Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 1, July 2025, pp. 167~177
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i1.pp167-177

Unraveling the relationships among essential oil compounds in *Aquilaria* species using GC-MS and GC-FID techniques

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 Info

Article history:

Received Nov 14, 2024
Revised Mar 10, 2025
Accepted Mar 26, 2025

Keywords:

Agarwood
Aquilaria essential oil
Feature reduction
Multicollinearity
Pearson correlation analysis
Significant compounds

ABSTRACT

Agarwood, a prized non-timber resource from the *Aquilaria* genus, is highly valued for its aromatic and medicinal properties, playing a significant role in the healthcare, fragrance, and pharmaceutical industries. This research analyzes essential oils from four *Aquilaria* species-*A. beccariana, A. malaccensis, A. crassna, and A. subintegr*a-using gas chromatography-mass spectrometry (GC-MS) and gas chromatography-flame ionization detection (GC-FID). The primary objective is to optimize classification efficiency by reducing computational time and reducing multicollinearity through feature selection. Pearson correlation analysis revealed strong relationships among six chemical compounds-β-selinene (A), dihydro-β-agarofuran (B), δ-guaiene (C), 10-epi-γ-eudesmol (D), γ-eudesmol (E), and pentadecanoic acid (F). Through feature selection, the three most significant compounds-dihydro-β-agarofuran (B), γ-eudesmol (D), and 10-epi-γ-eudesmol (E)-were identified, achieving a remarkable 90.02% reduction in computational time (from 0.0403 to 0.0040 seconds). These findings highlight the effectiveness of structured feature selection in refining essential oil profiling and enhancing species classification accuracy. Future research directions include exploring machine learning-based feature selection techniques to further streamline feature reduction processes and expand the scope of essential oil authentication. This study contributes to advancing the scientific understanding and practical utilization of agarwood essential oils, paving the way for more efficient and reliable analytical frameworks.

Corresponding Author:

Nurlaila Ismail
Advanced Signal Processing Research Interest Group, Faculty of Electrical Engineering
Universiti Teknologi MARA
Shah Alam, Selangor, 40450, Malaysia
Email: nurlaila0583@uitm.edu.my

1. INTRODUCTION

Agarwood, also referred to as eaglewood or gaharu, is a valuable non-timber forest resource obtained from certain species of the *Aquilaria* genus. This fragrant wood resin is widely utilized in various sectors such as healthcare, religious ceremonies, and fragrance production because of its distinct characteristics [1]. The importance of *Aquilaria* species lies in their ability to produce agarwood, which contains both volatile and non-volatile compounds with significant medicinal applications [2].

The resin produced by *Aquilaria* species plays a crucial role in agarwood manufacturing, making it a highly sought-after commodity [3]. However, the identification and evaluation of *Aquilaria* species have

historically relied on subjective human assessments, which are often time-consuming and inconsistent [4]. This challenge underscores the need for more objective and reliable identification techniques.

Agarwood is valued not only for its aromatic qualities but also for its medicinal properties, making it a staple in traditional remedies [1], [5]. Due to its rarity and high economic value, verifying the authenticity of agarwood is essential to prevent counterfeiting and ensure product integrity. Techniques like DNA barcoding have been employed to authenticate agarwood and maintain its quality standards [6].

However, existing approaches often involve extensive computation or fail to optimize compound selection efficiently. The growing demand for this resinous wood has raised concerns about overharvesting, threatening the sustainability of *Aquilaria* species [7]. Therefore, accurate species identification is critical for quality control and conservation efforts.

This research aims to identify the most significant compounds in the essential oils of *Aquilaria* species by analyzing peak area (%) data from GC-MS coupled with GC-FID. The data, provided by the BARCE at UMPSA, includes samples from four *Aquilaria* species: *Aquilaria Beccariana, Aquilaria Malaccensis, Aquilaria Crassna, and Aquilaria Subintegra*. The analysis, conducted using SPSS software version 26, will apply feature selection to reduce the number of compounds, minimize time consumption, and enhance efficiency.

By systematically identifying the most influential essential oil components, this research seeks to refine essential oil classification and improve industrial applications such as agarwood authentication and quality control. Researchers have utilized various analytical techniques and statistical methods to select significant compounds of essential oils while excluding redundant ones [7], [8]. Malaysia and Indonesia are among the leading countries cultivating *Aquilaria* trees for agarwood production. These regions are recognized for their extensive agarwood trade, contributing significantly to the global market. Table 1 provides an overview of different *Aquilaria* species and their distribution displays the overview of different types of *Aquilaria* species and their global distribution [9]-[11].

Table 1. The overview of different types of *Aquilaria* species and their global distribution

Aquilaria species	Origin	Agarwood producing	Reference
Aquilaria microcarpa	Malaysia, Indonesia, Singapore	Yes	[9]
Aquilaria parvifolia	Philippines	No	[9]
Aquilaria rostrata	Malaysia	Yes	[9], [10]
Aquilaria rugosa	Vietnam	No	[9]
Aquilaria crassna	Cambodia, Malaysia, Thailand, Vietnam, Laos, India	Yes	[9], [10]
Aquilaria cumingiana	Borneo, Indonesia, Philippines	Yes	[9], [10]
Aquilaria decemcostata	Philippines	Yes	[9]
Aquilaria filaria	Papua New Guinea, Indonesia, Thailand, Philippines	Yes	[9], [10]
Aquilaria hirta	Malaysia, Indonesia, Thailand	Yes	[9], [10]
Aquilaria khasiana	Pakistan, India	No	[9], [10]
Aquilaria malaccensis	Malaysia, Indonesia, Thailand, Philippines, Singapore, Myanmar, India, Bangladesh, Bhutan, Laos	Yes	[9], [10]
Aquilaria sinensis	China	No	[9], [10]
Aquilaria subintegra	Malaysia, Thailand	Yes	[9], [10]
Aquilaria urdanetensis	Philippines	No	[9]
Aquilaria yunnanensis	China	No	[9]
Aquilaria grandiflora	China	Yes	[9]
Aquilaria ophispermum	Indonesia	No	[9]
Aquilaria pentandra	Bhutan, Laos, Thailand, Myanmar	No	[9]
Aquilaria acuminata	Papua New Guinea, Indonesia, Thailand, Philippines	Yes	[9], [10]
Aquilaria apiculata	Philippines	Yes	[9], [10]
Aquilaria baillonii	Cambodia, Thailand, Laos, Vietnam	No	[9], [10]
Aquilaria banaensis	Vietnam	No	[9], [10]
Aquilaria beccariana	Malaysia, Indonesia, Borneo	Yes	[9], [10]
Aquilaria brachyantha	Malaysia, Philippines	Yes	[9], [10]
Aquilaria citrinicarpa	Philippines	No	[9]

A key approach to analyzing essential oil composition involves the use of Pearson correlation matrices to identify the relationship between environmental conditions and the chemical profiles of essential oil [11]. This approach allows researchers to explore ecological preferences and variations in essential oil components, facilitating the identification of compounds that are most affected by specific environmental

factors. Through advanced statistical analysis, such as Pearson's canonical correlation analysis (CCA), researchers can determine which essential oil compounds exhibit the strongest associations with prevailing environmental conditions [7], [11].

A study analyzing essential oils from *Apiaceae* fruits utilized gas chromatographic-mass spectrometric analysis, a methodology that closely aligns with the approach adopted in this research [12]. By integrating analytical techniques with statistical evaluation, the study successfully determined oil compositions and identified significant compounds through correlation-based selection. These findings highlight the importance of combining chemical profiling with statistical analysis to refine the process of compound identification and enhance the accuracy of essential oil characterization.

Reducing the number of variables in an analysis can help lower computational costs and improve overall efficiency [13], [14]. Pearson's correlation coefficient will be used to measure the statistical relationship or association between two continuous variables [15]. This test provides valuable insights into the magnitude and direction of correlations, allowing researchers to pinpoint significant chemical compounds. These identified compounds will serve as benchmarks for classifying different *Aquilaria* species, ensuring a more structured and data-driven approach to essential oil characterization.

2. METHOD

The methodology employed in this study systematically analyzes the chemical composition of *Aquilaria* essential oil samples, focusing on optimizing feature selection to reduce processing time and address multicollinearity. The approach was designed to ensure accurate data collection and the precise identification of significant chemical compounds. By refining the analysis process, the study aims to enhance computational efficiency while maintaining precision in chemical profiling.

The discussion is divided into two main subsections: 2.1 sample data collection and 2.2 correlation analysis for feature selection. The first subsection describes the procedures for collecting and characterizing samples, ensuring consistency and reliability in data acquisition. The second subsection details the statistical techniques used to assess relationships among the chemical compounds, enabling the identification of significant features. By focusing on essential variables, the methodology minimizes redundancy, streamlines the analysis and significantly reduces computational time thereby improving the overall efficiency of the profiling process.

2.1. Sample data collection

The *Aquilaria* essential oil samples were created by the BARCE at UMPSA. The essential oil were extracted using GC-MS coupled with the GC-FID technique. This method allows for precise identification and quantification of essential oil components, ensuring reliable data for further analysis. A total of 720 samples, each containing six different chemical compounds, were analyzed. The compounds are β-selinene, dihydro-β-agarofuran, δ-guaiene, 10-epi-γ-eudesmol, γ-eudesmol and pentadecanoic acid, identified as acronyms A, B, C, D, E, and F. The dataset comprises essential oil samples from four *Aquilaria* species: *Aquilaria beccariana* (AB), *Aquilaria malaccensis* (AM), *Aquilaria crassna* (AC), and *Aquilaria subintegra* (AS). These species were selected due to their significance in agarwood production and their distinct chemical compositions.

The peak area (%) measurements obtained from the GC-MS and GC-FID analyses indicate the relative concentration levels of each compound in different *Aquilaria* species. This quantitative information was essential for categorizing significant chemical compounds and assessing their variability among species. By analyzing these variations, this research aimed to identify significant chemical compounds for distinguishing different *Aquilaria* species and optimizing feature selection.

2.2. Correlation analysis for feature selection

Feature selection is a critical step that involves reducing the number of input variables to enhance computational efficiency and minimize processing time [14]. Correlation analysis, specifically Pearson's correlation coefficient, is commonly used to evaluate the relationship between variables. A positive correlation indicates that two features increase or decrease together, while a negative correlation suggests an inverse relationship where one feature increases as the other decreases [15].

Table 2 presents the interpretation of correlation coefficients, which are used to measure the strength and direction of the relationship between two variables. The table provides a general interpretation of the size of the correlation coefficient and the corresponding strength of the relationship. According to Table 2, a correlation coefficient between 0.8 and 1.0 indicates a very strong relationship between the variables, while a coefficient between 0.6 and 0.79 suggests a strong relationship.

A moderate relationship is indicated by a coefficient between 0.4 and 0.59, and a weak relationship is represented by a coefficient between 0.2 and 0.39. Finally, a coefficient between 0.0 and 0.19 is interpreted

as a weak or no relationship between the variables [16]. These classifications provide a guideline for selecting the most relevant features while eliminating less significant ones.

Table 2. Interpreting a correlation coefficient [16]

Size of the correlation	Coefficient general interpretation
0.8 to 1.0	Very strong relationship
0.6 to 0.79	Strong relationship
0.4 to 0.59	Moderate relationship
0.2 to 0.39	Weak relationship
0.0 to 0.19	Weak or no relationship

Highly correlated independent features are considered redundant, leading to increased processing time and inefficient computations [17]. Therefore, it is crucial to identify and eliminate highly correlated features to streamline data analysis while retaining the most informative compounds [18]. Feature selection methods are designed to decrease computation time, enhance efficiency, and provide deeper insights into the data [19].

By employing feature selection techniques, researchers can pinpoint influential features, discard irrelevant or redundant ones, and establish a new feature set that improves computational efficiency [20]. According to the findings in [21], statistical analysis using a Pearson correlation matrix was successfully employed to determine the insecticidal properties of essential oils from Piper species. The study found that the insecticidal activity was primarily attributed to specific compounds, such as camphene, α-pinene, and β-pinene, highlighting the effectiveness of using correlation analysis to identify significant compounds. In this research, compounds with correlation coefficients above 0.8 were considered redundant and were removed to optimize computational resources while preserving essential chemical information. By applying this method, the dataset was refined to retain the most significant compounds while ensuring that computational efficiency was maximized without unnecessary complexity.

3. RESULTS AND DISCUSSION

The research findings highlight the chemical composition essential oils derived from four *Aquilaria* species, with a focus on identifying the most significant chemical compounds through Pearson correlation analysis. This approach enables effective feature selection, optimizing the analytical process, to efficiency and reduce computational time. The findings are illustrated using figures, tables, and visual tools such as the correlation matrix and Venn diagram, which highlight the relationships between different chemical compounds. These visual aids effectively illustrate the relationships among chemical compounds, providing clear insights into their interdependencies. Such representations simplify the identification of essential features while distinguishing them from redundant ones. Understanding these correlations is crucial for refining the dataset and improving computational efficiency.

The discussion is organized into two subsections: 3.1. overview of *Aquilaria* essential oil feature selection, which outlines the feature selection process, and 3.2. analysis of the correlation matrix and implications for feature selection, which examines the identified correlations and their role in eliminating redundancy features. By reducing redundancy in the dataset, this approach enhances computational efficiency while preserving the integrity of the chemical analysis.

3.1. Overview of *Aquilaria* essential oil feature selection

Correlation analysis is used to select features of *Aquilaria* essential oil from agarwood trees, specifically focusing on four species*: Aquilaria Beccariana, Aquilaria Malaccensis, Aquilaria Crassna*, and *Aquilaria Subintegra*. Table 3 displays that the dataset consists of 720 samples collected from these four species. Six key chemical compounds-β-selinene, dihydro-β-agarofuran, δ-guaiene, 10-epi-γ-eudesmol, γ-eudesmol, and pentadecanoic acid-denoted as A, B, C, D, E, and F, were selected for this research due to their consistent presence across all four *Aquilaria* species, and their prominent peak areas (%) in the gas chromatography results [22].

These compounds have been widely acknowledged in previous research as critical markers of agarwood quality, as they play a significant role in shaping the distinctive fragrance and chemical composition of *Aquilaria* essential oils [23]. Notably, compounds A, C, and E are frequently linked to the characteristic aroma and resinous content of agarwood oils, making them important compounds for distinguishing between species. Additionally, compounds B and D have been identified as compounds that

contribute to the biological activities of agarwood oils [24]. While pentadecanoic acid is less commonly encountered, it serves as a secondary marker for species-specific variations. Therefore, these six compounds were selected based on their reliable presence and significant peak areas in GC-MS coupled with GC-FID analysis across the species. These compounds provided the input for the classification system, whereas the chemical composition of the four *Aquilaria* species was determined as the output. SPSS version 26 was used to conduct the analysis.

Table 3. Datasets of chemical compounds and peak areas for each *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

3.2. Analysis of the correlation matrix and implications for feature selection

The correlation matrix presented in Figure 1 highlights notable relationships among significant compounds in the essential oils of *Aquilaria* species. The Pearson correlation coefficient quantifies the strength and direction of the linear relationship between two variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation. Pearson's correlation coefficient (r) can be calculated using the equation shown in (1) [25].

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \Sigma(y-\bar{y})^2}} \quad (1)$$

Where x and y are values of the variable, x represents the independent variable, and y represents the dependent variable; $\bar{x}$ and $\bar{y}$ are the mean values of the variable x and y. This formula enables researchers to evaluate relationships between different compounds and determine whether certain features provide redundant information. Highly correlated compounds can be removed to improve efficiency without sacrificing the integrity of the dataset.

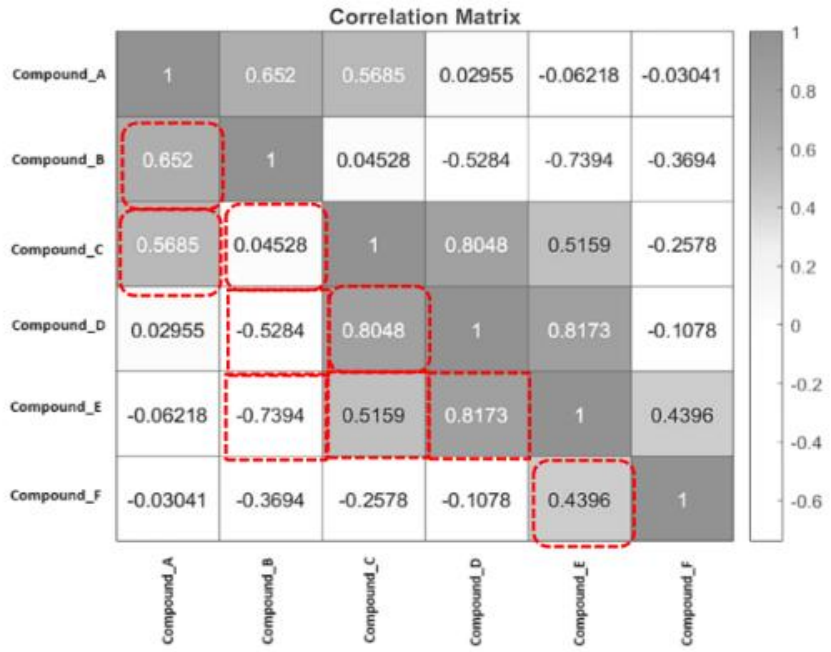

Figure 1. Correlation matrix for significant compounds in essential oils of *Aquilaria* species

A strong positive correlation (r=0.652) is observed between compound A and compound B, indicating that these compounds may capture similar characteristics. Similarly, compound A showed a moderate to strong correlation with compound C (r=0.5685), while compound B also correlates with compound C (r=0.4528). These findings highlight the necessity of feature selection methods to minimize redundancy and enhance computational efficiency.

Negative correlations are also notable in the dataset. Compound B exhibits a moderate negative correlation with compound D (r=-0.5284) and a strong negative correlation with compound E (r=-0.7394). These inverse relationships suggest potential biochemical interactions or mutual exclusivity in *Aquilaria* essential oil composition.

Additionally, compound C and compound D share a very strong positive correlation (r=0.8048), as did compound D and compound E (r=0.8173). Such high correlations indicate significant interrelationships among these compounds, suggesting possible redundancies. Retaining highly correlated features can increase computational time and complicate the analysis, making feature selection crucial for optimizing the dataset.

To address this, feature selection techniques were applied to remove highly correlated compounds, ensuring a streamlined dataset while maintaining essential chemical information. A similar approach was reported in [7], where correlation analysis linked the chemical composition of *Centaurea* essential oils with antimicrobial activity, demonstrating the utility of correlation analysis in reducing redundancy and enhancing efficiency.

Table 4 provides a summary of how well each compound can explains others in the dataset. Compound A can effectively explains compounds B and C, indicating that it captures key characteristics shared with these two compounds. Similarly, compounds B and C exhibit strong explanatory power, influencing compounds A, D, and E. This shows that these two compounds have a broad influence within the dataset and may represent critical features.

Table 4. Interpretation of Pearson's correlations

Compounds	Can explain (compounds)
A	B, C
B	A, D, E
C	A, D, E
D	B, C, E
E	B, C, D, F
F	E

Compound D also demonstrates strong explanatory power, as it can explain compounds B, C, and E well. This suggests that it plays a central role in connecting other features within the dataset. Compound E stands out as the most interconnected, as it can explain four other compounds-B, C, D, and F indicating that it carries significant information about the dataset's structure. Finally, compound F, while more limited, still explains compound E effectively, which highlights a specific but important relationship between these two compounds.

However, compounds B and C have similar explanatory power, particularly for compounds A, D, and E. To address the issue of multicollinearity, compound B has been chosen as the representative feature from this pair due to its broader explanatory significance and connections. This step reduces redundancy in the dataset and ensures a more efficient analysis. Thus, the most significant compounds for the Pearson correlation are B, D, and E. Using these three key compounds, it is possible to explain all six compounds that distinguish the *Aquilaria* essential oil.

A Venn diagram in Figure 2 visually represents the overlapping correlations among key compounds in *Aquilaria* essential oils. It highlights compounds B, E, and D as central compounds, effectively capturing shared chemical characteristics. Compound B, exhibiting both strong positive and negative correlations, serves as a useful proxy for understanding chemical relationships. Likewise, compound E's significant positive correlations suggest its importance in defining fragrance and biological properties. Compound D further reinforces its role as a key component due to its strong correlations with other compounds.

The overlaps among compounds B, E, and D emphasize their influence and representation in the dataset. These findings align with previous research [26], [27] which identified sesquiterpenes as major constituents in agarwood essential oil. Prioritizing these compounds in feature selection enhances efficiency, minimizes computational demands, and optimizes the chemical analysis of *Aquilaria* essential oils.

Feature selection plays a crucial role in optimizing computational efficiency. The computational time before feature selection was 0.0403 seconds, whereas after applying feature selection, it was significantly reduced to 0.0040 seconds. This demonstrates a substantial 90.02% improvement in computational efficiency.

The reduction in computation time indicates that feature selection successfully eliminates redundant or less relevant features, leading to a more efficient data processing pipeline. This improvement is particularly beneficial for large datasets, where unnecessary features can significantly slow down computations. The results highlight that by applying feature selection techniques, the model not only improves in terms of processing speed but also becomes more scalable and resource-efficient. These findings confirm the effectiveness of the applied feature selection method, reinforcing its importance in enhancing computational efficiency and optimizing chemical compound analysis tasks.

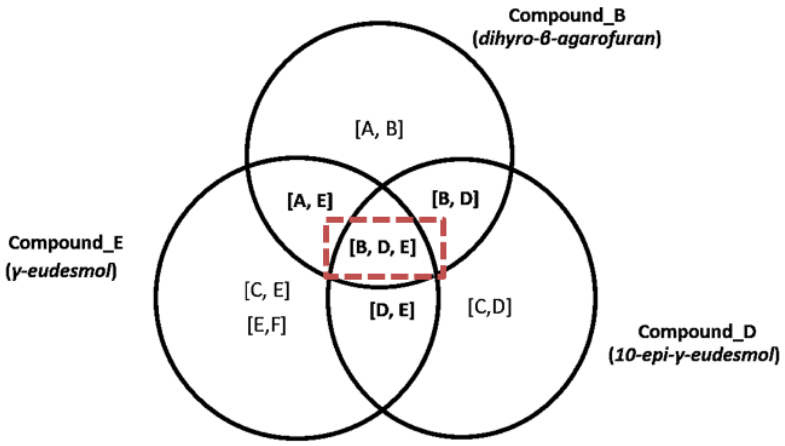

Figure 2. Venn diagram of the correlation matrix

3.3. Discussion

This study successfully identified and refined key chemical compounds in *Aquilaria* essential oils using GC-MS and GC-FID techniques. By employing Pearson correlation analysis, the six initially identified compounds-A, B, C, D, E, and F-were reduced to the three most significant chemical compounds: B, D, and E. This reduction enhances analytical efficiency and reduces computational demands by eliminating redundant features, thereby improving species classification accuracy, and optimizing essential oil profiling.

The impact of feature selection on computational efficiency was also evaluated. The computational time before feature selection was 0.0403 seconds, which was reduced to 0.0040 seconds after feature selection, achieving a 90.02% improvement in computational efficiency. This substantial reduction demonstrates the effectiveness of removing unnecessary features, as it not only reduces computational load but also streamlines data processing. These results confirm that structured feature selection methods can effectively enhance analysis efficiency, making data handling more scalable and resource-efficient.

The findings support the hypothesis that feature selection techniques improve analysis efficiency by addressing multicollinearity and eliminating redundant compounds. This confirms findings [7], [26], [27] that highlights sesquiterpenes (e.g., γ-eudesmol) as key contributors to agarwood's fragrance and medicinal properties. By prioritizing these compounds during feature selection, the study enhances efficiency, while minimizing computational demands. These results reinforce the importance of structured feature selection in essential oil analysis.

Beyond computational improvements, the identification of dihydro-*β*-agarofuran (B), *γ*-eudesmol (D), and 10-epi-*γ*-eudesmol (E) as significant chemical compounds highlights their potential significance in agarwood authentication. These compounds contribute to the fragrance, medicinal properties, and resin formation in *Aquilaria* species, which are crucial factors for industrial applications. By streamlining the dataset while maintaining essential chemical information, this research contributes to more effective essential oil characterization.

Despite these contributions, the research relies on traditional statistical methods for feature selection, which, while effective, may not fully optimize compound selection under complex conditions. Future research should integrate machine learning techniques, such as principal component analysis (PCA) or deep learning, to further enhance the selection of essential oil compounds. These advanced methods can detect more intricate relationships between chemical components, providing a scalable approach to essential oil profiling [28].

One limitation of this research is that the analysis was conducted on a dataset of 720 samples from four *Aquilaria* species. While this provides valuable insights, future research should expand the dataset to

include more species and environmental variations to ensure broader applicability. Additionally, future studies could explore how environmental factors influence compound variability, further improving the robustness of essential oil classification models. In summary, this research successfully refines the feature selection process, significantly improving computational efficiency in *Aquilaria* essential oil analysis. The integration of AI-driven classification methods in future research could further enhance the precision and reliability of essential oil profiling, benefiting industries such as perfumery, traditional medicine, and quality authentication.

4. CONCLUSION

The study employed advanced analytical techniques, including GC-MS coupled with GC-FID, to analyze the essential oils from four *Aquilaria* species: *Aquilaria beccariana, Aquilaria malaccensis, Aquilaria crassna,* and *Aquilaria subintegra*. By applying Pearson's correlation analysis and feature selection, the initial set of six chemical compounds—*β*-selinene (A), dihydro-*β*-agarofuran (B), *δ*-guaiene (C), 10-epi-*γ*-eudesmol (D), *γ*-eudesmol (E), and pentadecanoic acid (F)—was refined to the three most significant compounds: dihydro-*β*-agarofuran (B), *γ*-eudesmol (D), and 10-epi-*γ*-eudesmol (E). These compounds are critical to the fragrance and medicinal properties of agarwood, making them valuable markers for quality assessment and species classification. The reduction of redundant features resulted in a remarkable 90.02% improvement in computational efficiency, with processing time decreasing from 0.0403 seconds to 0.0040 seconds. This enhancement is particularly impactful for industrial applications, such as essential oil authentication, perfumery, and pharmaceutical standardization, where rapid and precise analysis is essential for quality control. Although this research successfully refines feature selection, it is limited to a dataset comprising four *Aquilaria* species. Future research should expand the dataset to include more species and geographical variations to ensure broader applicability. Additionally, while Pearson correlation effectively identifies linear relationships, advanced machine learning techniques such as PCA and deep learning could uncover more complex patterns and nonlinear interactions within the data. In summary, this research establishes a refined approach to *Aquilaria* essential oil classification, demonstrating that strategic feature selection can significantly enhance analytical efficiency. Future advancements in AI-driven methods will further revolutionize essential oil authentication, paving the way for more accurate and scalable industry applications.

ACKNOWLEDGMENTS

The authors would like to express their deep appreciation to the Journal Support Fund (JSF) of Universiti Teknologi MARA and the Institute of Postgraduate Studies, UiTM for their financial support. The authors also extend their gratitude to the members of the Advanced Signal Processing Research Interest Group for their invaluable feedback and support. Additionally, the authors are grateful for the significant assistance and expertise provided by the BioAromatic Research Centre of Excellence (BARCE) at Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) in the crucial task of data extraction.

FUNDING INFORMATION

The authors would like to express their deep appreciation to the Journal Support Fund (JSF) of Universiti Teknologi MARA and the Institute of Postgraduate Studies, UiTM for their financial support.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nur Athirah Syafiqah Noramli	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Noor Aida Syakira Ahmad Sabri	✓	✓	✓			✓		✓		✓	✓			
Muhammad Ikhsan Roslan	✓	✓	✓	✓			✓			✓				
Nurlaila Ismail	✓			✓						✓	✓	✓		✓
Zakiah Mohd Yussoff	✓			✓			✓	✓		✓				
Mohd Nasir Taib	✓	✓		✓						✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] T. Yan, S. Yang, Y. Chen, Q. Wang, and G. Li, "Chemical profiles of cultivated agarwood induced by different techniques," *Molecules*, vol. 24, no. 10, p. 1990, May 2019, doi: 10.3390/molecules24101990.

[2] T. Zhang, T. Zhu, P. Xiong, H. Huo, Z. Tari, and W. Zhou, "Correlated differential privacy: feature selection in machine learning," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2115–2124, Mar. 2020, doi: 10.1109/TII.2019.2936825.

[3] S. Y. Lee, W. L. Ng, M. N. Mahat, M. Nazre, and R. Mohamed, "DNA barcoding of the endangered *Aquilaria* (Thymelaeaceae) and its application in species authentication of Agarwood Products traded in the market," *PLoS ONE*, vol. 11, no. 4, p. e0154631, Apr. 2016, doi: 10.1371/journal.pone.0154631.

[4] F. Chen *et al.*, "Comprehensive comparisons between grafted kynam agarwood and normal agarwood on traits, composition, and in vitro activation of AMPK," *Molecules*, vol. 28, no. 4, p. 1667, Feb. 2023, doi: 10.3390/molecules28041667.

[5] 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.

[6] Y. Sano and M. Inoue, "Selective felling as incidental less-destructive harvesting: agarwood collection by Punan Aput in North Kalimantan, Indonesia," *Tropics*, vol. 33, no. 1, pp. 33–55, Jun. 2024, doi: 10.3759/tropics.MS22-01.

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

[8] S. Ezekiel, A. A. Alshehri, L. Pearlstein, X.-W. Wu, and A. Lutz, "IoT anomaly detection using multivariate," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 4, pp. 1662–1669, Feb. 2020, doi: 10.35940/ijitee.d1323.029420.

[9] M. F. Elias, H. Ibrahim, and W. R. W. Mahamod, "A review on the Malaysian *Aquilaria* species in Karas Plantation and agarwood production," *International Journal of Academic Research in Business and Social Sciences*, vol. 7, no. 4, pp. 1–9, May 2017, doi: 10.6007/ijarbss/v7-i4/2911.

[10] 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.

[11] O. S. Tsiftsoglou, R. Stagiopoulou, N. Krigas, and D. Lazari, "Exploring the ecological preferences and essential oil variability in wild-growing populations of the endangered local greek endemic thymus holosericeus (lamiaceae)," *Plants*, vol. 12, no. 2, p. 348, Jan. 2023, doi: 10.3390/plants12020348.

[12] B. Lončar *et al.*, "Do climate conditions affect the quality of the apiaceae fruits' essential oils?," *Horticulturae*, vol. 10, no. 6, p. 577, Jun. 2024, doi: 10.3390/horticulturae10060577.

[13] Y. Zhu, T. Li, and W. Li, "An efficient hybrid feature selection method using the artificial immune algorithm for high-dimensional data," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–21, Oct. 2022, doi: 10.1155/2022/1452301.

[14] T. Nguyen-Da, Y. M. Li, C. L. Peng, M. Y. Cho, and P. Nguyen-Thanh, "Tourism demand prediction after COVID-19 with dDeep learning hybrid CNN–LSTM—case study of Vietnam and Provinces," *Sustainability (Switzerland)*, vol. 15, no. 9, p. 7179, Apr. 2023, doi: 10.3390/su15097179.

[15] A. Al Tawil, L. Almazaydeh, B. Alqudah, A. Z. Abualkishik, and A. A. Alwan, "Predictive modeling for breast cancer based on machine learning algorithms and features selection methods," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, pp. 1937–1947, Apr. 2024, doi: 10.11591/ijece.v14i2.pp1937-1947.

[16] T. S. Wahyuni and K. K. Purwanto, "Students' conceptual understanding on acid-base titration and its relationship with drawing skills on a titration curve," *Journal of Physics: Conference Series*, vol. 1440, no. 1, p. 012018, Jan. 2020, doi: 10.1088/1742-6596/1440/1/012018.

[17] B. R. Bhamare, P. Jeyanthi, and R. Subhashini, "Aspect category extraction for sentiment analysis using multivariate filter method of feature selection," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 2138–2143, Sep. 2019, doi: 10.35940/ijrte.C4566.098319.

[18] S. Feng and M. F. Duarte, "Graph autoencoder-based unsupervised feature selection with broad and local data structure preservation," *Neurocomputing*, vol. 312, pp. 310–323, Oct. 2018, doi: 10.1016/j.neucom.2018.05.117.

[19] A. C. Siregar and B. C. Octariadi, "Classification of Sambas Traditional Fabric 'Kain Lunggi' using texture feature," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 4, p. 389, Oct. 2019, doi: 10.22146/ijccs.49782.
[20] W. Zhou, C. Wu, J. Wang, X. Yu, and Y. Yi, "Double regularized matrix factorization for image classification and clustering," *Eurasip Journal on Image and Video Processing*, vol. 2018, no. 1, p. 49, Dec. 2018, doi: 10.1186/s13640-018-0287-5.
[21] M. C. Á. Murillo, L. E. C. Suarez, and J. A. C. Salamanca, "Chemical composition and insecticidal properties of essential oils of piper septuplinervium and P. subtomentosum (Piperaceae)," *Natural Product Communications*, vol. 9, no. 10, pp. 1527–1530, Oct. 2014, doi: 10.1177/1934578x1400901031.
[22] N. A. S. A. Sabri *et al.*, "Statistical analysis for chemical compound based on several species of *Aquilaria* essential oil," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 4, pp. 3663–3673, Aug. 2024, doi: 10.11591/ijece.v14i4.pp3663-3673.
[23] Y. Z. H.-Y. Hashim, N. I. Ismail, and P. Abbas, "Analysis of chemical compounds of agarwood oil from different species by gas chromatography mass spectrometry (Gcms)," *IIUM Engineering Journal*, vol. 15, no. 1, pp. 1–6, May 2014, doi: 10.31436/iiumej.v15i1.469.
[24] R. Ning *et al.*, "First report on inhibitory effect against osteoclastogenesis of Dihydro-β-agarofuran-type sesquiterpenoids," *Journal of Agricultural and Food Chemistry*, vol. 70, no. 2, pp. 554–566, Jan. 2022, doi: 10.1021/acs.jafc.1c06862.
[25] M. Kampik, K. Bodzek, A. Piaskowy, A. Pilśniak, and M. Fice, "An analysis of energy consumption in railway signal boxes," *Energies*, vol. 16, no. 24, p. 7985, Dec. 2023, doi: 10.3390/en16247985.
[26] N. M. Azah *et al.*, "Chemometric study of selected agarwood oils by gas chromatography–mass spectrometry," *Journal of Tropical Forest Science*, vol. 26, no. 3, pp. 382–388, 2014.
[27] N. Ismail, M. A. Nor Azah, M. Jamil, M. H. F. Rahiman, S. N. Tajuddin, and M. N. Taib, "Analysis of high quality agarwood oil chemical compounds by means of SPME /GC-MS and Z-score technique," *Malaysian Journal of Analytical Sciences*, vol. 17, no. 3, pp. 403–413, 2013, [Online]. Available: http://umpir.ump.edu.my/6304/1/fist-2013-saiful-analysis_of_high_quality.pdf.
[28] J. Yin and X. Du, "High-dimensional reliability method accounting for important and unimportant input variables," *Journal of Mechanical Design*, vol. 144, no. 4, Apr. 2022, doi: 10.1115/1.4051982.

BIOGRAPHIES OF AUTHORS

Nur Athirah Syafiqah Noramli received her B. Sc. (Hons) in computer science from Universiti Teknologi MARA (UiTM) Cawangan Melaka Kampus Jasin. She is currently pursuing her studies as a postgraduate student at the Faculty of Electrical Engineering, at Universiti Teknologi MARA (UiTM) Shah Alam, Selangor, Malaysia. Her research interests include advanced signal processing, machine learning, and deep learning. She can be contacted at email: athirah.noramli1@gmail.com.

Noor Aida Syakira Ahmad Sabri received her Bachelor of Engineering (Hons) in electronic engineering from Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia, in 2022. Currently, she is pursuing postgraduate studies at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. Her research interests focus on advanced signal processing and machine learning. She can be contacted at email: aidasyakiraaa01@gmail.com.

Muhammad Ikhsan Roslan earned his Master of Science in Electronic Systems Design Engineering from Universiti Sains Malaysia (USM), Penang, Malaysia, in 2022 with first-class honors. He is currently a server validation engineer specializing in IP-level validation at UST (M) Sdn. Bhd, while also pursuing full-time postgraduate studies at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. With a strong passion for research in engineering, particularly in artificial intelligence, he combines academic excellence with practical experience, showcasing a dedicated commitment to advancing the field. He can be contacted at email: muhammadikhsanroslan@gmail.com

Assoc. Prof. Ir. Ts. Dr. Nurlaila Ismail received her Ph.D. in electrical engineering from Universiti Teknologi MARA, Malaysia. She is currently a senior lecturer at Faculty of Electrical Engineering, Universiti Teknologi MARA Shah Alam, Malaysia. Her research interests include advanced signal processing and artificial intelligence. She can be contacted at email: nurlaila0583@uitm.edu.my.

Assoc. Prof. Ts. Dr. Zakiah Mohd Yusoff received her Bachelor's Degree in electrical engineering and Ph.D. in electrical engineering from Universiti Teknologi MARA Shah Alam, in 2009 and 2014, respectively. She is a senior lecturer who is currently working at Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia. In Mei 2014, she joined Universiti Teknologi MARA as a teaching staff. Her major interests include process control, system identification, and essential oil extraction systems. She can be contacted at email: zakiah9018@uitm.edu.my.

Prof. Ir. Ts. Dr. Haji Mohd Nasir Taib received the degree in electrical engineering from the University of Tasmania, Hobart, Australia, the M.Sc. degree in control engineering from Sheffield University, UK, and the Ph.D. degree in instrumentation from the University of Manchester Institute of Science and Technology, UK. He is currently an Honorary Professor at Universiti Teknologi MARA (UiTM), Malaysia. He heads the Advanced Signal Processing Research Interest Group, Faculty of Electrical Engineering, UiTM. He has been a very active researcher and over the years had author and/or co-author many papers published in refereed journals and conferences. He can be contacted at email: dr.nasir@uitm.edu.my.