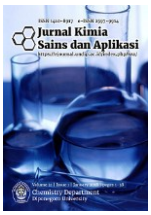
ISSN: 1410-8917
Jurnal Kimia Sains & Aplikasi
e-ISSN: 2597-9914

Jurnal Kimia Sains dan Aplikasi 29 (4) (2026): 296-305

Jurnal Kimia Sains dan Aplikasi
Journal of Scientific and Applied Chemistry

Journal homepage: http://ejournal.undip.ac.id/index.php/ksa



Detection of Adulteration in Coffee Products Using FTIR Spectroscopy and Multivariate Analysis

Sonny Widiarto [1,*], La Zakaria [2], Safitri Haurotul Jamalat [1], Dian Septiani Pratama [1], Irwan Saputra [3]

[1] Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Bandar Lampung, Indonesia
[2] Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Bandar Lampung, Indonesia
[3] Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, Jakarta, Indonesia

* Corresponding author: sonny@fmipa.unila.ac.id

https://doi.org/10.14710/jksa.29.4.296-305

Article Info

Article history:

Received: 18th December 2025
Revised: 11th April 2026
Accepted: 22nd April 2026
Online: 25th May 2026

Keywords:
Coffee Authentication; FTIR Spectroscopy; PCA; PLS; Chemometrics

Abstract

Food fraud, particularly in coffee, is increasingly prevalent worldwide and poses significant risks to consumers. This study aimed to analyze the purity of coffee products in the Bandar Lampung market and evaluate the validity of Fourier Transform Infrared (FTIR) spectroscopy and chemometric methods. The research methodology includes preparing reference standards, coffee sampling from various sellers in Bandar Lampung, establishing calibration and validation sets, performing FTIR analysis, and data processing using Principal Component Analysis (PCA) and Partial Least Squares (PLS) with Minitab software. A total of five coffee samples collected from the Bandar Lampung market were analyzed, and this work should be considered a preliminary investigation due to the limited sample size. PCA was employed as an exploratory tool to classify coffee types and identify potential adulterants. The PCA results indicated that samples A, B, and C clustered closely with the robusta coffee standard, while samples D and E showed slight deviation, exhibiting spectral characteristics associated with corn powder. These findings are consistent with the regional context of Lampung, one of the major coffee-producing regions in Indonesia, where robusta is predominantly cultivated. In addition, PCA suggests that corn is the dominant adulterant compared to rice. Based on these observations, the quantitative analysis was performed using a PLS model developed from robusta coffee and corn adulterant mixtures. The model demonstrated high apparent internal performance under the present experimental conditions ($R^2 > 0.999$), with relatively low calibration and prediction errors. However, these results should be interpreted with caution due to the limited number of samples and calibration design, which may increase the risk of overfitting. The estimates are therefore specific to the studied mixtures and experimental conditions. Predictions from the PLS model indicate that robusta coffee content in the sampled products ranges from approximately 71.7% to 98.2%, reflecting variability in the composition of the analyzed samples. The developed model is intended for quantifying corn adulteration in coffee samples and demonstrates the potential of FTIR-chemometric approaches for rapid coffee authentication.

1. Introduction

Indonesia is an agrarian country rich in natural resource potential, particularly in the agricultural sector. This sector consists of five sub-sectors: food crops, plantations, livestock, forestry, and fisheries. Among these, the plantation sub-sector thrives, as most plantation crops flourish in tropical climates such as those found in Indonesia. Coffee is one of the key plantation products of the country. According to the Directorate General of Plantations, Indonesia ranks fourth globally as a coffee producer, behind Brazil,

Vietnam, and Colombia, with a total coffee plantation area of approximately 1.262 million hectares. The majority of this area (98.14%) is comprised of smallholder plantations, while large estates account for only 1.86% [1]. Among the coffee types cultivated, robusta coffee dominates, covering 79.36% or approximately 968.88 thousand hectares, while arabica coffee accounts for 20.64%, with an average area of 251.94 thousand hectares. Coffee plantations in Indonesia are primarily located on the island of Sumatra (approximately 60%), while the remaining plantations are distributed across Java (14%), Sulawesi (12%), Nusa Tenggara (10%), and Kalimantan (3%). Specifically, in Sumatra, Lampung Province is the second-largest coffee producer, yielding 117,311 tons annually [2].

Coffee (*Coffea* sp.) is a flagship plantation crop that significantly contributes to the country's foreign exchange earnings [3]. With its high economic value, coffee is a vital source of income for over 1.5 million coffee farmers in Indonesia [4]. Moreover, coffee is one of the most popular beverages worldwide and among the most widely traded agricultural commodities.

Perceptions of coffee quality vary from producers to consumers. At the exporter and importer levels, coffee quality is influenced by several factors, including bean size, defect count, regulations, product availability, characteristics, and pricing. In coffee powder processing, quality is determined by aspects such as moisture content, stability of characteristics, origin, price, biochemical components, and flavor quality. At the consumer level, coffee choices are influenced by price considerations, aroma, taste preferences, health impacts, and environmental and social concerns [5]. Purity is a key determinant of coffee quality. Especially during imports, effective monitoring of supply is challenging, often relying on taste tests as indicators of product quality. Standards for maximum acceptable carbohydrate content have been established through ISO analysis and national regulations in England (BSCPIA/BSCMA 1995) and France (Syndicat Français des Fabricants de Café Soluble 1999) [6].

Currently, food fraud, particularly in coffee commodities, is on the rise in Indonesia, resulting in adverse effects for consumers [7]. Visually detecting mixed materials in ground coffee is difficult, particularly when adulterants have similar color and particle size. Various substances have been reported as coffee adulterants, including corn, rice, maize, soybeans, chicory, barley, and coffee husks, which are often used to reduce production costs [8]. In this study, corn and rice were selected as representative adulterants due to their widespread availability, relatively low cost, and physical characteristics that resemble ground coffee, making them suitable for intentional mixing. In addition, both commodities are commonly produced in the Lampung region, providing a practical basis for their potential use in local coffee adulteration practices. The economic incentive to substitute or partially replace coffee with these materials is driven by the price difference between coffee and these agricultural products, which can lead to increased profit margins for unscrupulous producers. Therefore, the development of a rapid and reliable method for detecting such adulteration is essential to ensure product authenticity and protect consumers.

The choice of analytical methods for identifying, discriminating, and authenticating the purity of natural products focuses on the chemical components present in these materials. Several laboratory techniques, including High-Performance Liquid Chromatography (HPLC), Gas Chromatography (GC), Thin Layer Chromatography (TLC), and various spectroscopic techniques (UV-Vis, FTIR, and NMR), combined with chemometric methods, are available for this purpose [9].

The use of Fourier Transform Infrared (FTIR) spectroscopy as an analytical technique is highly advantageous due to its efficiency, speed, and cost-effectiveness [10, 11]. FTIR spectra can differentiate among various materials, even when additional substances, such as corn and rice powders, are mixed with coffee powder. The FTIR spectrum can effectively distinguish between pure coffee and coffee mixed with additives, even when the exact chemical composition is unknown. However, the complexity of FTIR spectra can pose challenges for direct interpretation and visualization, making multivariate analysis techniques, such as chemometrics, essential for simplifying this process [12]. Multivariate analysis is a statistical approach used to understand data structures in high-dimensional spaces [13].

In cases of coffee adulteration, rice powder and corn powder are frequently used as fillers because of their low cost, availability, and visual similarity to ground coffee [14]. The presence of such adulterants can compromise the chemical composition, taste, and integrity of coffee products. To address this issue, this study employs FTIR spectroscopy combined with chemometric techniques, specifically Principal Component Analysis (PCA) and Partial Least Squares (PLS), to quantitatively detect and differentiate adulterated coffee samples. This approach aims to support efforts in safeguarding product quality and consumer trust through accurate, non-destructive analytical methods.

This research focuses on identifying both the type of coffee (robusta or arabica) and detecting potential adulterants in the coffee samples available in the Bandar Lampung market. Robusta and arabica are the two most widely cultivated coffee varieties in Indonesia, with robusta accounting for nearly 80% of total production [15]. Ensuring the authenticity of these coffee types is essential, given their differing quality attributes and market values [16].

In this study, PCA was employed as an exploratory tool to classify coffee samples and identify potential adulterants. This approach provides a basis for selecting the most relevant coffee type and adulterant for further quantitative analysis. Therefore, this study primarily focuses on the analysis of robusta coffee adulterated with corn using a PLS regression model, while arabica coffee and rice are included as reference materials for comparative and exploratory purposes.

2. Experimental

This research applied FTIR spectroscopy combined with chemometric analysis to evaluate the authenticity of coffee products. The methodology included preparing reference standards, collecting market samples, acquiring spectra, and performing multivariate data analysis using PCA and PLS. Model development and validation were conducted to assess the accuracy and reliability of the proposed approach. The experimental procedures and analytical methods are described in detail in the following subsections.

2.1. Materials and Tools

The materials used in this study consisted of arabica coffee beans, robusta coffee beans (purchased from certified coffee bean traders), rice, corn, distilled water, and five commercial coffee powder samples obtained from the Bandar Lampung market.

It should be noted that the number of samples used in this study was limited (n = 5). In standard chemometric practice, a larger sample size is generally required to develop a robust and statistically reliable calibration model. Therefore, the calibration model presented in this work should be regarded as preliminary and exploratory in nature.

The instruments and equipment used in this study included bowls, a blender, baking trays, sieves, a mortar and pestle, spatulas, an analytical balance, plastic bottles, scissors, labels, an FTIR spectrophotometer, a personal computer, and Minitab 19 software.

2.2. Preparation of Pure Coffee Standards

The preparation of pure coffee samples involved selecting medium-roasted arabica and robusta coffee beans. The beans were then ground into powder [17].

2.3. Preparation for Coffee Adulterants

The preparation of adulterants was carried out using corn and rice. Corn powder was prepared from husked dry corn through milling followed by sieving. Rice powder was prepared by washing the rice, followed by milling and sun-drying; after drying, the powder was sieved [18]. To ensure comparable thermal treatment and minimize spectral differences resulting from processing conditions, both rice and corn standards were roasted prior to grinding under the same conditions as the coffee standard (210°C for 10 minutes) [19].

2.4. Preparation of Calibration and Validation Sets

The calibration and validation sets were prepared by mixing arabica or robusta coffee with corn powder or rice powder at predetermined concentrations. The validation set was prepared independently of the calibration set using mixtures of known composition. These validation samples were not selected based on PCA results but were specifically designed to evaluate the predictive performance and generalization capability of the PLS model. This procedure was carried out to assess the ability of the model to identify coffee powder and detect adulterants in commercial market samples.

A calibration set consists of samples with known categories or compositions that are used to construct a chemometric classification model [20]. In this study, the calibration set included concentration levels of 0, 20, 40, 60, 80, and 100%. The 0% concentration corresponded to pure roasted corn or rice powder, whereas the 100% concentration corresponded to pure arabica or robusta coffee. The compositions used in the calibration set are presented in Table 1.

The calibration model was developed using six mixtures with varying composition ratios. However, it should be noted that the number of samples used in this study was relatively limited for chemometric modeling, potentially reducing the robustness and generalizability of the resulting PLS model. Therefore, the developed model should be considered preliminary.

Table 1 presents the proportions of arabica or robusta coffee mixed with roasted corn or rice powder at various concentrations, ranging from 100% pure coffee to 100% pure roasted adulterant, used to prepare the calibration set. Each concentration level was prepared in duplicate. Calibration and validation sets were constructed using independently prepared mixtures to avoid pseudo-replication and ensure model robustness.

Table 1. Mixture composition used for the calibration set

Arabica/robusta coffee (mg)	Corn/rice powder (mg)	Concentration (%)	Total mass (mg)	Category
100	0	100	100	Pure coffee
80	20	80	100	Coffee mixture
60	40	60	100	Coffee mixture
40	60	40	100	Coffee mixture
20	80	20	100	Coffee mixture
0	100	0	100	Pure corn/rice powder

2.5. Fourier Transform Infrared (FTIR) Spectrum Measurement

The calibration and validation sets were prepared by mixing corn or rice powder with arabica or robusta coffee at predetermined composition levels. These sample sets were used to evaluate the ability of the chemometric model to identify coffee powder adulteration in commercial market samples.

All samples were analyzed using an Agilent Cary 630 FTIR Spectrophotometer equipped with a diamond attenuated total reflectance (ATR) crystal. Powdered samples were directly placed onto the crystal surface and measured under consistent pressure to ensure reproducible contact. Spectral acquisition and preprocessing were performed using the instrument software, including ATR correction and pathlength normalization, to improve spectral comparability among samples. FTIR spectra were recorded in the range of 4000–650 cm^{-1} with a spectral resolution of 4 cm^{-1}. Each spectrum was obtained by averaging 32 co-added scans to improve the signal-to-noise ratio, while background spectra were collected under identical conditions prior to sample analysis. Each sample was measured in duplicate [9].

Prior to analysis, all samples were stored under controlled conditions to minimize moisture variation, since moisture can significantly influence FTIR spectra, particularly in the O–H stretching region.

2.6. Chemometric Analysis

Prior to PLS modeling, autoscaling was applied to the spectral data to ensure that all variables contributed equally to the analysis. PLS regression was then performed using the nonlinear iterative partial least squares (NIPALS) algorithm.

The optimal number of latent variables (LVs) was determined based on the minimum root mean square error of cross-validation (RMSECV) value. Based on this criterion, two latent variables were selected for the final model to minimize overfitting while maintaining model robustness. Cross-validation was carried out using the leave-one-out method.

An independent validation set of separately prepared mixtures was used to evaluate the predictive performance of the model, thereby minimizing the risk of pseudo-replication and overfitting. This approach is consistent with commonly recommended practices in chemometric modeling. Model performance was evaluated using RMSECV, RMSEC, and RMSEP, along with residual analysis to assess prediction errors.

3. Results and Discussion

This section presents the results of FTIR spectroscopy combined with chemometric analysis for the identification of adulteration in coffee products. PCA was first employed to evaluate spectral variations and clustering patterns among samples, followed by PLS regression for quantitative prediction of coffee purity and adulterant concentration. The performance of the developed models was further evaluated to assess their predictive capability and potential applicability for coffee authentication.

3.1. Sampling Locations

This research aimed to assess the authenticity of coffee products available in the Bandar Lampung City market, located in the southern region of Sumatra Island, Indonesia. The study involved several key steps: preparing standard samples of arabica coffee, robusta coffee, roasted rice powder, and roasted corn powder; collecting commercial coffee powder samples from various local markets in Bandar Lampung (5.4294° S, 105.2625° E); and performing qualitative and quantitative analysis using FTIR spectroscopy. All standard materials, including arabica and robusta coffee beans, rice, and corn, were obtained from certified coffee shops and traditional markets within the city.

The commercial coffee samples were randomly selected from different locations, as illustrated in Figure 1. The relatively small number of samples used in this study represents a significant limitation. In chemometric modeling, a larger dataset (typically more than 20 mixtures) is generally recommended to ensure statistical validity and to minimize the risk of overfitting. FTIR spectra were recorded for all samples, and the resulting data were analyzed using Minitab 19 software to develop and validate the chemometric models.

3.2. FTIR Spectral Analysis of Coffee and Adulterants

FTIR spectroscopy was employed to analyze the pure standards of arabica coffee, robusta coffee, roasted rice powder, and roasted corn powder. The resulting IR spectra, shown in Figure 2, reveal several distinctive peaks that enable differentiation between coffee and potential adulterants based on their chemical composition. A prominent peak at 3272 cm^{-1}, observed with stronger intensity in rice and corn powders than in coffee samples, is associated with O–H stretching vibrations, likely due to the higher starch content in these adulterants. In contrast, a characteristic peak at 3008 cm^{-1}, attributed to C=C stretching in cis double bonds from lipid compounds, appears only in arabica and robusta coffee, highlighting the lipid-rich profile unique to coffee.

Figure 1. Sampling locations in Bandar Lampung, Indonesia. Base map source: Badan Informasi Geospasial (BIG) [21]

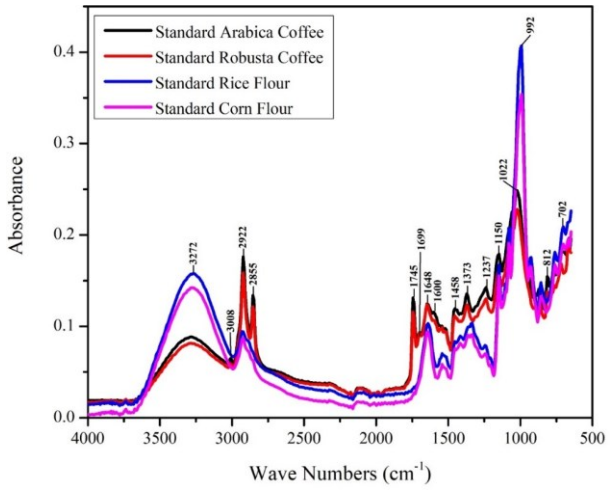


Figure 2. FTIR spectra of coffee and adulterant standards

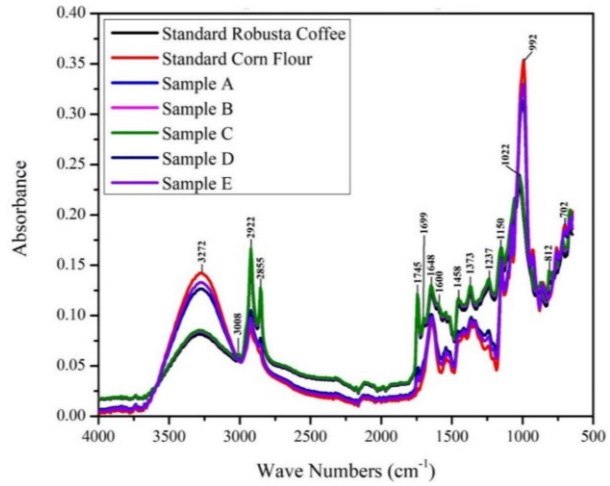


Figure 3. FTIR spectra of robusta coffee, corn flour, and samples A–E

In the 2922 cm^{-1} region, all four samples exhibit peaks; however, the absorbance is notably higher in coffee, suggesting a greater concentration of aliphatic C– H groups. Similarly, at 2855 cm^{-1}, absorbance bands appear exclusively in the coffee samples, corresponding to symmetric and asymmetric C–H stretching in methyl and methylene groups, primarily found in lignin, a structural polymer abundant in coffee beans.

A series of carbonyl-related peaks further distinguishes coffee from the adulterants. The band at approximately 1745 cm^{-1} is associated with C=O stretching vibrations, although this region is complex and may include contributions from lipids, chlorogenic acid derivatives, caffeine-related compounds, and other constituents. Therefore, assigning this peak to a single compound, such as quinic acid, should be interpreted cautiously [22]. Peaks at 1699 cm^{-1} and 1648 cm^{-1} are also stronger in coffee and correspond to C=O vibrations from free fatty acids and other carbonyl-containing compounds. The 1600 cm^{-1} peak likely reflects contributions from secondary metabolites such as caffeine, trigonelline, and nicotinic acid, which are characteristic of coffee.

The region from 1458 to 1373 cm^{-1} exhibits more intense bands in coffee, attributed to aliphatic C–H bending vibrations. Additional peaks at 1237, 1150, and 1022 cm^{-1}, observed only in coffee samples, are linked to C–O stretching vibrations in esters and polysaccharides such as chlorogenic acid and lignin. In contrast, rice and corn powders exhibit higher absorbance in the 992–702 cm^{-1} region, which is commonly associated with glycosidic C–O–C stretching vibrations in amylose and amylopectin, the primary components of starch.

The band observed at approximately 812 cm^{-1} shows relatively higher absorbance in coffee samples and may be attributed to out-of-plane C–H bending vibrations of alkenes. However, this band is not uniquely specific and should be considered as a supporting spectral marker, possibly related to unsaturated components such as fatty acids present in coffee [23].

3.3. Characterization of Coffee Samples Using FTIR Spectroscopy

The FTIR spectra of the coffee samples (Figure 3) revealed peaks at wavenumbers consistent with those of the coffee and adulterant reference standards. Samples A, B, and C showed strong absorbance at 3008 cm^{-1} and 1237 cm^{-1}, characteristic of robusta coffee, indicating close similarity to the pure robusta reference. In contrast, samples D and E exhibited prominent peaks at 3272 cm^{-1} and 992 cm^{-1}, corresponding to starch-rich materials such as corn flour, suggesting possible adulteration. While the peak positions across the samples were generally similar, differences in absorbance intensity indicate variations in the concentration of specific components. These intensity differences provide insight into the relative purity of the samples and the potential presence of adulterants.

3.4. Chemometric Analysis Using Principal Component Analysis (PCA)

PCA is a multivariate technique that reduces a dataset by transforming correlated variables into a smaller number of uncorrelated variables, known as principal components (PCs), which capture the maximum variance [24]. The first principal components (PC1 and PC2) account for the largest proportion of variance in the data and thus reflect the main underlying structure. Higher-order components, on the other hand, tend to capture less systematic variation and may represent noise [25].

In this study, PCA was employed as a qualitative tool to examine clustering patterns of the coffee samples relative to four reference standards: arabica coffee, robusta coffee, rice powder, and corn powder. The dominant standard component was later used to construct calibration and validation sets for quantifying adulteration levels.

FTIR spectra in the 4000–650 cm^{-1} region served as the basis for PCA, with specific wavenumbers selected for their diagnostic peaks. Transmittance values were converted to absorbance, and both single and duplicate measurements were included to enhance data reliability. The resulting eigenvalues, summarized in Table 2, illustrate the proportion of variance explained by each PC and support the discrimination between pure and adulterated coffee samples.

Table 2. PCA eigenvalues and explained variance

Component	Eigenvalue	Proportion	Cumulative
PC1	8.0355	0.670	0.670
PC2	2.8432	0.237	0.907
PC3	0.7720	0.064	0.971
PC4	0.2439	0.020	0.991
PC5	0.0488	0.004	0.995
PC6	0.0259	0.002	0.997
PC7	0.0201	0.002	0.999
PC8	0.0068	0.001	1.000
PC9	0.0029	0.000	1.000
PC10	0.0008	0.000	1.000
PC11	0.0000	0.000	1.000
PC12	0.0000	0.000	1.000

Eigenvalues reflect the proportion of variance explained by each PC in a dataset [26]. A commonly used criterion considers PCs with eigenvalues ≥ 1 as significant contributors to data variation, while those < 1 are considered less influential [27].

As shown in Table 2, the first principal component (PC1) has the highest eigenvalue of 8.0355, accounting for 67.0% of the total variance. The second component (PC2) explains an additional 23.7%, with an eigenvalue of 2.8432. Together, PC1 and PC2 capture 90.7% of the total variation, which is sufficient for meaningful data interpretation and dimensionality reduction.

To determine the optimal number of PCs to retain, a scree plot is used (Figure 4). This plot displays the eigenvalues against the component numbers, helping to visualize the point at which additional components contribute minimally to explaining the variance. Based on the scree plot, PC3 has an eigenvalue of 0.7720, which is below the threshold of 1, indicating limited contribution to data variation and reduced reliability for interpretation. Therefore, only PC1 and PC2, both with eigenvalues greater than 1, were selected. This reduction suggests that the original 12 variables can be effectively represented in two dimensions, as illustrated in the score plot (Figure 5).

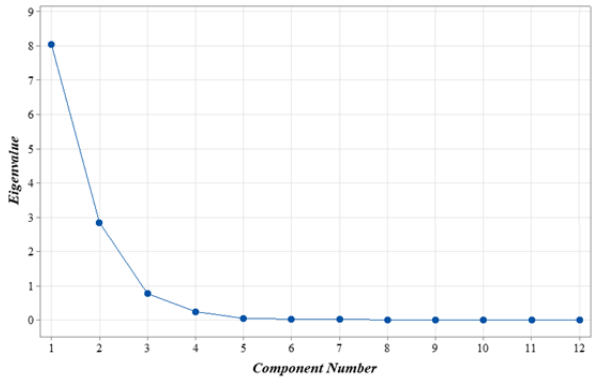


Figure 4. Scree plot of PCA eigenvalues from selected wavenumber peaks

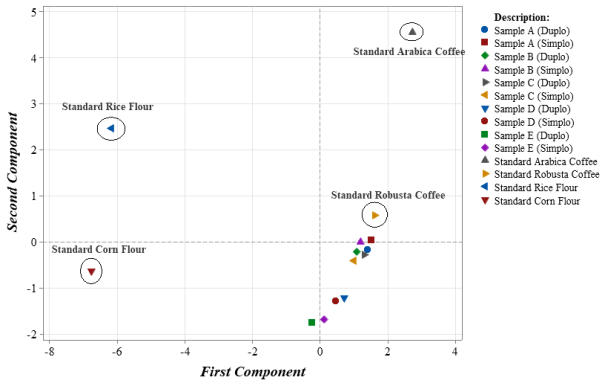


Figure 5. PCA score plot of coffee and adulterant standards and samples A–E

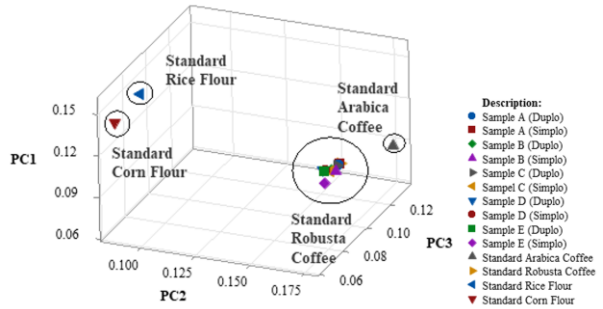


Figure 6. 3-dimensional PCA plot

The PCA score plot visualizes sample grouping based on variance along PC1 and PC2, which accounted for the largest proportion of total variance [26]. As shown in Figure 5, samples A, B, and C (including their duplicates) clustered closely with the robusta coffee standard, suggesting strong compositional similarity. In contrast, samples D and E appear more distant from the robusta standard, with the duplicate of sample E positioned nearer to the corn powder standard, indicating possible adulteration. The proximity of sample E to the corn reference in the PCA space may indicate similarities in spectral characteristics. Nevertheless, this observation should be interpreted cautiously, as no complementary analytical methods were used to confirm adulteration.

The plot clearly shows how samples and standards are grouped by spectral similarity. The closer the position of points, the higher the similarity between the sample and reference [28]. A 3-dimensional PCA plot (Figure 6) provides additional visualization, especially for the differentiation between coffee and corn powder standards.

3.5. Chemometric Analysis Using Partial Least Squares (PLS)

Based on Figures 7 and 8, differences in absorbance at observed peaks are noted, particularly at wavenumbers 3008 and 1237 cm^{-1}. As the concentration increases, the resulting spectrum becomes closer to that of robusta coffee. Conversely, at wavenumbers 3272 and 992 cm^{-1}, lower concentrations result in a spectrum resembling that of corn powder. The variations in absorbance and the resulting spectrum are likely influenced by the presence and proportion of the mixture of robusta coffee and corn powder concentrations.

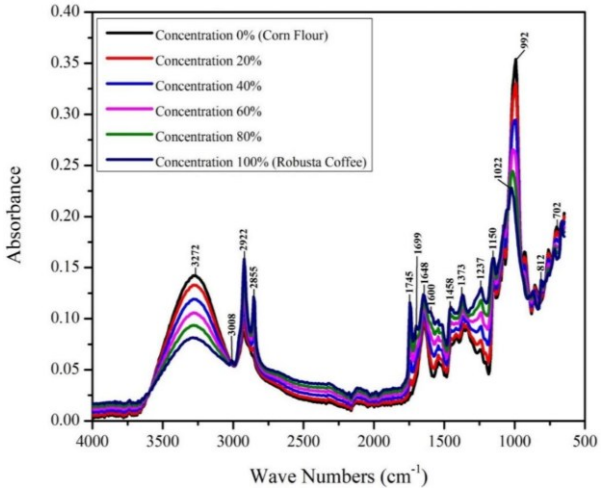


Figure 7. FTIR spectra of the calibration set

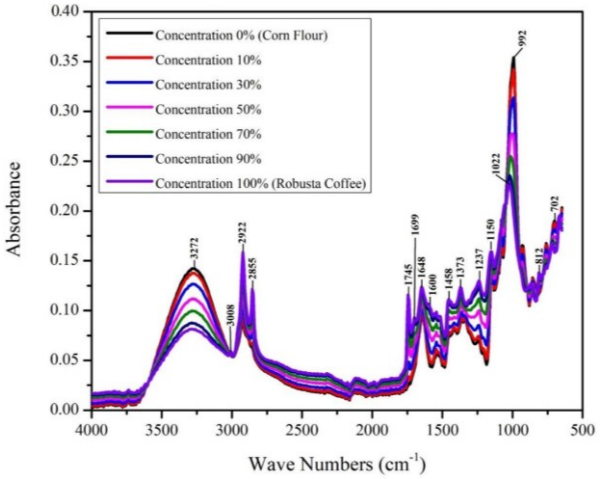


Figure 8. FTIR spectra of the validation set

Data from the calibration and validation sets are processed using the PLS method. This method is typically represented in a graph showing the correlation between actual concentration (x-axis) and predicted concentration (y-axis), obtained through a combination of FTIR spectroscopy and chemometric analysis [29]. The aim of using the PLS model is to mitigate the impact of numerous irrelevant predictors on data variability, thereby improving predictive accuracy and model performance [30]. The effectiveness of the PLS multivariate calibration technique is assessed using parameters like R-Square (coefficient of determination) and Root Mean Square Error (RMSE) [31]. An R-Square value closer to 1 indicates a stronger relationship between the predictors and the response, whereas a lower RMSE value indicates reduced prediction error and better model quality [32].

3.6. Development of the PLS Calibration Model

The calibration set was used to develop a predictive model that establishes correlations between the spectral data (predictors) and adulterant concentration (response variable). PLS regression produced a high correlation coefficient (R) of 0.999981 and a Root Mean Square Error of Calibration (RMSEC) of 0.14%. The PLS model showed a very high degree of fit to the calibration data, with low calibration error. However, these results should be interpreted with caution, as the limited number of samples may increase the risk of overfitting and lead to an overestimation of model performance. The extremely high R^2 values and low RMSEC observed in this study are likely influenced by the small size of the calibration dataset. In practical spectroscopic applications, such near-perfect linearity is uncommon and may indicate that the model is overfitted to the available data. Consequently, the reported performance primarily reflects the model's ability to fit the calibration dataset rather than its predictive capability for independent samples. The very low RMSEC demonstrates minimal error during calibration, confirming the model's reliability in quantifying adulteration levels [33]. The calibration analysis curve is shown in Figure 9.

3.7. Evaluation of Calibration Model Using Validation Set with PLS

The validation set was employed to assess the predictive accuracy of the calibration model. Using PLS regression, the validation results showed a high correlation coefficient (R) of 0.999926 and a Root Mean Square Error of Prediction (RMSEP) of 0.66%. An R^2 value close to 1 indicates that the model effectively explains most of the variation in the response variable, while the low RMSEP reflects minimal error in predicting adulterant concentrations during validation [34]. The validation curve is presented in Figure 10.

The PLS plots (Figures 9 and 10) show a strong linear relationship between predicted and actual values. However, this apparent agreement should be interpreted with caution, as the limited number of samples may contribute to an overfitting effect, resulting in an artificially high degree of linearity.

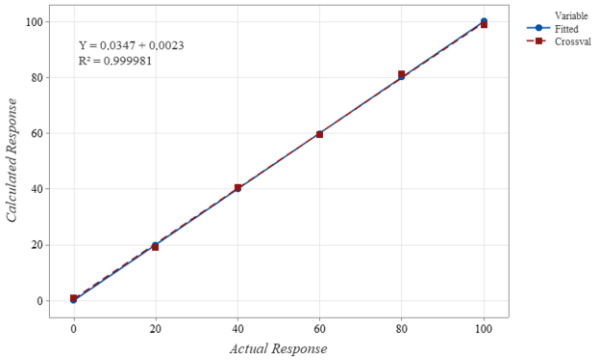


Figure 9. PLS response plot (calibration set)

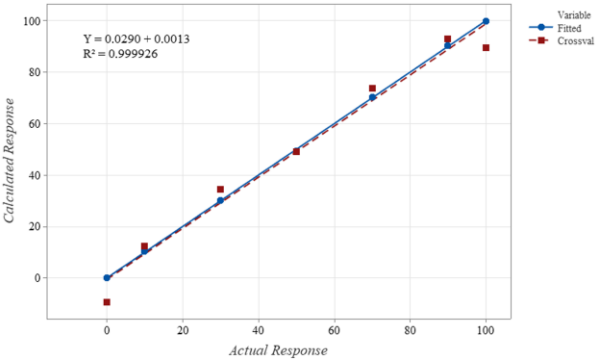


Figure 10. PLS response plot (validation set)

The limited number of calibration samples (n = 6) represents a significant limitation of this study. In chemometric modeling, such a small dataset increases the likelihood of overfitting, where the model closely fits the calibration data but may not generalize well to new samples. This may explain the extremely high R^2 values obtained in this study. Therefore, the reported model performance should be interpreted with caution, as it may overestimate the true predictive capability of the model.

In addition to peak identification, the quantitative analysis is based on variations in absorbance intensity across the FTIR spectra. Changes in absorbance at specific wavenumber regions, particularly those associated with carbonyl and carbohydrate-related vibrations, contribute to the differentiation between pure coffee and adulterated samples. These spectral variations are captured by the PLS model, which establishes a multivariate relationship between spectral intensity and adulterant concentration.

To further evaluate model performance, the PLS model was assessed using cross-validation, yielding a low RMSECV value, which is consistent with the low RMSEC and RMSEP values obtained. Residual analysis showed no systematic trend when plotted against the reference concentration, indicating that the model errors were randomly distributed.

Cross-validation was conducted using a leave-one-out approach, which provides an internal safeguard against overfitting, particularly for small datasets. The selection of two latent variables based on the minimum RMSECV further supports the robustness of the model.

However, it should be noted that the calibration design, consisting of six concentration levels (0–100% at 20% intervals), may inherently favor smooth interpolation. Therefore, the high predictive performance observed in this study reflects the controlled experimental conditions and limited compositional space. Accordingly, the precision of predicted values should not be overinterpreted, and results are reported to one decimal place, consistent with the RMSEP value (≈ 0.66%).

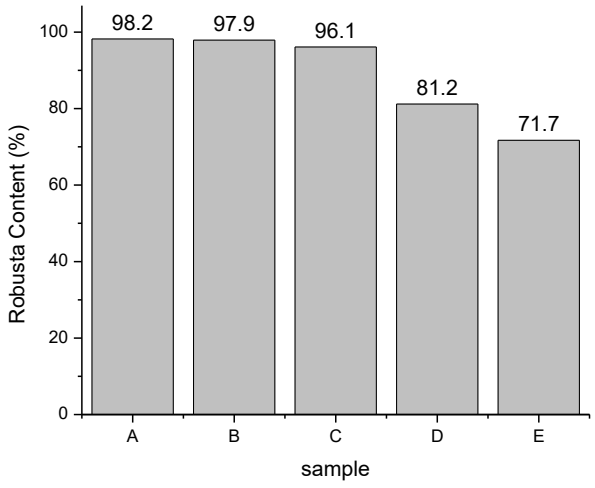


Figure 11. Bar chart of estimated robusta content in coffee samples

Figure 11 presents the robusta content of individual samples using a bar chart, providing a direct comparison of the estimated values across the analyzed samples. These values were derived from the PLS model, which establishes a multivariate relationship between FTIR spectral intensity and the concentration of reference standards. The predicted robusta content ranged from approximately 71.7% to 98.2%, with samples A–C showing higher values and samples D and E exhibiting lower values, indicating greater compositional variability.

It should be noted that these estimates reflect the similarity of the samples to the robusta–corn mixtures defined in the PLS model, as rice was not included in the quantitative analysis. Furthermore, the qualitative trend observed in the PCA score plot, where samples D and E are positioned closer to the corn reference, is consistent with the lower robusta content predicted by the PLS model, thereby supporting the agreement between the exploratory and quantitative analyses.

4. Conclusion

This study demonstrates that FTIR spectroscopy combined with chemometric techniques can be applied as a rapid approach for evaluating the purity of coffee samples. PCA results indicated that the main variation in the dataset could be effectively described by two principal components. The samples were generally grouped near the robusta reference, although some samples showed spectral similarities to corn, suggesting possible compositional differences that should be interpreted cautiously. The estimated robusta content in coffee samples from the Bandar Lampung market ranged from approximately 71.7% to 98.2%, indicating variability in sample composition. However, given the limited number of calibration samples used in this study, the developed PLS model may be prone to overfitting and should be considered preliminary. Overall, the proposed method shows promising potential as a rapid screening tool for coffee authentication. Further studies involving a larger and more diverse dataset, along with additional validation methods, are required to improve the reliability and general applicability of the model.

Acknowledgement

The authors gratefully acknowledge the financial support provided by Universitas Lampung through the BLU (Badan Layanan Umum) research funding scheme in 2024, under Contract No. 901/UN26.21/PN/2024 dated 6 August 2024. This support was instrumental in the successful completion of this study.

References

[1] Jongkers Tampubolon, Albina Ginting, Hotden L. Nainggolan, Juli Ritha Tarigan, Indonesian Coffee Development Path: Production and International Trade, *Asian Journal of Agricultural Extension, Economics & Sociology*, 41, 12, (2023), 316-328 https://doi.org/10.9734/ajaees/2023/v41i122335

[2] Anggita Tresliyana Suryana, Harianto Harianto, Yusman Syaukat, Harmini Harmini, Geographic scope of the global value chain for Indonesian coffee,

Coffee Science, 18, (2023), e182169 https://doi.org/10.25186/.v18i.2169

[3] Rita Andini, Murna Muzaifa, Leni Marlina, Muhammad Ikhsan Sulaiman, Rachman Jaya, Ali M. Muslih, Heru P. Widayat, Making biodiversity work for coffee production. A case study of Gayo Arabica coffee in Indonesia, *MOJ Ecology & Environmental Sciences*, 6, 4, (2021), 156-162 https://doi.org/10.15406/mojes.2021.06.00228

[4] Zsuzsanna Bacsi, Mária Fekete-Farkas, Muhammad Imam Ma'ruf, Coffee Yield Stability as a Factor of Food Security, *Foods*, 11, 19, (2022), 3036 https://doi.org/10.3390/foods11193036

[5] Carla Martino Bemfeito, Angélica Sousa Guimarães, Alberto Lima de Oliveira, Bruna Fernandes Andrade, Luiza Maria Amaral Frossard de Paula, Carlos José Pimenta, Do consumers perceive sensory differences by knowing information about coffee quality?, *LWT - Food Science and Technology*, 138, (2021), 110778 https://doi.org/10.1016/j.lwt.2020.110778

[6] Valdeir Viana Freitas, Larissa Lorrane Rodrigues Borges, Márcia Cristina Teixeira Ribeiro Vidigal, Marcelo Henrique dos Santos, Paulo Cesar Stringheta, Coffee: A comprehensive overview of origin, market, and the quality process, *Trends in Food Science & Technology*, 146, (2024), 104411 https://doi.org/10.1016/j.tifs.2024.104411

[7] Meinilwita Yulia, Diding Suhandy, Quantification of Corn Adulteration in Wet and Dry-Processed Peaberry Ground Roasted Coffees by UV–Vis Spectroscopy and Chemometrics, *Molecules*, 26, 20, (2021), 6091 https://doi.org/10.3390/molecules26206091

[8] Jing-Ya Xie, Jin Tan, Front-face synchronous fluorescence spectroscopy: a rapid and non-destructive authentication method for Arabica coffee adulterated with maize and soybean flours, *Journal of Consumer Protection and Food Safety*, 17, (2022), 209-219 https://doi.org/10.1007/s00003-022-01396-8

[9] Deni Agus Triawan, Devi Ratnawati, Ria Nurwidiyani, Doni Notriawan, Morina Adfa, Rekianto Setiawan Bahar, Mohamad Rafi, Data fusion of FTIR and UV–Vis spectra combined with chemometrics to classification garden balsam leaves (*Impatiens balsamina* L.) extract from two different location in Bengkulu Province, *Vegetos*, (2024), https://doi.org/10.1007/s42535-024-00960-x

[10] Hemerson D. dos Santos, Elisangela F. Boffo, Coffee beyond the cup: analytical techniques used in chemical composition research—a review, *European Food Research and Technology*, 247, (2021), 749-775 https://doi.org/10.1007/s00217-020-03679-6

[11] 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

[12] Reema Valand, Sangeeta Tanna, Graham Lawson, Linda Bengtström, A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations, *Food Additives & Contaminants: Part A*, 37, 1, (2020), 19-38 https://doi.org/10.1080/19440049.2019.1675909

[13] Diding Suhandy, Meinilwita Yulia, Classification of *Lampung robusta* Specialty Coffee According to Differences in Cherry Processing Methods Using UV Spectroscopy and Chemometrics, *Agriculture*, 11, 2, (2021), 109 https://doi.org/10.3390/agriculture11020109

[14] Wilma C. Josefa, Daniel F. Cipriano, Aline T. Toci, Jair C. C. Freitas, Solid-state NMR investigation on the presence of adulterants in coffee samples, *Food Research International*, 206, (2025), 116054 https://doi.org/10.1016/j.foodres.2025.116054

[15] Fitrio Ashardiono, Agus Trihartono, Optimizing the potential of Indonesian coffee: a dual market approach, *Cogent Social Sciences*, 10, 1, (2024), 2340206 https://doi.org/10.1080/23311886.2024.2340206

[16] Woranitta Sahachairungrueng, Chanyanuch Meechan, Nutchaya Veerachat, Anthony Keith Thompson, Sontisuk Teerachaichayut, Assessing the Levels of Robusta and Arabica in Roasted Ground Coffee Using NIR Hyperspectral Imaging and FTIR Spectroscopy, *Foods*, 11, 19, (2022), 3122 https://doi.org/10.3390/foods11193122

[17] Deni Agus Triawan, Ghufira Ghufira, Morina Adfa, Mohamad Rafi, Pencirian Kopi Robusta Bengkulu dari Kemungkinan Bahan Pencampur Menggunakan Kombinasi Spektroskopi FTIR dan Kemometrik, *agriTECH*, 42, 4, (2022), 321-328 https://doi.org/10.22146/agritech.64432

[18] Cinthia de Carvalho Couto, Otniel Freitas-Silva, Edna Maria Morais Oliveira, Clara Sousa, Susana Casal, Near-Infrared Spectroscopy Applied to the Detection of Multiple Adulterants in Roasted and Ground Arabica Coffee, *Foods*, 11, 1, (2022), 61 https://doi.org/10.3390/foods11010061

[19] Elin Yusibani, Ikramullah, Evi Yufita, Zulkarnain Jalil, Endi Suhendi, The Effect of Temperature and Roasting Time on The Physical Properties of Arabica and Robusta Gayo Coffee Bean, *Journal of Applied Agricultural Science and Technology*, 7, 2, (2023), 100-108 https://doi.org/10.55043/jaast.v7i2.75

[20] Hai-Peng Wang, Pu Chen, Jia-Wei Dai, Dan Liu, Jing-Yan Li, Yu-Peng Xu, Xiao-Li Chu, Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues, *TrAC Trends in Analytical Chemistry*, 153, (2022), 116648 https://doi.org/10.1016/j.trac.2022.116648

[21] Geospasial Badan Informasi, in, Badan Informasi Geospasial, 2024,

[22] Yegane Tarandakzad, Rasool Khodabakhshian, Mehdi Khojastehpour, Hadi Mahdavianmehr, FTIR spectroscopy identification of fraud in coffee powder: a study on preprocessing methods, *Food Chemistry*, 518, (2026), 148777 https://doi.org/10.1016/j.foodchem.2026.148777

[23] Mauricio Flores-Valdez, Ofelia Gabriela Meza-Márquez, Guillermo Osorio-Revilla, Tzayhri Gallardo-Velázquez, Identification and Quantification of Adulterants in Coffee (*Coffea arabica* L.) Using FT-MIR Spectroscopy Coupled with Chemometrics, *Foods*, 9, 7, (2020), 851 https://doi.org/10.3390/foods9070851

[24] Abdelkrim Rebiai, Hadia Hemmami, Soumeia Zeghoud, Bachir Ben Seghir, Imane Kouadri, Laouini Salah Eddine, Noureddine Elboughdiri, Saad Ghareba, Djamel Ghernaout, Nadir Abbas, Current Application of Chemometrics Analysis in Authentication of Natural Products: A Review, *Combinatorial Chemistry & High Throughput Screening*, 25, 6, (2022), 945-972 https://doi.org/10.2174/1386207324666210309102239

[25] Oxana Rodionova, Sergey Kucheryavskiy, Alexey Pomerantsev, Efficient tools for principal component analysis of complex data— a tutorial, *Chemometrics and Intelligent Laboratory Systems*, 213, (2021), 104304 https://doi.org/10.1016/j.chemolab.2021.104304

[26] J. Renwick Beattie, Francis W. L. Esmonde-White, Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra, *Applied Spectroscopy*, 75, 4, (2021), 361-375 https://doi.org/10.1177/0003702820987847

[27] Raúl González-Domínguez, Ana Sayago, Ángeles Fernández-Recamales, An Overview on the Application of Chemometrics Tools in Food Authenticity and Traceability, *Foods*, 11, 23, (2022), 3940 https://doi.org/10.3390/foods11233940

[28] James N. Miller, Jane C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 5th ed. ed., Pearson Education Limited, London, United Kingdom, 2005,

[29] Mitsalina Fildzah Arifah, Irnawati Irnawati, Ruslin Ruslin, Khoirun Nisa, Anjar Windarsih, Abdul Rohman, The Application of FTIR Spectroscopy and Chemometrics for the Authentication Analysis of Horse Milk, *International Journal of Food Science*, 2022, 1, (2022), 7643959 https://doi.org/10.1155/2022/7643959

[30] Xiaolei Zhang, Jie Yang, Advanced chemometrics toward robust spectral analysis for fruit quality evaluation, *Trends in Food Science & Technology*, 150, (2024), 104612 https://doi.org/10.1016/j.tifs.2024.104612

[31] Mina Sadrara, Mohammadreza Khanmohammadi Khorrami, Gasoline quality prediction using the alternating conditional expectation (ACE) algorithm and ATR-FTIR, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 310, (2024), 123903 https://doi.org/10.1016/j.saa.2024.123903

[32] Aimen El Orche, Casimir Adade Adade, Hafid Mefetah, Amine Cheikh, Khalid Karrouchi, Miloud El Karbane, Mustapha Bouatia, Chemometric Analysis of UV-Visible Spectral Fingerprints for the Discrimination and Quantification of Clinical Anthracycline Drug Preparation Used in Oncology, *BioMed Research International*, 2021, 1, (2021), 5580102 https://doi.org/10.1155/2021/5580102

[33] Keshav Kumar, Partial Least Square (PLS) Analysis: Most Favorite Tool in Chemometrics to Build a Calibration Model, *Resonance*, 26, (2021), 429-442 https://doi.org/10.1007/s12045-021-1140-1

[34] Rafael C. Castro, David S. M. Ribeiro, João L. M. Santos, Ricardo N. M. J. Páscoa, Authentication/discrimination, identification and quantification of cinnamon adulterants using NIR spectroscopy and different chemometric tools: A tutorial to deal with counterfeit samples, *Food Control*, 147, (2023), 109619 https://doi.org/10.1016/j.foodcont.2023.109619