ISSN: 1410-8917
Jurnal Kimia Sains & Aplikasi
e-ISSN: 2597-9914

Jurnal Kimia Sains dan Aplikasi 28 (9) (2025): 512-521
Jurnal Kimia Sains dan Aplikasi
Journal of Scientific and Applied Chemistry
Journal homepage: http://ejournal.undip.ac.id/index.php/ksa

The Employment of ATR-FTIR Spectroscopy for Quantification of Turpentine in Cajuput Oil

Klarisa Sabila [1], Mazaya Amajida [1], Hisbullah [1], Syaifullah Muhammad [1,2], Hesti Meilina [1,2,*]

[1] Department of Chemical Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
[2] ARC PUI-PT Nilam Aceh, Universitas Syiah Kuala, Banda Aceh, Indonesia

* Corresponding author: hesti.meilina@usk.ac.id

https://doi.org/10.14710/jksa.28.9.512-521

Article Info

Article history:

Received: 15th June 2025
Revised: 16th November 2025
Accepted: 25th November 2025
Online: 08th December 2025

Keywords:
Cajuput Oil; Turpentine; FTIR; Chemometrics; PLS; PCA

Abstract

Cajuput oil is widely used for its therapeutic properties; however, its high economic value makes it vulnerable to adulteration with cheaper solvents, such as turpentine. This study employed ATR-FTIR spectroscopy combined with chemometric methods to qualitatively classify and quantitatively predict turpentine adulteration in cajuput oil. Cajuput oil samples were adulterated with turpentine at concentrations ranging from 0% to 10% (v/v) in 0.5% increments and analyzed in the mid-infrared region (4000–400 cm^{-1}). Spectral pretreatments, including multiplicative scatter correction (MSC), smoothing, baseline correction, second derivative, and standard normal variate (SNV), were applied prior to chemometrics analysis. FTIR spectra revealed the appearance of C=C stretching bands at 1647 and 1508 cm^{-1} only in the adulterated samples. In contrast, Principal Component Analysis (PCA) with SNV and MSC pretreatments provided clear clustering of samples according to turpentine concentration, with cumulative variances reaching up to 90%. Partial Least Squares (PLS) using MSC, smoothing, baseline correction, and SNV pretreatments yielded excellent calibration and cross-validation performance, with R^2 values of 0.98–0.99, and low SEC/SECV and RMSEC/RMSECV values. These results demonstrate that ATR-FTIR spectroscopy combined with appropriate chemometric pretreatments offers a rapid, solvent-free, and reliable approach for authentication and quality control of cajuput oil adulterated with turpentine.

1. Introduction

Indonesia, a tropical nation with rich biodiversity, is a prominent source of various essential oils with extensive applications across multiple industries. These aromatic essential oils serve as primary ingredients in cosmetics [1], aromatherapy [2], pharmaceuticals [3], and food products [4]. Among the diverse types of essential oils produced, cajuput oil derived from cajuput globulus of the Myrtaceae family stands out due to its unique aroma, health benefits, and high economic value [5]. The importance of cajuput oil in Indonesia is highlighted by its annual production, estimated at 650 tons, which is a significant contribution to the global production of around 3,500 tons annually [6]. The high demand for cajuput oil in Indonesia has caused the country to become one of its largest importers.

Cajuput oil is highly valued for its therapeutic properties, primarily due to its main chemical component, 1,8-cineole (also known as eucalyptol). This compound has shown effectiveness in respiratory treatments, provides calming effects, prevents insect bites, relieves joint pain, and supports digestive health [7]. Recent studies have also highlighted the antiviral potential of 1,8-cineole, demonstrating through in silico analyses its inhibitory effects against COVID-19 by binding to specific viral proteins [8]. The widespread health benefits associated with cajuput oil have led to a significant increase in consumer demand, making it an essential commodity in the Indonesian healthcare market.

However, the surge in demand has brought about challenges, particularly in ensuring the purity and authenticity of cajuput oil products available to consumers. Adulteration has become a prevalent issue, where some producers exploit the demand by introducing cheaper substitutes or diluting essential oils with other substances to lower production costs [9]. This practice not only undermines the medicinal efficacy of cajuput oil but also poses health risks, such as allergies and skin irritation, especially among consumers unable to distinguish between pure and adulterated products [10]. Thus, there is a critical need for reliable and efficient methods to authenticate cajuput oil, ensuring that products on the market are both safe and beneficial for consumers.

Traditional methods for evaluating the purity of cajuput oil primarily involve physicochemical analysis. Researchers commonly employ techniques such as gas chromatography-mass spectrometry (GC-MS) and measurements of specific gravity, optical rotation, alcohol solubility, and refractive index [11]. While effective, these methods often involve complex sample preparation, the use of hazardous chemicals, instrumental limitations, data handling challenges, environmental impact, and high costs [12]. In response to these limitations, researchers have increasingly turned to Fourier Transform Infrared (FTIR) spectroscopy combined with chemometrics as a promising alternative. FTIR spectroscopy enables rapid analysis [13] and non-destructive analysis of chemical composition by passing infrared radiation through a sample, where certain wavelengths are absorbed, while others are transmitted [14]. The resulting spectral data provide clues into the molecular structure of the sample, making it suitable for distinguishing between pure and adulterated cajuput oil.

Given the robust results from previous studies, this research employs FTIR spectroscopy in conjunction with chemometric methods, specifically Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression, to investigate the purity and quality of cajuput oil. PCA is instrumental in reducing data dimensionality [15], allowing for a clearer distinction between pure and adulterated Cajuput oil samples based on spectral characteristics. PCA generates score plots that help visualize the clustering of data [16]. This technique is particularly useful for evaluating the effectiveness of data pretreatment methods, as it shows how well they improve the separation of pure and adulterated samples. Pretreatment is a critical step in chemometrics that helps eliminate unwanted spectral variations, thereby improving the accuracy and interpretability of PCA results [17].

Conversely, PLS regression is employed as a calibration tool to quantify adulterants, such as turpentine, in cajuput oil samples [18]. This study specifically targets turpentine adulteration, as it is one of the more commonly encountered adulterants in essential oils [19]. Through the application of PLS regression, this study aimed to develop a model capable of accurately predicting turpentine concentration in cajuput oil, thereby providing a reliable tool for quality control. The accuracy of the model was evaluated using key indicators, including the coefficient of determination (R^2) and the standard error of calibration and prediction (SEC/SEP), to ensure its reliability and applicability in practical contexts [20].

This study aims to address the existing gaps in cajuput oil authentication through three primary objectives. First, it seeks to characterize the FTIR spectra of pure and adulterated cajuput oil samples by analyzing functional groups and absorbance intensities, which are crucial for identifying distinct spectral fingerprints associated with each sample type. Second, the study evaluates the impact of pretreatment on data classification by analyzing PCA-generated score plots, which provide a quantitative assessment of data separation quality. Finally, it assesses the accuracy of PLS calibration models in predicting turpentine concentrations within adulterated samples, aiming to establish a robust, cost-effective method for cajuput oil purity testing.

This research contributes to the expanding body of literature on essential oil authentication methods, specifically within the context of cajuput oil, which remains underexplored despite its substantial economic and therapeutic value. Chemometric methods, including Principal Component Analysis (PCA) and Partial Least Squares (PLS), offer dependable strategies for processing complex FTIR data, significantly enhancing sample differentiation capabilities and delivering accurate quantitative results, thereby mitigating the limitations inherent in relying solely on direct spectral analysis [21].

By employing FTIR-chemometric techniques, this study not only provides an innovative approach to the quality control of cajuput oil but also addresses the broader demand for efficient, environmentally sustainable authentication methods in the essential oil industry. These findings are expected to benefit producers, regulatory agencies, and consumers by ensuring access to authentic, high-quality cajuput oil products. Additionally, the methodological framework established in this study may serve as a foundation for future research on essential oil authentication, particularly for oils that face challenges similar to those posed by adulteration.

2. Experimental

2.1. Sample Preparation and Adulteration Proceedings

Cajuput oil samples were obtained from local distillation in Aceh. In this study, the oil was adulterated with turpentine, a common industrial solvent. The turpentine, sourced from PT Samiraschem Indonesia, is a natural solvent derived from the distillation of pine resin. Adulteration was performed by adding turpentine to cajuput oil at concentrations ranging from 0% to 10% (v/v) in increments of 0.5%, yielding 21 sample variations. Each preparation contained approximately 10 mL of cajuput oil to maintain consistency. All 21 concentrations were prepared in triplicate, resulting in a total of 63 samples. Prior to analysis, the samples were stored in glass containers under controlled conditions to prevent chemical alteration or contamination.

2.2. FTIR Measurement

The prepared samples were analyzed using Fourier Transform Infrared (FTIR) spectroscopy to obtain their respective absorption spectra. FTIR analysis was performed within the mid-infrared region, covering wavenumbers from 4000 to 400 cm^{-1}, to examine the absorption spectra of each sample. Measurements were conducted with a SHIMADZU IR-21 FTIR spectrometer using a scanning range optimized to capture key spectral features relevant to both cajuput oil and turpentine. Spectral acquisition parameters were carefully adjusted to ensure high-resolution data collection across the defined turpentine concentration gradient (0–10%, v/v) in 0.5% increments. Each sample spectrum was recorded and stored in an Excel format for subsequent data processing and chemometric analysis. The collected spectral data were subsequently processed using chemometric techniques to detect distinct variations associated with adulteration. FTIR analysis provided the initial dataset used as a reference for both quantitative and qualitative evaluation of adulteration levels.

2.3. Spectral Pretreatment

To enhance spectral clarity and ensure data consistency, several pretreatment techniques were applied to the FTIR spectra prior to chemometric analysis. These included second derivative transformation, smoothing, baseline correction, Multiplicative Scatter Correction (MSC), and Standard Normal Variate (SNV) transformation. The application of these pretreatments standardized the spectral data, minimized artifacts, and improved the interpretability of the patterns associated with different levels of adulteration [22].

2.4. Statistical Determination Using Multivariate Chemometrics

Chemometric analysis was performed using The Unscrambler X (CAMO Software) version 10.4.43636.111, employing both qualitative and quantitative approaches to characterize the levels of adulteration and spectral patterns. The wavenumber range used for multivariate analysis was 4000–400 cm^{-1}. This range was selected because it includes the key functional group vibrations relevant to both the main components (cajuput oil) and the adulterant (turpentine). This range captures important molecular information, such as the stretching and bending vibrations of functional groups, which are crucial for distinguishing pure cajuput oil from adulterated samples.

Principal Component Analysis (PCA) was conducted as a qualitative analysis to interpret the distribution and grouping of spectra, facilitating differentiation between pure and adulterated samples. PCA allowed for the visualization of spectral variability and the identification of distinct clusters associated with varying turpentine concentrations. Partial Least Squares (PLS) was chosen as the quantitative analysis method due to its superior ability to handle collinearity in the data and provide more reliable predictions of concentration values in complex datasets. PLS can simultaneously model the relationship between the spectra and the concentration of turpentine, which makes it particularly suited for this type of analysis. For the PLS modelling, a total of 63 ATR-FTIR spectra were used to construct the calibration model. The same 63 spectra were subsequently subjected to an internal cross-validation method, as no separate external validation set was defined. PLS models were developed using pre-treated spectral data, with the following statistical parameters evaluated to assess model accuracy: coefficient of determination (R^2), standard error of calibration (SEC), root mean square error of calibration (RMSEC), standard error of cross-validation (SECV), and root mean square error of cross-validation (RMSECV).

3. Results and Discussion

3.1. Characterizations of Pure Cajuput Oil and Adulterated Cajuput Oil

This section presents and discusses the FTIR spectra of both pure cajuput oil and cajuput oil samples adulterated with turpentine. The analysis aims to distinguish pure cajuput oil from adulterated samples by identifying spectral variations induced by the incorporation of turpentine at varying concentrations. Figure 1 illustrates the FTIR spectra of pure cajuput oil and the adulterated samples across the defined turpentine concentration range.

As shown in Figure 1, the FTIR spectral profiles of both the 0% and 10% cajuput oil samples exhibited similar overall patterns. However, noticeable variations in absorbance values and intensities at specific wavenumbers indicate the presence of adulterant compounds in the 10% sample. Further insights into the functional group differences between the samples are presented in Table 1.

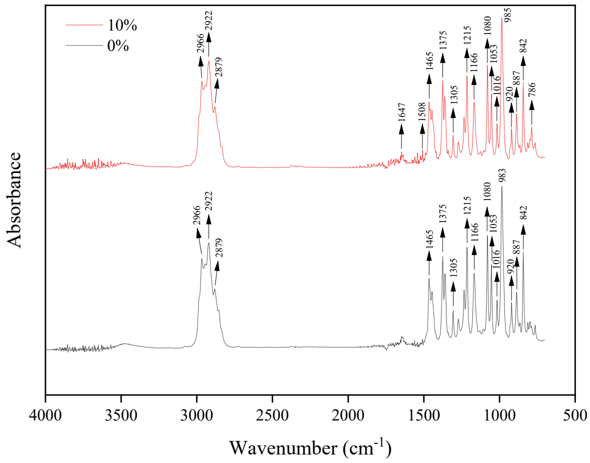

Figure 1. Spectra of pure cajuput oil (0% turpentine) and adulterated cajuput oil (10% turpentine)

Table 1. Identification of functional groups in cajuput oil samples with varying turpentine concentrations

No.	Functional groups	0% turpentine		10% turpentine	
		Wavenumber (cm^{-1})	Intensity	Wavenumber (cm^{-1})	Intensity
1	C-H stretching	2966	strong	2966	strong
		2922	strong	2922	strong
		2879	strong	2879	strong
2	C=C stretching	-	-	1647	medium
		-	-	1508	medium
3	C-H bending	1465	medium	1465	medium
		1375	medium	1375	medium
4	O-H bending	1305	medium	1305	medium
5	C-O stretching	1215	strong	1215	strong
		1166	strong	1166	strong
		1080	strong	1080	strong
		1053	strong	1053	strong
		1016	strong	1016	strong
6	C=C bending	983	strong	985	strong
		920	strong	920	strong
		887	strong	887	strong
		842	medium	842	medium
		-	-	786	medium

The principal component of cajuput oil is 1,8-cineole, which is represented by the chemical formula $C_{10}H_{18}O$. Functional groups indicative of the presence of 1,8-cineole, detectable by FTIR, include C-H, O-H, and C-O bonds. Conversely, the main component of turpentine, which acts as an adulterant, is α-pinene with the chemical formula $C_{10}H_{16}$. The chemical bonds of α-pinene include C-H and C=C bonds.

The results from both spectra show similar patterns, as shown in Figure 1. This was identified by peaks appearing within specific functional groups. The C-H functional group, observed in the 3000–2850 cm^{-1} wavenumber range, corresponds to stretching vibrations in alkane compounds. In adulterated cajuput oil (10% turpentine), peaks at 1647 cm^{-1} and 1508 cm^{-1} indicate the presence of C=C stretching bonds, signifying turpentine content [23]. Peaks at 1465 cm^{-1} and 1375 cm^{-1} correspond to moderate-intensity C-H bending vibrations, while the peak at 1305 cm^{-1} represents moderate-intensity O-H bending. Additionally, the C-O functional group, observed within the 1000–1300 cm^{-1} range, suggests the presence of 1,8-cineole, whereas the C=C bending functional group appears within the 995–665 cm^{-1} range, further indicating the turpentine content. A key difference between the two spectra is the presence of the C=C stretching group, which appears exclusively in the adulterated cajuput oil sample, whereas the C=C bending group is present in both samples.

Visually differentiating the spectra, particularly in terms of the functional groups, of pure cajuput oil and adulterated samples poses significant challenges. To overcome these limitations, a comprehensive set of investigative approaches has been adopted, incorporating chemometric techniques to enhance spectral interpretation [24]. Chemometrics, which integrates mathematics, chemistry, and statistical analysis, provides a powerful framework for detecting subtle differences in spectral data that are not readily observable by visual inspection alone [25]. This analytical strategy enables a more accurate assessment of functional group variations and supports the verification of adulterant presence within the samples. The following sections describe the chemometric methods employed to achieve these objectives.

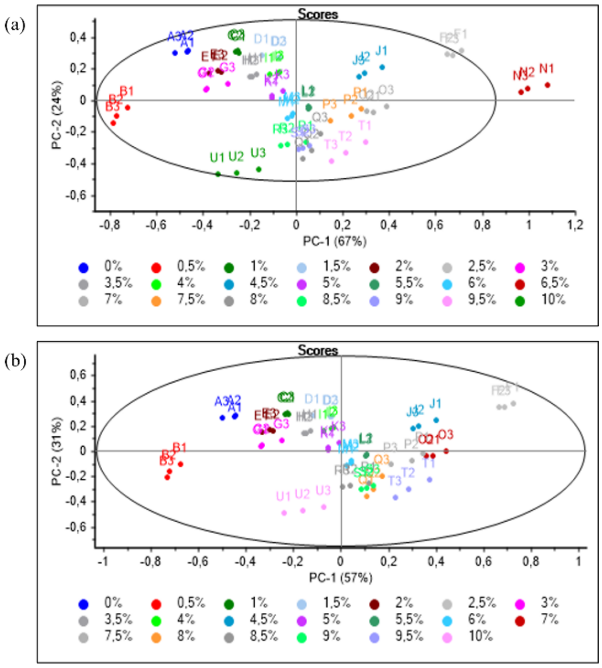

Figure 2. Hotelling T^2 ellipse of (a) raw spectra without removal of outliers, (b) removal of outliers

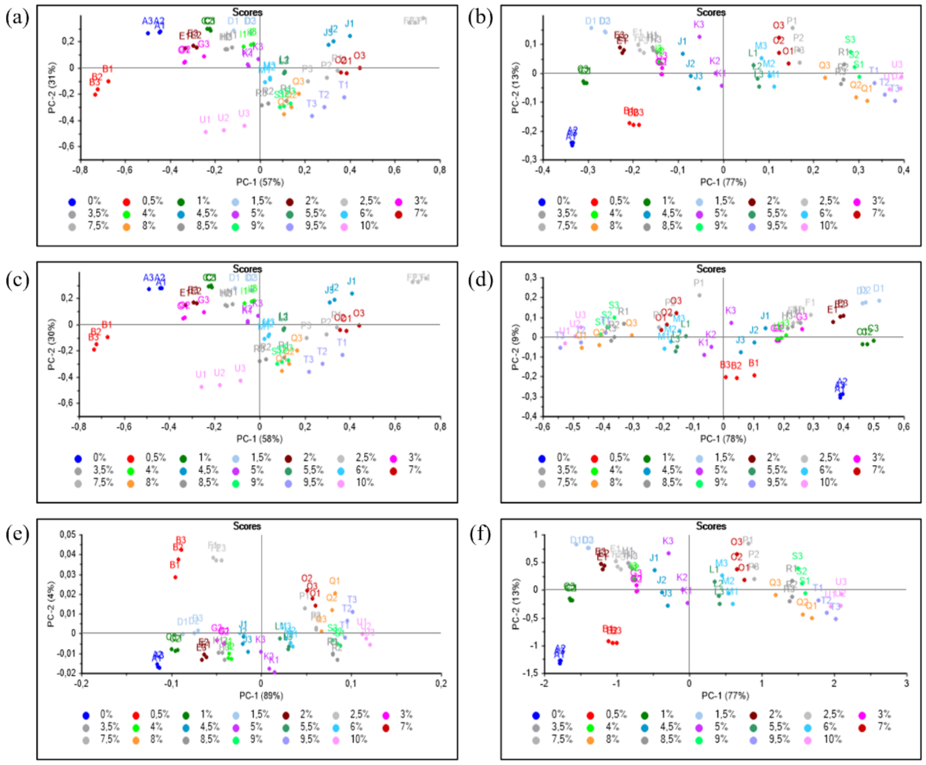

Figure 3. PCA score plots of ATR-FTIR spectra with different spectral pretreatments: (a) raw, (b) MSC, (c) smoothing, (d) baseline correction, (e) 2nd derivative, and (f) SNV

3.2. Principal Component Analysis (PCA)

The analysis of PCA clustering began with an essential preliminary step, such as outlier removal. This process was implemented to prevent extreme data points from distorting dimensionality reduction. By excluding such outliers, PCA becomes more robust, and the derived components more faithfully capture the actual patterns within the data [26]. This provides a more reliable basis for subsequent interpretations and analyses.

Outlier removal is a process of identifying data points that fall outside the ellipse and is conducted using the PCA method in conjunction with Hotelling's T^2 ellipse [27]. Hotelling's T^2 ellipse was utilized to assess whether the data were sufficiently well-behaved for further analysis [28]. In this process, a significance level of 10% (90% confidence level) was applied to determine the threshold for outlier identification, ensuring that only data within acceptable limits were retained for subsequent analysis.

The results of Hotelling's T^2 ellipse in Figure 2(a) prior to outlier removal, reveal a total of 63 data points. Three data points, N1, N2, and N3, at a turpentine concentration of 6.5% were identified as outliers. Outliers are defined as data points that exhibit patterns significantly different from those of the majority of the dataset [29]. After performing outlier removal using Hotelling's T^2 ellipse, as shown in Figure 2(b), the dataset was reduced from 63 to 60 samples. This refinement allowed the data analysis process to proceed to the clustering analysis phase using PCA.

Clustering analysis is a data grouping technique that utilizes PCA with spectral data pre-treatment methods to improve the accuracy of the resulting models [30]. The pretreatment of spectral data adjusts spectral algorithms to enhance data quality by reducing wave interference and noise, ultimately producing a more reliable and stable model [31]. Following pretreatment, the data underwent PCA using the Unscrambler software, allowing for more precise clustering of the samples. The pretreatment methods applied in this analysis include MSC, smoothing, baseline correction, second derivative transformation, and SNV, all of which are instrumental in reducing spectral variability and enhancing model interpretability. These methods are widely recognized for their efficacy in producing high-quality spectral data for robust multivariate analyses, as documented in several studies [32]. Figure 3 shows a plot of the PCA scores after outlier removal. The detailed results of the PCA scores for several pre-treatment techniques are presented in Table 2.

Table 2. Cumulative variance of PCA scores with different pretreatment techniques

Pretreatment	PC-1 (%)	PC-2 (%)	Cumulative variance (%)
Raw spectra	57	31	88
MSC	77	13	90
Smoothing	58	30	88
Baseline	78	9	87
2nd derivative	89	4	93
SNV	77	13	90

PCA clustering analysis results with the applied pretreatments demonstrated that certain methods, specifically SNV and MSC, effectively clustered turpentine concentrations within the data. For each PCA pretreatment, the total clustering capability was assessed using cumulative variance values, calculated as the sum of the variance explained by the first two principal components (PC-1 and PC-2). The high cumulative variance scores indicate that PC-1 and PC-2 together capture the essential information and variability in the original spectral data, aligning with findings from prior research that highlight the effectiveness of these components in spectral clustering tasks [33]. Based on cumulative variance, the 2nd derivative, SNV, and MSC pretreatments achieved the highest total variances at 93%, 90%, and 90%, respectively, as shown in Figure 3(e), 3(f), and 3(b).

Beyond cumulative variance in PCA, each pretreatment entails distinct advantages and trade-offs. SNV and MSC effectively normalize scattering and path-length variation, thereby enhancing cluster separation in PCA while preserving the linearity of the spectra–concentration relationship in PLS (as reflected by high calibration and validation R^2). Smoothing and baseline correction reduce noise and baseline drift, yielding more stable PLS predictions without removing relevant intensity information. In contrast, the second derivative sharpens narrow band features often manifested as higher PCA variance yet may amplify noise and diminish absolute intensity information required for regression, ultimately lowering R^2 in PLS.

Further analysis of clustering using SNV and MSC pretreatments revealed two primary clusters based on PC-1, with PC-1<0, representing turpentine concentrations from 0% to 5%, and another with PC-1>0, corresponding to concentrations from 5.5% to 10%. This clustering along PC-1 signifies that SNV and MSC pretreatments are effective in distinguishing between lower and higher concentrations of the adulterant. Additionally, clustering along PC-2 provided further resolution, with PC-2<0 corresponding to turpentine concentrations between 0% and 0.5%, and PC-2>0 representing concentrations from 1% to 5%. Within the PC-1 range, values between 0 and 0.5 clustered samples containing 5.5–7.5% turpentine, while PC-1>0.2 represented higher concentrations ranging from 8–10%.

The ability of SNV and MSC pretreatments to generate well-defined clusters across the 0% to 10% concentration range supports their efficacy in enhancing PCA-based spectral interpretation. This clustering performance is consistent with the high cumulative variance observed, confirming that these algorithms are particularly suited for the spectral analysis of complex mixtures such as essential oils [34]. The identification of distinct clusters related to different concentration levels provides a robust framework for detecting adulteration in cajuput oil, as it indicates that SNV and MSC are capable of reliably distinguishing between cajuput oil and adulterant compounds across a range of concentrations. These results are consistent with previous studies highlighting the critical role of appropriate spectral pretreatment in enhancing data quality and clustering accuracy in chemometric analyses [35].

3.3. Partial Least Squares (PLS)

PLS is a multivariate regression method widely used in chemical analysis, particularly for building predictive models based on spectral data [36]. In this study, PLS was employed to quantify turpentine in cajuput oil accurately by utilizing infrared spectra as predictor variables. PLS models were analyzed through calibration and validation processes. The parameters used to determine the influence of the wave number range on PLS analysis are the coefficient of determination (R^2), SEC, RMSEC, SECV, and RMSECV.

The calibration and cross-validation results for the different spectral pretreatments are summarized in Table 3. In general, the PLS models exhibited high predictive ability for most pretreatments, with R^2 values ranging from 0.98 to 0.99 for both calibration and validation in the MSC, smoothing, baseline correction, and SNV pretreatments. These high R^2 values indicate that a large proportion of the variation in turpentine concentration is explained by the FTIR spectral information and that the models have strong predictive capability. In addition, the SEC/RMSEC and SECV/RMSECV values obtained for these pretreatments are very low, confirming that the prediction errors are small.

Table 3. Performance of PLS models using different spectral pretreatments

Pretreatment	Calibration			Validation		
	R^2	SEC	RMSEC	R^2	SECV	RMSECV
Raw spectra	0.74	0.0063	0.0062	0.74	0.0063	0.0062
MSC	0.98	0.00081	0.0008	0.98	0.00081	0.0008
Smoothing	0.99	0.001	0.001	0.99	0.001	0.001
Baseline	0.98	0.001	0.001	0.98	0.001	0.001
2nd derivative	0.58	0.0002	0.0002	0.58	0.0002	0.0002
SNV	0.98	0.0043	0.0043	0.98	0.0043	0.0043

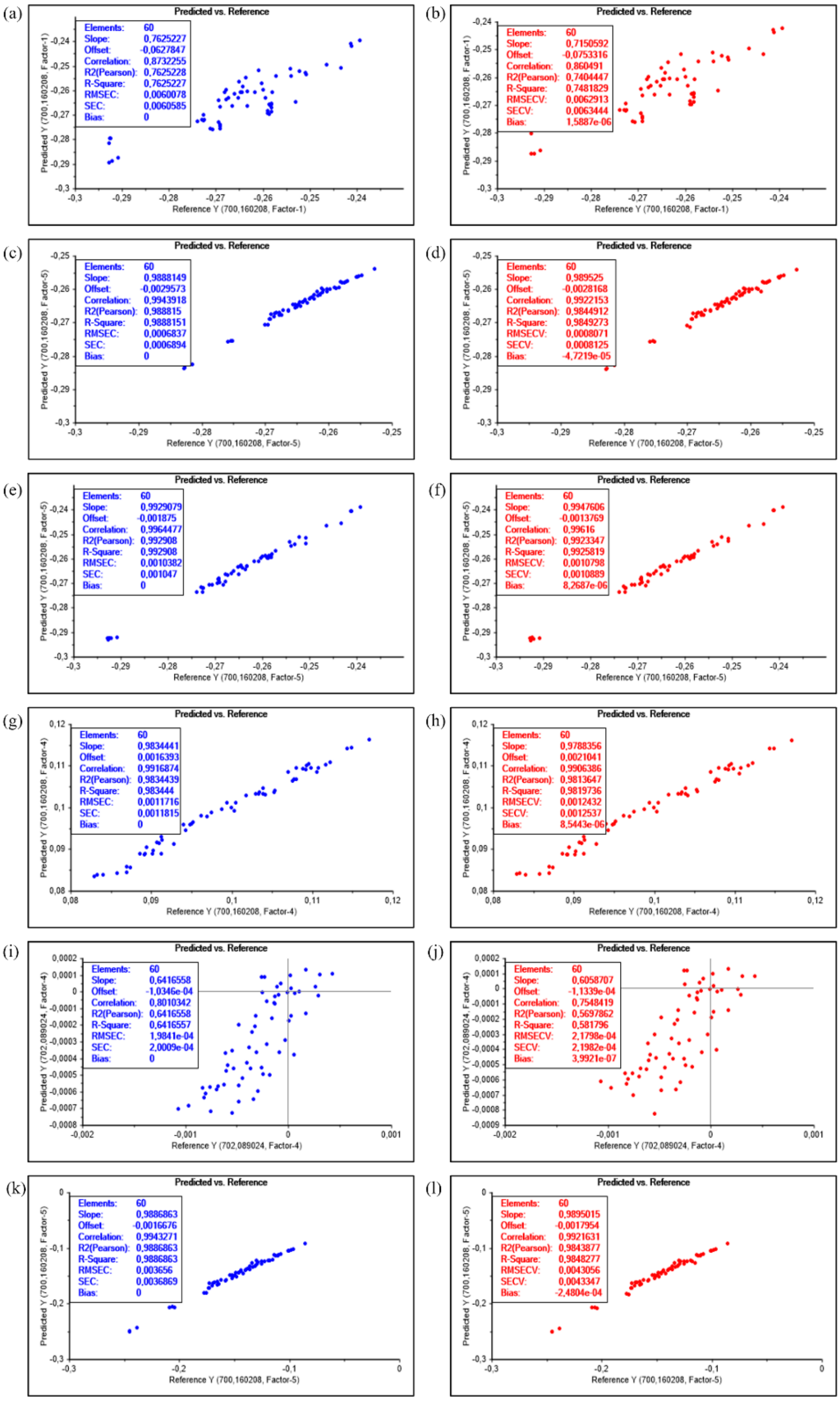

Figure 4. Reference vs predicted plots of PLS models for calibration (left column) and cross-validation (right column) using different spectral pretreatments: raw spectra (a-b), MSC (c-d), smoothing (e-f), baseline correction (g-h), 2nd derivative (i-j), and SNV (k-l)

The comparison between calibration and validation errors also suggests that the models are stable and do not suffer from severe overfitting. For example, the MSC pretreatment yielded an R^2 of 0.98 for both calibration and validation, with SEC and SECV values of 0.0006 and 0.00081, respectively. A similar trend is observed for the smoothing and baseline pretreatments, where the differences between calibration and validation errors are minimal, indicating that the PLS models can be reliably applied to new samples within the same population.

In contrast, the 2nd derivative pretreatment resulted in noticeably lower R^2 values for both calibration (0.64) and validation (0.58), despite its very small error values. The relatively low R^2 indicates that a smaller portion of the variance in the reference turpentine concentration is captured by the model, resulting in weaker predictive performance compared to the other pretreatments. This suggests that the 2nd derivative transformation may introduce additional noise or distortions in the spectra that are not fully compensated by the PLS models.

The scatter plots of calibration and validation data for each pretreatment, shown in Figure 4, further support these findings. For the MSC, smoothing, baseline correction, and SNV pretreatments, the data points are closely clustered around the 1:1 line, indicating good agreement between predicted and reference values in both calibration and validation sets. By contrast, the 2nd derivative pretreatment shows a wider dispersion of points, consistent with its lower R^2 values. Taken together, these results demonstrate that the application of appropriate spectral pretreatments, particularly MSC, smoothing, and baseline correction, significantly improves the performance and robustness of the PLS models for turpentine quantification in cajuput oil.

The application of chemometrics, mathematical and statistical techniques for analyzing chemical data, enhances the interpretative power of FTIR spectroscopy. When combined with chemometrics, FTIR spectroscopy has shown potential for the rapid and accurate classification of essential oil samples, even in cases where adulterants are present at low concentrations [37]. This study has several limitations that need to be addressed. The number of samples used was insufficient to represent cajuput oil produced with different distillation methods and commercially available cajuput oil. This limited sample size may affect the ability of PCA and PLS models to optimally recognize patterns owing to the lack of representation of a wide range of data variations. Furthermore, models built with a small number of samples are prone to overfitting, which can reduce the prediction accuracy and generalization ability of the model. A combination of FTIR spectroscopy and chemometrics generates complex spectral data. This requires expertise to ensure that the analysis is performed accurately and precisely.

The results of this study contribute to the qualitative and quantitative analysis of essential oils, particularly cajuput oil. Although the samples used do not encompass variations from different distillation processes or commercial samples on the market, this study demonstrates that chemometrics approaches based on PCA and PLS can be effectively applied to such analyses. However, limitations of this research include the relatively small sample size and the lack of variation in the samples, which may limit the generalizability of the findings to a broader range of cajuput oil samples from different regions or production methods. Additionally, the study focused only on the adulteration with turpentine, and other types of adulterants or environmental factors influencing the oil composition were not considered. To further explore the chemical characteristics of these findings, further research involving more samples from various types of cajuput oil is needed. Overall, this study proves that cajuput oil can be analyzed both qualitatively and quantitatively using a chemometrics approach, namely PCA and PLS. By demonstrating the effectiveness of the employed methodological approach, this study opens up opportunities for quality control of cajuput oil using FTIR spectrum-based chemometrics methods.

4. Conclusion

Spectral characterization of pure and adulterated Cajuput oil revealed that pure Cajuput oil does not exhibit C=C stretching bonds, whereas the adulterated sample demonstrated the presence of these bonds at wavenumbers of 1647 cm^{-1} and 1508 cm^{-1}. Furthermore, the PCA method utilizing SNV and MSC pretreatments was the most effective for clustering, achieving a cumulative variance of 90% (PC-1:77% and PC-2:13%). In addition, the performance of PLS models using multiple scatter correction (MSC), smoothing, baseline correction, and standard normal variate (SNV) pretreatments showed excellent predictive accuracy and robustness, whereas the 2nd pretreatment produced lower R^2 values and more scattered predictions, indicating poorer suitability for quantitative FTIR–PLS analysis of turpentine in cajuput oil.

Acknowledgments

I would also like to express my gratitude to Syiah Kuala University, especially the Department of Chemical Engineering, for the academic support and facilities provided during this research. I am sincerely grateful to all individuals who contributed to the completion of this research. Special thanks are due to Dr. Ir. Hesti Meilina, S.T., M.Si., for her valuable guidance, advice, and continuous support throughout the research process.

References

[1] Aswandi Aswandi, Cut Rizlani Kholibrina, Harlinda Kuspradini, Essential Oils for Cosmetics Application, in: E.T. Arung, W. Fatriasari, I.W. Kusuma, H. Kuspradini, K. Shimizu, Y.-u. Kim, N.I.W. Azelee, Z. Edis (Eds.) *Biomass-based Cosmetics: Research Trends and Future Outlook*, Springer Nature Singapore, Singapore, 2024, https://doi.org/10.1007/978-981-97-1908-2_7

[2] Henri Boby, L'aromathérapie, une technique complémentaire en médecine périopératoire: Aromatherapy, a complementary technique in perioperative medicine, *Oxymag*, 35, 183, (2022), 10-14 https://doi.org/10.1016/j.oxy.2022.02.005

[3] Sara Baptista-Silva, Sandra Borges, Oscar L. Ramos, Manuela Pintado, Bruno Sarmento, The progress of essential oils as potential therapeutic agents: a review, *Journal of Essential Oil Research*, 32, 4, (2020), 279-295 https://doi.org/10.1080/10412905.2020.1746698

[4] Charles Oluwaseun Adetunji, Javad Sharifi-Rad, Application of essential oils in the food industry, in: *Applications of Essential Oils in the Food Industry*, Elsevier, 2024, https://doi.org/10.1016/B978-0-323-98340-2.00001-8

[5] Phuong Ha Tran, Thi Thanh Tam Vu, Thi Diem Tran Phan, Van Mien Nguyen, Thi Nghia Minh Ngo, Canh Viet Cuong Le, That Huu Dat Ton, Chemical compositions and biological properties of the leaf essential oil of three *Melaleuca* species, *World Academy of Sciences Journal*, 6, 6, (2024), 67 https://doi.org/10.3892/wasj.2024.282

[6] A. Rimbawanto, N. K. Kartikawati, F. Latumahina, Conservation and utilization of *Melaleuca cajuputi* sub sp *cajuputi*, an indigenous species in Moluccas Island, Indonesia, *IOP Conference Series: Earth and*

Environmental Science, 800, 1, (2021), 012040 https://doi.org/10.1088/1755-1315/800/1/012040

[7] Zdeňka Navrátilová, Cineol (eukalyptol) v léčbě respiračních onemocnění (Cineole (eucalyptol) in the treatment of respiratory diseases), *Česká a Slovenská Farmacie*, 73, 3, (2024), 181-186 https://doi.org/10.36290/csf.2024.028

[8] Saghar Ketebchi, Maryam Papari Moghadamfard, A review on the effective natural compounds of medicinal plants on the COVID-19, *Natural Product Research*, 39, 4, (2025), 834-847 https://doi.org/10.1080/14786419.2024.2309322

[9] Jan Schripsema, Sônia Maria da Silva, Denise Dagnino, Differential NMR and chromatography for the detection and analysis of adulteration of vetiver essential oils, *Talanta*, 237, (2022), 122928 https://doi.org/10.1016/j.talanta.2021.122928

[10] Tzi Bun Ng, Evandro Fei Fang, Alaa El-Din Ahmed Bekhit, Jack Ho Wong, Chapter 2 - Methods for the Characterization, Authentication, and Adulteration of Essential Oils, in: V.R. Preedy (Ed.) *Essential Oils in Food Preservation, Flavor and Safety*, Academic Press, San Diego, 2016, https://doi.org/10.1016/B978-0-12-416641-7.00002-X

[11] S. Idrus, M. S. Radiena, Sumarsana, H. Smith, Quality and Chemical Composition of Cajuput Oil from Moluccas and Papua, *Journal of Physics: Conference Series*, 1463, (2020), 012016 http://doi.org/10.1088/1742-6596/1463/1/012016

[12] Adissu Alemayehu Asfaw, Juan Aspromonte, Kris Wolfs, Ann Van Schepdael, Erwin Adams, Overview of sample introduction techniques prior to GC for the analysis of volatiles in solid materials, *Journal of Separation Science*, 42, 1, (2019), 214-225 https://doi.org/10.1002/jssc.201800711

[13] K. Bounaas, N. Bouzidi, Y. Daghbouche, S. Garrigues, M. de la Guardia, M. El Hattab, Essential oil counterfeit identification through middle infrared spectroscopy, *Microchemical Journal*, 139, (2018), 347-356 https://doi.org/10.1016/j.microc.2018.03.008

[14] Daniel Cozzolino, Infrared Spectroscopy, in: *Electromagnetic Technologies in Food Science*, 2021, https://doi.org/10.1002/9781119759522.ch12

[15] Aymen Adil Lazim, Ali Farhan Nader, Huda Fannoosh Khazaal Alsaad, Mahdi Khamees Ghadeer, Application of Principal Component Analysis for Enhanced Well Log Interpretation: A Case Study from the Zubair Oil Field, South of Iraq, *The Iraqi Geological Journal*, 58, 1C, (2025), 83-94 https://doi.org/10.46717/igj.58.1C.8ms-2025-3-23

[16] Laís Feltrin Sidou, Endler Marcel Borges, Teaching Principal Component Analysis Using a Free and Open Source Software Program and Exercises Applying PCA to Real-World Examples, *Journal of Chemical Education*, 97, 6, (2020), 1666-1676 https://doi.org/10.1021/acs.jchemed.9b00924

[17] Jean-Michel Roger, Jean-Claude Boulet, Magida Zeaiter, Douglas N. Rutledge, 3.01 - Pre-processing Methods☆, in: S. Brown, R. Tauler, B. Walczak (Eds.) *Comprehensive Chemometrics (Second Edition)*, Elsevier, Oxford, 2020, https://doi.org/10.1016/B978-0-12-409547-2.14878-4

[18] Siong Fong Sim, Min Xuan Laura Chai, Amelia Laccy Jeffrey Kimura, Prediction of Lard in Palm Olein Oil Using Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Partial Least Squares Regression (PLSR) Based on Fourier-Transform Infrared (FTIR), *Journal of Chemistry*, 2018, 1, (2018), 7182801 https://doi.org/10.1155/2018/7182801

[19] Any Guntarti, Laela Hayu Nurani, Putri Lestari, Citra Ariani Edityaningrum, Lalu Muhammad Irham, Abdul Rohman, Authentication of Clove Leaf Oil in Products (*Syzygium aromaticum* (L.) Merr. & LM Perry) Using GC-MS and FTIR Methods Combined with Chemometric, *Malaysian Journal of Analytical Sciences*, 28, 3, (2024), 664-680

[20] Q. Ding, M. Yao, Sh Wu, M. Zeng, N. Xue, D. Wu, J. Xu, Quantitative Analysis of Al in Flour Products by Laser-Induced Breakdown Spectroscopy Combined with Partial Least Squares, *Journal of Applied Spectroscopy*, 89, 4, (2022), 712-718 https://doi.org/10.1007/s10812-022-01415-4

[21] Sonny Widiarto, Raisha Fauziyah, Triana Puji Astari, Ni Luh Gede Ratna Juliasih, Sutopo Hadi, La Zakaria, Irwan Saputra, Authentication of Processed Beef Sausage Products Using Chemometric Analysis Based on FTIR Spectrophotometry Data, *Jurnal Kimia Sains dan Aplikasi*, 28, 1, (2025), 39-46 https://doi.org/10.14710/jksa.28.1.39-46

[22] Puneet Mishra, Jean Michel Roger, Federico Marini, Alessandra Biancolillo, Douglas N. Rutledge, Parallel pre-processing through orthogonalization (PORTO) and its application to near-infrared spectroscopy, *Chemometrics and Intelligent Laboratory Systems*, 212, (2021), 104190 https://doi.org/10.1016/j.chemolab.2020.104190

[23] Ng Hui Ying, Gan Shi Min, Saliza Asman, Extraction of essential oil Citrus Hystrix by using ultrasound-assisted extraction, *Malaysian Journal of Analytical Sciences*, 27, 6, (2023), 1236-1248

[24] Meryeme El Maouardi, Kris De Braekeleer, Abdelaziz Bouklouze, Yvan Vander Heyden, Comparison of Near-Infrared and Mid-Infrared spectroscopy for the identification and quantification of argan oil adulteration through PCA, PLS-DA and PLS, *Food Control*, 165, (2024), 110671 https://doi.org/10.1016/j.foodcont.2024.110671

[25] Shuxia Guo, Jürgen Popp, Thomas Bocklitz, Chemometric analysis in Raman spectroscopy from experimental design to machine learning–based modeling, *Nature Protocols*, 16, 12, (2021), 5426-5459 https://doi.org/10.1038/s41596-021-00620-3

[26] Guihong Wan, Baokun He, Haim Schweitzer, The art of centering without centering for robust principal component analysis, *Data Mining and Knowledge Discovery*, 38, 2, (2024), 699-724 https://doi.org/10.1007/s10618-023-00976-y

[27] Faizan E. Mustafa, Ahmed Ijaz, Maaruf Muhammad, Khalid Muhammad, Detection of Cracks in the Industrial System Using Adaptive Principal Component Analysis and Wavelet Denoising, *2024 IEEE International Conference on Industrial Technology (ICIT)*, Bristol, United Kingdom, 2024 https://doi.org/10.1109/ICIT58233.2024.10540880

[28] Ross Sparks, Monitoring Highly Correlated Multivariate Processes Using Hotelling's T^2 Statistic: Problems and Possible Solutions, *Quality and*

Reliability Engineering International, 31, 6, (2015), 1089-1097 https://doi.org/10.1002/qre.1656

[29] Wong Hui Shein, Anwar Fitrianto, A Comparative Study of Outliers Identification Methods in Univariate Data Set, *Advanced Science Letters*, 23, 2, (2017), 1422-1427 https://doi.org/10.1166/asl.2017.8366

[30] Pengjie Zhang, Bing Liu, Xihui Mu, Jiwei Xu, Bin Du, Jiang Wang, Zhiwei Liu, Zhaoyang Tong, Performance of Classification Models of Toxins Based on Raman Spectroscopy Using Machine Learning Algorithms, *Molecules*, 29, 1, (2024), 197 https://doi.org/10.3390/molecules29010197

[31] Dongze Li, Guifang Wu, Hai Ma, Zhao Liu, Guiquan Liu, Junhua Hou, Identification of Blueberry Beverage Using Vis/NIR Spectroscopy, *3rd International Conference on Mechanical, Electronic and Information Technology Engineering (ICMITE 2017)*, 2017 https://doi.org/10.1051/matecconf/201713900050

[32] Zhushanying Zhang, Hanwen Gu, Kaiwen Xie, He Jiang, Qinlan Xie, Jiming Sa, Pretreatment and combined method based on near infrared spectroscopy, *Laser & Optoelectronics Progress*, 58, 16, (2021), 1617001 http://dx.doi.org/10.3788/LOP202158.1617001

[33] Shazlyn Milleana Shaharudin, Norhaiza Ahmad, Choice of Cumulative Percentage in Principal Component Analysis for Regionalization of Peninsular Malaysia Based on the Rainfall Amount, Singapore, 2017 https://doi.org/10.1007/978-981-10-6502-6_19

[34] Snezana Agatonovic-Kustrin, Petar Ristivojevic, Vladimir Gegechkori, Tatiana M. Litvinova, David W. Morton, Essential Oil Quality and Purity Evaluation via FT-IR Spectroscopy and Pattern Recognition Techniques, *Applied Sciences*, 10, 20, (2020), 7294 https://doi.org/10.3390/app10207294

[35] A. Anne Frank Joe, A. Gopal, A study on various preprocessing algorithms used for NIR spectra, *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 7, 4, (2016), 2752-2757

[36] Le Chang, Jiali Wang, William Woodgate, Analysing spectroscopy data using two-step group penalized partial least squares regression, *Environmental and Ecological Statistics*, 28, 2, (2021), 445-467 https://doi.org/10.1007/s10651-021-00496-2

[37] Archasvi Tyagi, Anil K. Yadav, Akanksha Yadav, Lalita Saini, Vivek Kumar, Pooja Jain, Inam Mohammad, Mohammad Javed Ansari, Hesham Ali El Enshasy, Fagr Kh Abdel-Gawad, Sami Al Obaid, Shahida Anusha Siddiqui, Vijai Malik, Vibrational Spectroscopic Methods for the Identification and Distinction of Essential Oils in *Genus Ocimum* L.: A Chemometric Approach, *Journal of King Saud University – Science*, 34, (2022), 102355 https://doi.org/10.1016/j.jksus.2022.102355