fragrance and market value. It contains a complex mixture of sesquiterpenes, chromones, fatty acids, and aromatic compounds, which vary based on species and growth conditions [3]-[6]. Among these, sesquiterpenes and oxygenated derivatives are the primary contributors to its unique scent [1], [3], [4]. Traditionally, agarwood classification has relied on sensory evaluation based on colour, texture, and fragrance. However, such subjective approaches are prone to misclassification, adulteration, and fraud [1].

Advanced analytical techniques such as Gas Chromatography–Mass Spectrometry (GC-MS) and Gas Chromatography–Flame Ionization Detection (GC-FID) have been extensively applied to characterize the chemical constituents of agarwood oils and assist in authentication [1], [3], [4], [6]-[8]. Nevertheless, most prior studies emphasize qualitative profiling or exploratory multivariate analyses such as Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA), which primarily facilitate visualization or cluster separation rather than quantitative prediction [9], [10]. While these chemometric methods provide insight into general compositional trends, they do not establish direct quantitative relationships between specific compounds and species differentiation.

In this context, the present study employs stepwise regression analysis to address this methodological limitation. Stepwise regression offers a statistically rigorous means of identifying the most influential chemical variables by systematically including or excluding predictors based on significance thresholds [11]-[14]. Unlike PCA or PLS-DA, which compress data into latent dimensions and may obscure the effects of individual variables, stepwise regression retains interpretability by explicitly quantifying the contribution of each compound to species classification. This approach enables the development of a predictive model capable of identifying the significant chemical markers that underpin species differentiation.

Furthermore, this study is aligned with recent advances in data-driven chemical authentication, where analytical chemistry is increasingly combined with chemometric and machine-learning techniques to enhance prediction accuracy, automation, and objectivity [15]. Recent studies in natural product and essential oil authentication have demonstrated the potential of integrating GC-MS datasets with computational models such as artificial neural networks (ANN), support vector machines (SVM), and K-nearest neighbors (KNN) to achieve high classification accuracy [15]-[20]. Building on this trend, the integration of GC-MS and GC-FID profiling with stepwise regression in this study represents a robust, interpretable, and statistically validated framework for *Aquilaria* species authentication. This integration not only improves analytical precision but also contributes to reducing fraudulent trade and promoting the sustainable management and conservation of endangered *Aquilaria* species.

2.0 METHODOLOGY

2.1 Sample Collection and Experimental Setup

A total of 720 agarwood oil samples were analysed to identify the significant chemical compounds differentiating *Aquilaria* species. The study focused on four species: *A. beccariana* (AB), *A. malaccensis* (AM), *A. crassna* (AC), and *A. subintegra* (AS).

Six representative compounds β-selinene (A), dihydro-β-agarofuran (B), δ-guaiene (C), 10-epi-γ-eudesmol (D), γ-eudesmol (E), and pentadecanoic acid (F) were selected for quantitative analysis. The selection was based on three criteria [1], [20]- [23] :

1. Consistency: these compounds were detected reliably across all four sampled species, ensuring comparability.
2. Chemical and practical importance: they are widely recognized contributors to agarwood aroma, oil quality, and pharmacological activity.
3. Chemotaxonomic relevance: prior studies identified these sesquiterpenes as diagnostic markers for differentiating *Aquilaria* species.

GC-MS and GC-FID were used for chemical analysis [20], [24]. GC-MS provided qualitative identification of volatile compounds, while GC-FID quantified their relative concentrations [1], [4], [22], [24]. Each sample was analyzed in triplicate to ensure reproducibility and minimize experimental variation. The peak area percentages of the six compounds were recorded and used as input for statistical modelling.

The dataset, as summarized in Table 1, shows distinct variations in compound concentrations among the four *Aquilaria* species. β-Selinene (A) is present in all species but in relatively low amounts, with the highest concentration in *A. beccariana* (0.66%) and the lowest in *A. crassna* (0.11%). Dihydro-β-agarofuran (B)

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

has its highest abundance *in A. beccariana* (1.25%), while the other species have similar levels, ranging from 0.44% to 0.55%. δ-Guaiene (C) shows a significant concentration in A. malaccensis (2.02%), whereas the other species have much lower amounts, between 0.21% and 0.74%.

Table 1 Concentration of chemical compounds across *Aquilaria* species (%)

Code	Chemical Compounds	Indent, Mode	Peak Area (%)			
			AB	AM	AC	AS
A	β-selinene	FID, MS	0.66	0.56	0.11	0.37
B	dihyro-β-agarofuran	FID, MS	1.25	0.55	0.48	0.44
C	δ-guaiene	FID, MS	0.74	2.02	0.21	0.35
D	10-epi-γ-eudesmol	FID, MS	0.34	6.73	2.54	2.16
E	γ-eudesmol	FID, MS	0.26	2.17	0.95	1.85
F	Pentadecanoic acid	FID, MS	0.15	0.15	0.14	0.46

Among the detected compounds, 10-epi-γ-eudesmol (D) is notably abundant in *A. malaccensis*, with a peak area of 6.73%, which is significantly higher than its presence in the other species (0.34%–2.54%). γ-Eudesmol (E) follows a similar trend, showing the highest concentration in *A. malaccensis* (2.17%) and the lowest in *A. beccariana* (0.26%). Pentadecanoic acid (F) is present in all four species but in relatively low amounts, with its highest concentration detected in *A. subintegra* (0.46%) and the lowest in *A. crassna* (0.14%). These differences in compound distribution suggest species-specific variations in essential oil composition, which genetic, environmental, or physiological factors could influence.

To identify the most significant compounds contributing to species differentiation, stepwise regression analysis was conducted using SPSS software. This technique systematically selects or eliminates predictor variables based on their statistical significance, ensuring that only the most relevant compounds are considered in the model [11]-[14]. By applying this technique, the study aims to determine which essential oil components play a crucial role in distinguishing different *Aquilaria* species. This information is valuable for quality control, authentication, and commercial applications of agarwood-derived essential oils.

2.2 Agarwood oil extraction and analysis

The extraction of agarwood essential oil was performed at the BioAromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) using a hydro-distillation technique. The distillation process was optimized to ensure the gradual release of volatile compounds from the wood matrix. The extracted oil was subsequently diluted with analytical-grade dichloromethane (DCM) to enhance the solubility of the volatile components and improve detection accuracy during chromatographic analysis [22]

Identification and quantification of the chemical constituents were carried out using GC-MS coupled with GC-FID [8]. The GC-MS analysis was performed using an Agilent 7890B GC System coupled with an Agilent 5977A Mass Spectrometer Detector (MSD), as illustrated in Figure 1. The instrument was equipped with a DB-1ms column (30 m × 250 µm × 0.25 µm) and operated with helium as the carrier gas at a flow rate of 1.0 mL/min.

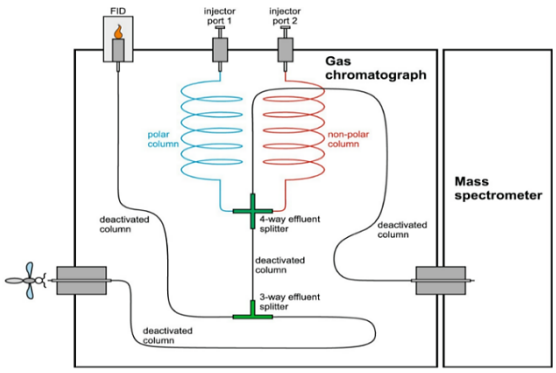


Figure 1 GC-MS Coupled with GC-FID System [24]

The oven temperature was initially set at 80°C, then increased at a rate of 3°C per minute to reach 250°C, where it was held for three minutes to allow complete compound elution. The mass spectrometer operated in Electron Impact (EI) mode at 70 eV, a standard ionization technique that produces reproducible fragmentation patterns for compound identification. Compound identities were verified by matching mass spectra against the National Institute of Standards and Technology (NIST) library, applying a minimum similarity threshold of 80% for reliable identification.

Table 2 Summary of GC-MS and GC-FID analytical parameters

Instrument	Column	Detector / Mode	Main Operating Conditions	Purpose
GC-MS	DB-1ms (30 m × 250 µm × 0.25 µm)	EI (70 eV)	Helium carrier gas (1.0 mL/min), 80–250°C at 3°C/min	Qualitative identification
GC-FID	DB-1ms (same)	FID (250°C)	Helium carrier gas (1.0 mL/min), 80–250°C at 3°C/min	Quantitative analysis

In addition to GC-MS, GC-FID analysis was conducted using the same chromatographic system, equipped with a Flame Ionization Detector (FID) operating at 250°C. Unlike GC-MS, which provides qualitative

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

identification based on mass spectra, GC-FID quantitatively determines compound concentrations through ionization of organic molecules [22], [24]. The peak areas (%) of the detected compounds were recorded and used to determine their relative abundances in different *Aquilaria* species. The detailed operating conditions of both instruments are summarized in Table 2.

2.3 Data Pre-Processing

Prior to statistical analysis, the dataset underwent a data pre-processing stage to ensure quality, consistency, and reliability. The dataset comprised peak area percentages obtained from GC-FID analysis, representing the relative abundances of six essential oil compounds across four *Aquilaria* species.

To enhance data reliability, outlier detection and removal were conducted to eliminate extreme values that could distort statistical outcomes [25], [26]. Following this, data normalization was performed using the min–max scaling method to standardize the range of peak area values while preserving their original distribution. This procedure minimized variability arising from differences in sample preparation or instrumental conditions, ensuring comparability between samples and species [25].

To prevent redundancy among correlated predictors, multicollinearity diagnostics were carried out using the Variance Inflation Factor (VIF) approach [27]-[33]. Compounds exhibiting excessively high VIF values were considered for exclusion or transformation to maintain the statistical independence of variables and improve model accuracy.

Through these pre-processing procedures, the dataset was refined and optimized for stepwise regression analysis, providing a robust foundation for identifying the most influential chemical compounds contributing to *Aquilaria* species differentiation.

2.4 Stepwise Regression Analysis

Stepwise regression analysis was employed to identify the most significant essential oil compounds contributing to the differentiation of *Aquilaria* species. This technique refines predictive models by systematically including or excluding variables based on their statistical significance [11], [12], [34]. In this study, stepwise regression was applied to the peak area percentages of six chemical compounds (β-selinene, dihydro-β-agarofuran, δ-guaiene, 10-epi-γ-eudesmol, γ-eudesmol, and pentadecanoic acid) across four *Aquilaria* species. The analysis determined which compounds exhibited the strongest influence on species classification.

The regression process integrated both forward selection and backward elimination procedures. Variables were included in the model when their significance level satisfied $p \leq 0.05$ and excluded when $p \geq 0.10$, ensuring that only statistically relevant predictors were retained [13], [14], [34], [35]. The adjusted R^2 and standard error of estimate were monitored during each iteration to optimize model performance and avoid overfitting.

All analyses were conducted using IBM SPSS Statistics, following a standardized workflow (Figure 2). The dataset was first structured with the dependent variable (Y) representing *Aquilaria* species classification and independent variables (X_1–X_6) representing the six identified chemical compounds. The algorithm iteratively evaluated the contribution of each variable to model fit, retaining only those with significant predictive influence.

.

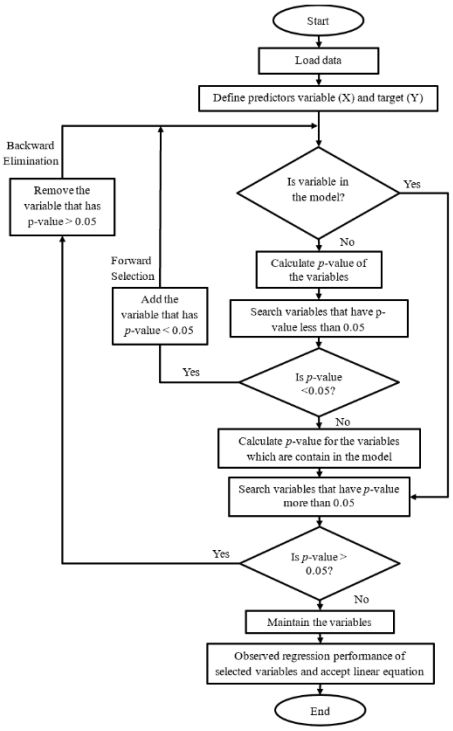


Figure 2 Stepwise regression analysis process

Upon completion, the final model generated a linear regression equation describing the relationship between the selected essential oil compounds and *Aquilaria* species differentiation. Model validity was confirmed through residual analysis and replicate data evaluation, ensuring compliance with regression assumptions and overall robustness. This approach provided a statistically sound framework for distinguishing *Aquilaria* species and enhanced interpretability of the underlying chemical patterns.

3.0 RESULTS AND DISCUSSION

3.1 Results

The stepwise regression analysis successfully identified three significant chemical compounds (γ-eudesmol (Compound E), δ-guaiene (Compound C), and 10-

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

epi-γ-eudesmol (Compound D)) as statistically significant differentiators among the four *Aquilaria* species. As shown in Table 3, the stepwise selection process began with Compound E, which had the highest statistical significance, followed by Compound C and Compound D. This method ensured that only the most relevant predictors were retained while eliminating less influential variables. The variable inclusion criteria were defined by a probability-of-F-to-enter threshold of ≤ 0.050 and a probability-of-F-to-remove threshold of ≥ 0.100, thereby maintaining statistical precision and minimizing model complexity.

Table 3 Regression variable selection

Variables Entered/Removed[a]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	Compound_E	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Compound_C	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Compound_D	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Species

The model summary in Table 4 illustrates the improvement in predictive power as more compounds were introduced. Initially, when only Compound E was included, the R^2 value was 0.267, meaning the model explained 26.7% of the variation in species classification. While this indicates a moderate correlation, it also suggests that additional variables were needed to enhance predictive accuracy. Once Compound C was added, the R^2 value increased significantly to 0.988, suggesting that this compound plays a major role in species differentiation. This sharp increase highlights δ-guaiene's strong influence in distinguishing the species when combined with γ-eudesmol.

Table 4 Model summary

Model Summary[d]				
Model	**R**	**R^2**	**Adjusted R^2**	**Std. Error of the Estimate**
1	0.516[a]	0.267	0.261	0.965
2	0.994[b]	0.988	0.987	0.126
3	0.995[c]	0.990	0.990	0.112

a. Predictors: (Constant), Compound_E

b. Predictors: (Constant), Compound_E, Compound_C

c. Predictors: (Constant), Compound_E, Compound_C, Compound_D

d. Dependent Variable: Species

The final step of the regression analysis involved adding Compound D, further improving the model's performance. With all three selected compounds included, the R^2 value reached 0.990, confirming that the model could now explain 99% of the variance in species classification. This near-perfect value indicates that these three compounds together provide a highly accurate basis for distinguishing between *Aquilaria* species. Additionally, the adjusted R^2 value remained nearly the same at 0.990, confirming that adding additional predictors was statistically justified and indicating that no model overfitting occurred.

The regression model was further validated using quantitative error metrics. The Root Mean Square Error (RMSE) was 0.0506, and the Mean Absolute Error (MAE) was 0.0395, both indicating minimal deviation between observed and predicted values. These low error magnitudes confirm the model's strong predictive performance and reliability.

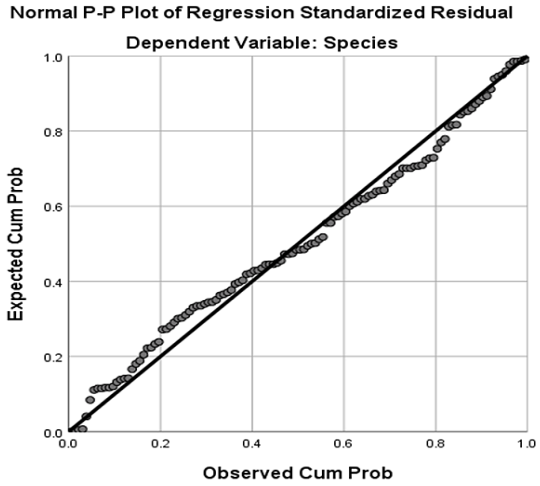


Figure 3 Normal probability plot (P-P plot) of residuals

Residual diagnostics were conducted to ensure that the regression assumptions were satisfied. The normal probability plot (P–P plot) of standardized residuals (Figure 3) shows that the observed cumulative probabilities closely align with the diagonal reference line, indicating that the residuals are normally distributed. This confirms that the model adheres to the assumptions of linearity, independence, and homoscedasticity. The close alignment of data points with the diagonal reference line indicates that the residuals are normally distributed, confirming the adequacy and reliability of the regression model.

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

To illustrate interspecies variation in chemical composition, Figure 4 presents the average peak area (%) of six major compounds detected across the four *Aquilaria* species. Distinct compositional differences were observed, with *A. malaccensis* exhibiting markedly higher concentrations of 10-epi-γ-eudesmol (Compound D) and γ-eudesmol (Compound E), whereas δ-guaiene (Compound C) showed moderate but consistent variation among the species. These three compounds correspond to the significant predictors identified in the regression model, underscoring their diagnostic importance for species classification and chemotaxonomic differentiation. The clustered bar chart displays β-selinene (A), dihydro-β-agarofuran (B), δ-guaiene (C), 10-epi-γ-eudesmol (D), γ-eudesmol (E), and pentadecanoic acid (F), indicating that the elevated levels of Compounds C, D, and E in *A. malaccensis* reflect species-specific chemical profiles consistent with regression-based classification.

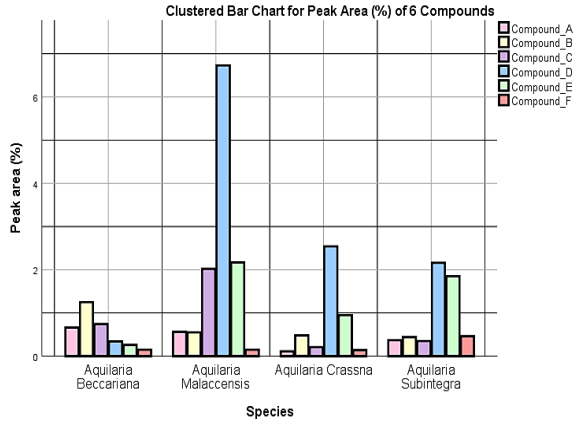


Figure 4 Compound distribution across *Aquilaria* species.

Overall, the results confirm that stepwise regression analysis is effective for identifying diagnostic chemical compounds in *Aquilaria* species classification. The high R^2 value (0.990), coupled with low RMSE (0.0506) and MAE (0.0395), demonstrates excellent model reliability and predictive precision. The significance of γ-eudesmol, δ-guaiene, and 10-epi-γ-eudesmol aligns with earlier reports [7], [36], reinforcing their role as chemotaxonomic markers. These findings provide a robust foundation for future applications in agarwood authentication, essential-oil quality assessment, and conservation-focused trade regulation.

3.2 Discussion

This study successfully demonstrated that stepwise regression is an effective chemometric approach for differentiating *Aquilaria* species based on their chemical composition. The analysis revealed γ-eudesmol, δ-guaiene, and 10-epi-γ-eudesmol as the most significant discriminating compounds, achieving an exceptionally high predictive accuracy (R^2 = 0.990; RMSE = 0.0506; MAE = 0.0395). These results confirm the robustness and reliability of the developed regression model. Consequently, the study establishes a quantitative and objective framework for agarwood authentication that overcomes the subjectivity of traditional sensory-based methods. Moreover, the identification of these sesquiterpenes aligns with previous reports recognizing them as diagnostic chemical markers and key contributors to agarwood's characteristic aroma [20], [22], [36]-[39]. This agreement reinforces the model's chemical validity and its potential for practical implementation in species authentication.

Building on these findings, it is essential to position stepwise regression within the broader context of existing analytical approaches. Conventional multivariate techniques such as PCA and PLS-DA have been widely applied to agarwood and other natural products [9], [10]. However, these methods primarily serve exploratory or visualization purposes and often lack optimized variable selection for prediction. In contrast, stepwise regression refines variable inclusion based on statistical significance, producing a more interpretable and targeted model [20], [40]. Nevertheless, recent advances in machine learning particularly ANN and SVM have achieved high accuracy in essential oil classification [16]-[18]. Therefore, future comparative studies should benchmark the performance of regression-based models against nonlinear classifiers such as ANN, PCA, or SVM using cross-validation to identify the most efficient predictive framework for *Aquilaria* species classification.

Despite the model's strong performance, its applicability under field conditions may be affected by external factors. Environmental variables including geographic origin, soil composition, and fungal infection, are known to influence agarwood chemistry [2], [39], [41]. Thus, expanding the dataset to incorporate multi-regional sampling and varying ecological conditions would enhance model generalizability. Additionally, hierarchical or mixed-effects regression frameworks could capture nested environmental variability, improving robustness against site-specific differences. Complementarily, ensemble-based methods such as random forests or gradient-boosted trees could model complex nonlinear dependencies while maintaining interpretability through variable-importance metrics.

Beyond model enhancement, integrating complementary analytical techniques presents a promising step toward scalable and automated agarwood authentication. Combining chromatographic techniques (GC–MS, GC–FID) with spectroscopic approaches such as Fourier Transform Infrared (FTIR) or Nuclear Magnetic Resonance (NMR) spectroscopy can improve both fingerprinting precision and analytical throughput [9]. For example, FTIR and Raman spectroscopy coupled with PCA and Hierarchical Cluster Analysis (HCA) have successfully distinguished authentic and adulterated *Rosa*

DOI: https://doi.org/10.11113/jurnalteknologi.v88.24590

damascena oils, while GC–MS and ATR-IR data effectively differentiated *Piper* species [42], [43]. Adapting these integrated frameworks for agarwood would reduce analytical time, increase throughput, and enable high-precision chemical screening, thus enhancing the efficiency of quality assurance and trade monitoring.

In summary, this study validates stepwise regression as a robust and interpretable statistical tool for *Aquilaria* species differentiation while highlighting the potential of hybrid analytical–computational frameworks for future development. Looking forward, integrating chemometric modelling with machine learning and spectroscopic data could yield adaptive, continuously improving predictive systems capable of handling large and diverse chemical datasets. Ultimately, such advancements will not only strengthen authentication reliability but also contribute to sustainable agarwood trade practices, standardized quality control, and the conservation of endangered *Aquilaria* species.

4.0 CONCLUSION

This study demonstrated the effectiveness of GC-MS coupled with GC-FID in identifying significant chemical compounds for differentiating *Aquilaria* species. Stepwise regression analysis successfully identified γ-eudesmol, δ-guaiene, and 10-epi-γ-eudesmol as the most influential predictors, achieving a high coefficient of determination ($R^2 = 0.990$), which confirms the model's strong predictive capability. These results establish a robust analytical framework for objective species authentication, thereby contributing to fraud prevention, quality assurance, and conservation of *Aquilaria* resources. From an industrial perspective, the proposed model can serve as a baseline for developing automated agarwood classification systems, supporting standardized quality control and sustainable trade practices. Its integration into commercial authentication workflows could streamline raw material verification in perfumery, pharmaceuticals, and traditional medicine industries. Future work should focus on expanding the dataset to include a wider range of geographical origins and environmental conditions to enhance model generalizability. Incorporating AI-driven predictive analytics such as deep learning and ensemble machine-learning algorithms could further refine classification accuracy and enable real-time, automated agarwood authentication. Overall, this research provides a scientific foundation for developing scalable, data-driven systems that advance the sustainable management and conservation of endangered *Aquilaria* species while strengthening the integrity of the global agarwood trade.

Acknowledgement

The authors gratefully acknowledge the financial support provided by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS), under grant number 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 at UMPSA for their expertise and significant contributions to the data extraction process, which played a pivotal role in the success of this work.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics declaration

Formal ethical approval was not obtained for this study. However, the authors confirm that the research was conducted in accordance with established ethical principles and good research practices, with due consideration given to participant rights, informed consent, confidentiality, and responsible data handling.

References

[1] D. Jia and S. Yi, "Classification of Hoi-An and Sin-Chew agarwood by components analysis of VOCs released in heat-treated agarwood using TD-GCMS and chemometric methods," *Bioresources*, Vol. 13, No. 2, pp. 2916–2931, 2018.
[2] 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: https://10.3390/molecules27113386.
[3] T. T. Kim Ngan *et al.*, "Chemical Components of Agarwood (Aquilaria crassna) Essential Oils Grown in Various Regions of Asia," *Asian Journal of Chemistry*, Vol. 32, No. 1, pp. 36–40, Nov. 2019, DOI: https:// 10.14233/ajchem.2020.22177.
[4] M.-R. Wang, W. Li, S. Luo, X. Zhao, C.-H. Ma, and S.-X. Liu, "GC-MS Study of the Chemical Components of Different Aquilaria sinensis (Lour.) Gilgorgans and Agarwood from Different Asian Countries," *Molecules*, Vol. 23, No. 9, p. 2168, Aug. 2018, DOI: https:// 10.3390/molecules23092168.
[5] 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 Modern Herbal Medicine*, Vol. 7, No. 1, pp. 1–10, 2024, DOI: https:// 10.53388/MHM2024001.
[6] M. N. Azah, N. Ismail, J. Mailina, M. N. Taib, M. Rahiman, and Z. M. Hafizi, "Chemometric Study of Selected Agarwood Oils by Gas Chromatography-Mass Spectrometry," *Journal of Tropical Forest Science*, Vol. 26, No. 3, pp. 382–388, 2014.
[7] Y. Zh-y Hashim, N. Ismail, and P. Abbas, "Analysis of Chemical Compounds of Agarwood Oil from Different Species by Gas Chromatography Mass Spectrometry (GCMS)," *Hashim et al. IIUM Engineering Journal*, Vol. 15, No. 1, pp. 1–6, 2014, DOI: https://doi.org/10.31436/iiumej.v15i1.469.
[8] H. E. Shaian, C. M. Aizal, and S. N. Tajuddin, "Chemical Composition of Agarwood Essential Oil (Aquilaria malaccensis) Upon Exposure Towards Heat Condition," *Malaysian Journal of Chemistry*, Vol. 23, No. 2, pp. 199–204, 2021.
[9] S. Ismail *et al.*, "Discriminative Analysis of Different Grades of Gaharu (Aquilaria malaccensis Lamk.) via 1H-NMR-Based Metabolomics Using PLS-DA and Random Forests Classification Models," *Molecules*, Vol. 22, No. 10, p. 1612, Sep. 2017, DOI: https:// 10.3390/molecules22101612.
[10] A. Porrello, A. Vaglica, D. Savoca, M. Bruno, and F. Sottile, "Variability in Crithmum maritimum L. Essential Oils' Chemical Composition: PCA Analysis, Food Safety, and Sustainability," *Sustainability (Switzerland)* , Vol. 16, No. 6, Mar. 2024, DOI: https://10.3390/su16062541.
[11] Z. Zhang, "Variable selection with stepwise and best subset approaches," *Ann Transl Med*, Vol. 4, No. 7, pp. 136–136, Apr. 2016, DOI: https://10.21037/atm.2016.03.35.
[12] M. G. Kim, "Geometric aspects of deletion diagnostics in multivariate regression," *J Appl Stat*, Vol. 42, No. 10, pp. 2073–2079, Oct. 2015, DOI: https://10.1080/02664763.2015.1016411.
[13] P. Pandit, P. Dey, and K. N. Krishnamurthy, "Comparative Assessment of Multiple Linear Regression and Fuzzy Linear Regression Models," *SN Comput Sci*, Vol. 2, No. 2, p. 76, Apr. 2021, DOI: https://10.1007/s42979-021-00473-3.
[14] A. E. Goodenough, A. G. Hart, and R. Stafford, "Regression with Empirical Variable Selection: Description of a New Method and Application to Ecological Datasets," *PLoS One*, Vol. 7, No. 3, pp. 1–10, Mar. 2012, DOI: https://10.1371/journal.pone.0034338.
[15] Aqib Fawwaz Mohd Amidon, Zakiah Mohd Yusoff, Nurlaila Ismail, and Mohd Nasir Taib, "KNN Euclidean Distance Model Performance on Aquilaria Malaccensis Oil Qualities," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, Vol. 48, No. 2, pp. 1–13, Jul. 2024, DOI: https://10.37934/araset.48.2.1628.
[16] N. Z. Mahabob, Z. M. Yusoff, A. F. M. Amidon, N. Ismail, and M. N. Taib, "Modeling of agarwood oil compounds based on linear regression and ANN for oil quality classification," *International Journal of Electrical and Computer Engineering*, Vol. 11, No. 6, pp. 5505–5514, Dec. 2021, DOI: https://10.11591/ijece.v11i6.pp5505-5514.
[17] 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," *Journal of Electrical & Electronic Systems Research*, Vol. 19, No. OCT2021, pp. 51–55, Oct. 2021, DOI: https://10.24191/jeesr.v19i1.007.
[18] M. A. Haqmi Abas, "Agarwood Oil Quality Classification using Support Vector Classifier and Grid Search Cross Validation Hyperparameter Tuning," *International Journal of Emerging Trends in Engineering Research*, Vol. 8, No. 6, pp. 2551–2556, Jun. 2020, DOI: https://10.30534/ijeter/2020/55862020.
[19] M. A. Aiman Ngadilan, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. Mohd Ali, and S. N. Tajuddin, "Radial basis function (RBF) tuned kernel parameter of agarwood oil compound for quality classification using support vector machine (SVM)," *2018 9th IEEE Control and System Graduate Research Colloquium, ICSGRC 2018 - Proceeding*, pp. 64–68, Jul. 2018, DOI: https://10.1109/ICSGRC.2018.8657524.
[20] 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," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 39, No. 1, p. 167, Jul. 2025, DOI: https://10.11591/ijeecs.v39.i1.pp167-177.
[21] 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)*, IEEE, Feb. 2025, pp. 10–15. DOI: https://10.1109/CSPA64953.2025.10933407.
[22] 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 Brief*, Vol. 53, p. 110209, Apr. 2024, DOI: https://10.1016/j.dib.2024.110209.
[23] N. Muhamad Sarih, N. Ismail, N. Akhtar, and S. N. Tajuddin, "Analysis of GC-FID and GC-MS Microwave-Assisted Hydrodistillation Extraction (MAHD) of Agarwood Chips," *The International Journal of Integrated Engineering*, Vol. 13, No. 6, Aug. 2021, DOI: https://10.30880/ijie.2021.13.06.017.
[24] 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, Dec. 2022, DOI: https://10.3389/fevo.2022.1042732.
[25] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Oriental journal of computer science and technology*, Vol. 13, No. 0203, pp. 78–81, Jan. 2021, DOI: https://10.13005/ojcst13.0203.03.
[26] K. Gibert, M. Sànchez–Marrè, and J. Izquierdo, "A survey on pre-processing techniques: Relevant issues in the context of environmental data mining," *AI Communications*, Vol. 29, No. 6, pp. 627–663, Dec. 2016, DOI: https://10.3233/AIC-160710.
[27] B. B. Xia, X. F. Zhao, and D. G. Feng, "Improve Steganalysis by MWM Feature Selection," in *Watermarking - Volume 2*, Vol. 2, InTech, 2012, pp. 243–258. DOI: https://10.5772/37988.
[28] K. P. Vatcheva and M. Lee, "Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies," *Epidemiology: Open Access*, Vol. 6, No. 2, pp. 1–9, 2016, DOI: https://10.4172/2161-1165.1000227.
[29] J. H. Kim, "Multicollinearity and misleading statistical results," *Korean J Anesthesiol*, Vol. 72, No. 6, pp. 558–569, Dec. 2019, DOI: https://10.4097/kja.19087.
[30] S. H. Huang, "Supervised feature selection: A tutorial," *Artif Intell Res*, Vol. 4, No. 2, pp. 22–37, Apr. 2015, DOI: https://10.5430/air.v4n2p22.
[31] H. J. Zainodin, G. Khuneswari, A. Noraini, and F. A. A. Haider, "Selected Model Systematic Sequence via Variance Inflationary Factor," *International Journal of Applied Physics and Mathematics*, Vol. 5, No. 2, pp. 105–114, 2015, DOI: https://10.17706/ijapm.2015.5.2.105-114.
[32] M. O. Akinwande, H. G. Dikko, and A. Samson, "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis," *Open J Stat*, Vol. 05, No. 07, pp. 754–767, 2015, DOI: https://10.4236/ojs.2015.57075.
[33] Y. Kang, P. Liu, F. Lv, Y. Zhang, Y. Yang, and J. Wei, "Genetic relationship and source species identification of 58 Qi-Nan germplasms of Aquilaria species in China that easily form agarwood," *PLoS One*, Vol. 17, No. 6, p. e0270167, Jun. 2022, DOI: https://10.1371/journal.pone.0270167.
[34] N. Roustaei, "Application and interpretation of linear-regression analysis," *Medical hypothesis discovery and innovation in ophthalmology*, Vol. 13, No. 3, pp. 151–159, Oct. 2024, DOI: https://10.51329/mehdiophthal1506.

[35] N. Shrestha, "Detecting Multicollinearity in Regression Analysis," *Am J Appl Math Stat*, Vol. 8, No. 2, pp. 39–42, Jun. 2020, DOI: https://10.12691/ajams-8-2-1.
[36] 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," *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 14, No. 2, p. 2254, Apr. 2024, DOI: https://10.11591/ijece.v14i2.pp2254-2261.
[37] A. Fawwaz Mohd Amidon, S. Mariatul Hazwa Mohd Huzir, Z. Mohd Yusoff, N. Ismail, and M. Nasir Taib, "The significance of artificial intelligent technique in classifying various grades of agarwood oil," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 29, No. 1, pp. 261–269, Jan. 2022, DOI: https://10.11591/ijeecs.v29.i1.pp261-269.
[38] A. H. 'Izzati H. Al-Hadi *et al.*, "Boxplot analysis of 4 grade agarwood essential oil for various grades," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 29, No. 1, pp. 238–244, Jan. 2023, DOI: https://10.11591/ijeecs.v29.i1.pp238-244.
[39] 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: https://10.1021/acs.jafc.4c06383.
[40] S. M. H. M. Huzir *et al.*, "Stepwise regression of agarwood oil significant chemical compounds into four quality differentiation," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 29, No. 2, p. 735, Feb. 2023, DOI: https://10.11591/ijeecs.v29.i2.pp735-741.
[41] S. R. Adhikari, K. Pokhrel, and S. D. Baral, "Economic Value of Agarwood and Its Prospects of Cultivation," *Int J Appl Sci Biotechnol*, Vol. 9, No. 1, pp. 23–31, Mar. 2021, DOI: https://10.3126/ijasbt.v9i1.35984.
[42] N. Cebi, M. Arici, and O. Sagdic, "The famous Turkish rose essential oil: Characterization and authenticity monitoring by FTIR, Raman and GC–MS techniques combined with chemometrics," *Food Chem*, Vol. 354, Aug. 2021, DOI: https://10.1016/j.foodchem.2021.129495.
[43] E. Al-Sayed, H. A. Gad, and D. M. El-Kersh, "Characterization of Four Piper Essential Oils (GC/MS and ATR-IR) Coupled to Chemometrics and Their anti- Helicobacter pylori Activity," *ACS Omega*, Vol. 6, No. 39, pp. 25652–25663, Oct. 2021, DOI: https://10.1021/acsomega.1c03777.