International Journal of Horticultural Science and Technology
Journal homepage: https://ijhst.ut.ac.ir

Investigating the Possibility of Using Near Infrared Spectroscopy (NIRS) Method in Coffee Beans Caffeine and Chlorogenic Acid (CGA) Determination by Near Infrared Spectroscopy (NIRS)

Kusumiyati Kusumiyati[1]*, Agus Arip Munawar[2], Andasuryani Andasuryani[3]

1 Master Program of Agronomy, Faculty of Agriculture, Universitas Padjadjaran, Sumedang 45363, Indonesia

2 Department of Agricultural Engineering, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

3 Department of Agricultural Engineering and Biosystem, Andalas University, Limau Manis, Padang, 25163, Indonesia

ARTICLE INFO

Article history:

Received: 2 May 2024,
Received in revised form: 13 November 2024,
Accepted: 23 November 2024

Article type:

Research paper

Keywords:

Arabica,
Non-destructive,
Quality,
Real-time detection,
Robusta

*Corresponding author's email: kusumiyati@unpad.ac.id

ABSTRACT

The use of near-infrared spectroscopy (NIRS) to predict coffee bean quality has shown significant promise due to its high efficiency and precision. This study aimed to employ NIRS to estimate caffeine and chlorogenic acid (CGA) levels in coffee beans from two distinct geographical regions. While NIRS models are often tailored for specific products, expanding their applicability could enhance productivity. The study utilized multivariate analysis on 50 samples, comprising both Arabica and Robusta coffee varieties. Results for the caffeine prediction model included a coefficient of correlation in the calibration set (R_{cal}) of 0.85, root mean square error in the calibration set (RMSEC) of 0.30, coefficient of correlation in the cross-validation set (R_{cv}) of 0.82, root mean square error in the cross-validation set (RMSECV) of 0.31, and a ratio of prediction to deviation (RPD) of 2.21. For CGA, the model produced values of 0.88 (R_{cal}), 0.61 (RMSEC), 0.88 (R_{cv}), 0.65 (RMSECV), and 2.18 (RPD). Key wavelengths associated with caffeine and water were identified at 1122, 1452, 1682, and 1950 nm, while CGA showed strong correlations at 1415, 1718, and 1909 nm. The study concluded that the model's accuracy was satisfactory, highlighting the potential of NIRS as a viable alternative to traditional laboratory methods for predicting caffeine and CGA levels in coffee beans.

Introduction

Coffee (*Coffea* sp.) is a high-value agricultural commodity that thrives in tropical regions, particularly in Indonesia. The islands of Java and Sumatra offer ideal geographical conditions for cultivating coffee plants due to their favorable climate and soil. Belonging to the Rubiaceae family, the genus *Coffea* comprises around 120 species, with *Coffea arabica L.* (Arabica) and *Coffea canephora L.* (Robusta) being the primary species cultivated for commercial bean production (Shokouh et al., 2019). These two species play pivotal roles in the coffee industry due to their distinct characteristics. Arabica, known for its smooth taste, vibrant acidity, and aromatic profile, thrives in high-altitude areas with abundant rainfall and fertile soil (Saurnida et al., 2023). Conversely, Robusta, recognized for its strong, bitter flavor and higher caffeine content, is a popular choice for espresso blends. The flavor and quality of coffee are influenced by a variety of chemical components, including alkaloids, phenolic acids, flavonoids, terpenoids, sterols, caffeine, chlorogenic acid (CGA), antioxidants, moisture content, and volatile

compounds (Alamri et al., 2022; Gobbi et al., 2023; Saud and Salamatullah, 2021). Understanding and analyzing these components are crucial for assessing the quality and taste characteristics of coffee beans.

Caffeine, a key component of coffee beans, has a significant impact on their taste profile, contributing to the acidity and sharpness of flavor (Santosa et al., 2021; Tsegay et al., 2020). Rigorous monitoring of caffeine content is crucial for maintaining consistent coffee quality. Similarly, chlorogenic acid (CGA) plays an essential role in shaping coffee's unique aroma and flavor, making the precise measurement of both components vital for ensuring optimal coffee bean quality. The levels of caffeine and CGA are influenced by a combination of internal and external factors (Purwoko et al., 2022). Internal factors, such as the species and variety of the coffee plant (Girma et al., 2020), play a pivotal role in determining these components. External variables, including environmental conditions and cultivation practices, further influence their concentration (Purwoko et al., 2022). Caffeine primarily contributes to the bitterness of coffee, while CGA is responsible for its sour taste (Awaliyah et al., 2022). Interestingly, a higher caffeine concentration is generally associated with superior coffee quality, whereas increased CGA levels may indicate lower quality.

Traditionally, high-performance liquid chromatography (HPLC) has been widely used to measure caffeine and CGA content (Awwad et al., 2021; Ayelign and Sabally, 2013; Sa et al., 2023). While HPLC is a robust and accurate analytical technique, it has several limitations. The method requires extensive sample preparation, including extraction and dissolution in a solvent before injection into the HPLC column. This labor-intensive process, coupled with the high cost of laboratory equipment and reagents, can be time-consuming and may restrict accessibility for budget-constrained laboratories. Additionally, errors in sample preparation or equipment operation can significantly affect the accuracy of the results, necessitating a high level of expertise for proper operation and data interpretation. Given these challenges, technological advancements have introduced alternative methods for detecting caffeine and CGA. One increasingly popular approach is near-infrared spectroscopy (NIRS) (Goisser et al., 2020; Kusumiyati et al., 2018, 2019a, 2020, 2021a; Rodríguez et al., 2023; Vitale et al., 2013). This non-destructive technology enables the rapid and efficient measurement of caffeine and CGA levels by analyzing the interaction between near-infrared light and coffee bean molecules. The molecular interactions produce distinct spectral patterns that can be identified by NIRS instruments, offering a more accessible and streamlined alternative to traditional methods.

Near-infrared spectroscopy (NIRS) offers numerous advantages to the coffee industry, providing a reliable, efficient, and user-friendly method for quality assessment without requiring complex preparation or specialized expertise (Kusumiyati et al., 2022a). The technology allows simultaneous detection of multiple components, enabling a comprehensive evaluation of coffee bean quality in a fraction of the time. NIRS is particularly effective in preventing species contamination among coffee samples, a prevalent challenge in regions with diverse coffee cultivation practices. By doing so, it plays a crucial role in helping industry stakeholders maintain the quality and purity of coffee beans. Numerous studies have demonstrated the effectiveness of NIRS in identifying various components in coffee, including CGA content, moisture levels, caffeine, and color (Ayu et al., 2020; Grassi et al., 2021; Levate et al., 2021; Juliano et al., 2021; Scholz et al., 2014; Tugnolo et al., 2021; Yusmanizar et al., 2019). However, ongoing research is necessary to refine the spectral data for more precise predictions.

This study aimed to develop an accurate NIRS calibration model for measuring caffeine and CGA content in Arabica and Robusta coffee beans sourced from two geographical regions. A non-destructive approach was employed, showcasing the potential of NIRS to rapidly and efficiently evaluate coffee quality while preserving the integrity of the samples. To achieve this, the current research compared two regression methods—partial least squares regression (PLSR) and principal component regression (PCR). The optimal regression method was then selected to construct a calibration model for quantifying caffeine and CGA content in coffee beans. By integrating these methods, the research highlights the promise of NIRS as a practical and innovative tool for coffee quality assessment.

Materials and Methods

Coffee bean samples

Each coffee sample used in the study weighed 50 g. A total of 50 coffee bean samples, comprising both Arabica and Robusta species, were collected from Java (West Java and East Java) and Sumatra (West Sumatra, South Sumatra, and Lampung) Islands. The number of samples from each island was selected randomly to ensure diversity. The same set of samples was used for both calibration and cross-validation processes. These samples

were harvested, sorted, and prepared for spectral measurements and laboratory analysis. All procedures were conducted at the Horticulture Laboratory, Faculty of Agriculture, Universitas Padjadjaran, Indonesia, to ensure consistency and accuracy in sample handling and analysis.

Collecting spectral data

Diffuse reflectance spectra data were collected using a NIR spectrometer (PSD NIRS i16 USK) with a wavelength range of 1000–2000 nm and a resolution of 0.2 nm. The acquisition process involved 32 scans per sample. The PSD NIRS i16, developed in 2016, has been widely tested for spectral data acquisition across various agricultural products, including rice, honey, mangoes, bananas, grapes, tomatoes, oranges, garlic, onions, soil samples, coffee beans, coffee powder, cocoa beans, and cocoa powder. The device was calibrated using the Thermo Nicolet Antaris™ II, a standard NIRS instrument, at Georg-August University in Goettingen, Germany (Hayati et al., 2021). This calibration ensured high accuracy and reliability for spectral measurements. The PSD NIRS i16 features photodiode sensors with an optical gain adjustable from 2x to a maximum of 16x, and its spectral range spans 1000–2500 nm. It can capture transmittance, reflectance, and absorbance spectrum data, with spectra files saved in *.spa or *.csv formats. To ensure optimal light penetration during measurements, coffee beans were evenly distributed in a cylindrical quartz sample container. After placement, the spectrometer averaged the spectra per sample, ensuring consistent and accurate data acquisition.

Laboratory analysis

Caffeine analysis

Quantifying caffeine in coffee beans is typically performed using high-performance liquid chromatography (HPLC) (Shimadzu, LC 20AT Prominence, Japan). This method is highly effective for separating and measuring various chemical compounds in coffee beans, including caffeine. The HPLC procedure involves several stages, including sample preparation and extraction, following the methodologies of Khasanov et al. (2005), Ciaramelli et al. (2019), and Adnan et al. (2020). Initially, the coffee beans were ground into a fine powder using a grinder. The resulting powder was sieved through a 0.355 mm mesh. To prepare the solution, 10 mg of the coffee powder was dissolved in 10 mL of distilled water. The mixture was stirred for 5–10 minutes at 90 °C to ensure thorough dissolution. The extract was then centrifuged at 5000 rpm for 5 minutes and filtered through a 0.45 µm filter. The supernatant was separated from the solution and used as the sample for HPLC analysis. For the elution process, a gradient method was applied to vary the concentration of the organic modifier. Two solutions were used: (A) 0.025 M sodium acetate with a pH of 6.8 and (B) a mixture of 0.025 M sodium acetate (pH 6.8) and acetonitrile in a 70:30 volume ratio. The column was initially equilibrated for 4 minutes with a 95:5 ratio of solutions A to B. The sample was then introduced into the column, and after 4 minutes, the gradient was adjusted to achieve a 15:85 ratio of solutions A to B. Detection was performed using a UV detector set at a wavelength of 280 nm. Caffeine levels in the samples were quantified and expressed as percentages (%), providing an accurate measure of caffeine content in the coffee beans.

CGA analysis

The quantification of chlorogenic acid (CGA) concentration in coffee beans is commonly carried out using high-performance liquid chromatography (HPLC) (Shimadzu, LC 20AT Prominence, Japan). The determination of CGA followed the methodologies outlined by Shan et al. (2014), Naveen et al. (2018), and Munawar et al. (2024). The preparation process began with grinding the coffee beans into a fine powder, which was then sieved similarly to the procedure used for caffeine measurement. For CGA extraction, the finely ground coffee was mixed with a solution of water and methanol in a 6:4 volume ratio. The mixture was thoroughly agitated and allowed to rest for 5 minutes to ensure proper extraction. The liquid portion, containing suspended particles, was collected and prepared for HPLC analysis. During the HPLC procedure, the mobile phase consisted of a solution containing 70% methanol. A UV detector set to a wavelength of 324 nm was used to identify and quantify CGA. The concentrations of CGA in the coffee samples were expressed as percentages (%), providing an accurate assessment of their CGA content.

Data analysis

Data analysis was conducted using partial least squares regression (PLSR) and principal component regression (PCR), two methods commonly employed in spectroscopy studies. Both techniques aim to develop accurate predictive models, reveal underlying patterns and trends, and analyze the variability within spectral data. PLSR focused on extracting

components strongly correlated with laboratory analysis results (dependent variables) from NIR spectral data (independent variables). In contrast, PCR reduced data dimensionality by combining principal components, which were then used to estimate caffeine and CGA values. To enhance the accuracy of the models, spectral pre-treatment with multiplicative scatter correction (MSC) was applied. This step minimized noise and variability in the coffee spectra, improving data quality for analysis. The analyses were performed using **The Unscrambler X 10.4** software (Camo Software AS, Oslo, Norway), following procedures outlined by Kusumiyati et al. (2021b, 2021c).

Results

Measured data analysis

Table 1 details the caffeine and CGA composition in coffee beans from two botanical varieties, Arabica and Robusta, cultivated in Java and Sumatra. Various factors influence the characteristics of coffee, including genetic makeup, growing environment, cultivation practices, postharvest handling, and analytical methodologies. In Java, caffeine content ranged from 2.14% to 4.54%, with an average of 3.28%. In Sumatra, caffeine content showed a broader range, from 1.48% to 4.24%, and a slightly lower average of 3.02%. Although the average caffeine values were relatively comparable between the two islands, Sumatra exhibited greater variability, reflected by a standard deviation (SD) of 1.67% compared to Java's SD of 0.67%. When categorized by botanical variety, Arabica coffee displayed a caffeine content range of 1.58% to 4.54%, with a mean of 3.10% and an SD of 0.69%. Robusta coffee, on the other hand, had a caffeine content range of 1.48% to 4.24%, with a mean of 3.13% and an SD of 0.73%.

Table 1. Laboratory analysis of caffeine and CGA in coffee beans.

Parameters	Value	Geographical		Botanical	
		Java	Sumatra	Arabica	Robusta
Caffeine (%)	Min	2.14	1.48	1.58	1.48
	Max	4.54	4.24	4.54	4.24
	Mean	3.28	3.02	3.10	3.13
	SD	0.67	1.67	0.69	0.73
CGA (%)	Min	6.12	6.42	6.12	6.42
	Max	12.48	11.32	12.48	10.72
	Mean	8.46	8.50	8.51	8.55
	SD	0.69	1.29	1.45	1.26

Min: Minimum, Max: Maximum, SD: Standard deviation.

Features of the spectra

NIRS spectra visually represented the degree to which materials absorbed or reflected light within the 1000–2000 nm wavelength range. Figure 1a illustrates the spectra obtained from coffee beans using diffuse reflectance (original spectra). These spectra comprise peaks and valleys, which indicate the light intensity detected at various NIRS wavelengths. Notably, the wavelength around 1400 nm is strongly associated with CGA composition, while the wavelength near 1900 nm is closely related to caffeine content. These associations align with findings from previous studies (Budiastra et al., 2018; Ribeiro et al., 2011), which reported similar correlations.

Agricultural commodities exhibit complex chemical compositions. Figure 2 highlights how regression coefficients identify the wavelengths with the greatest influence on modeling each coffee quality parameter. These key wavelengths are distinguished by prominent peaks and troughs. The spectral patterns for each sample type were visually categorized to identify similarities and differences. Figure 1b presents a PCA plot, where multiplicative scatter correction (MSC) was applied to the original absorbance spectra to enhance PCA results. MSC, a widely used normalization method, adjusts spectra by aligning them more closely with a predefined reference. By manipulating spectral size and offset, it allows determination of the dataset's mean (Windig et al., 2008). Additionally, Figure 1b demonstrates the potential to achieve complete spectral separation between Arabica and Robusta coffee beans from Java and Sumatra, with a 100% success rate.

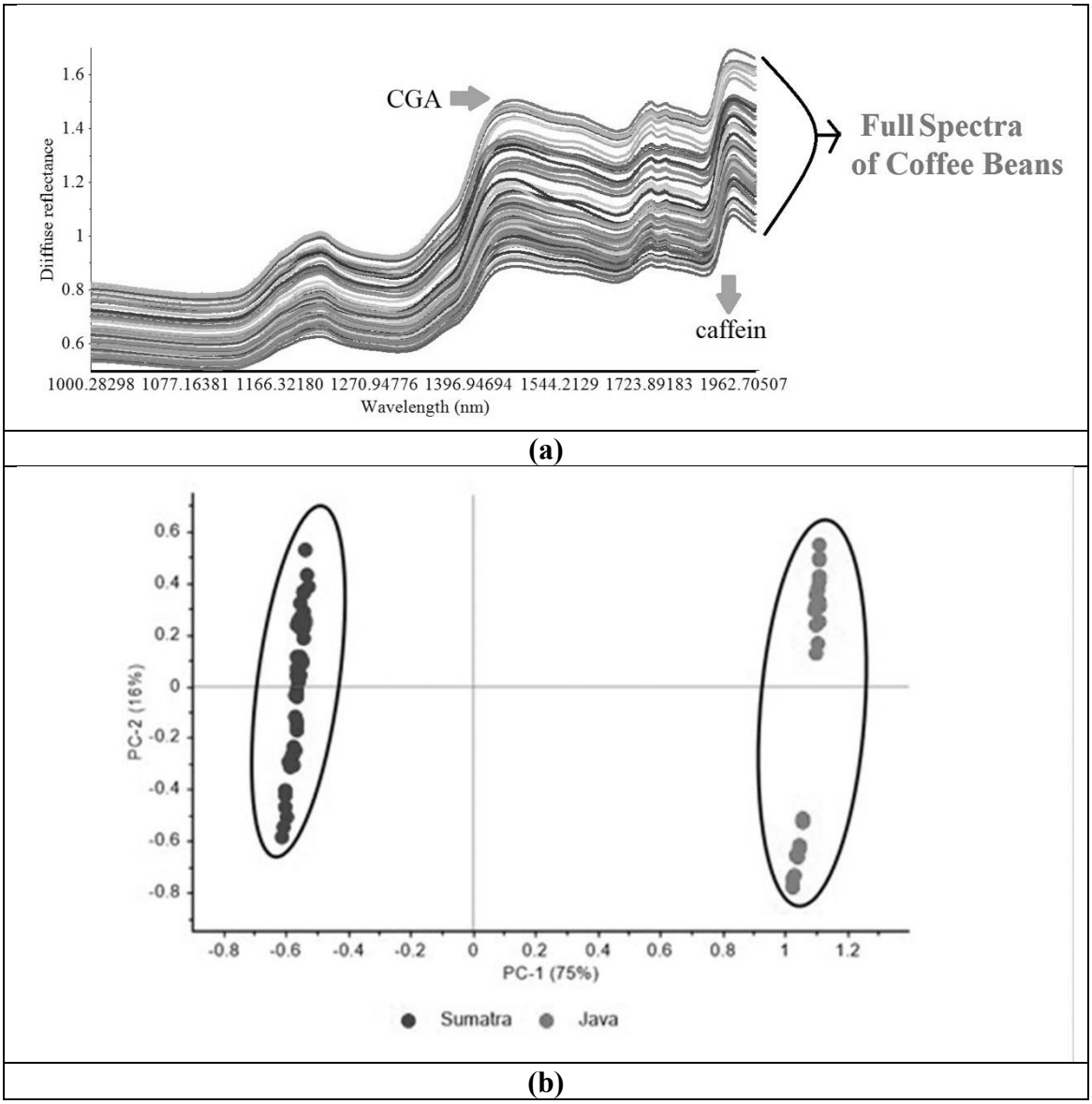
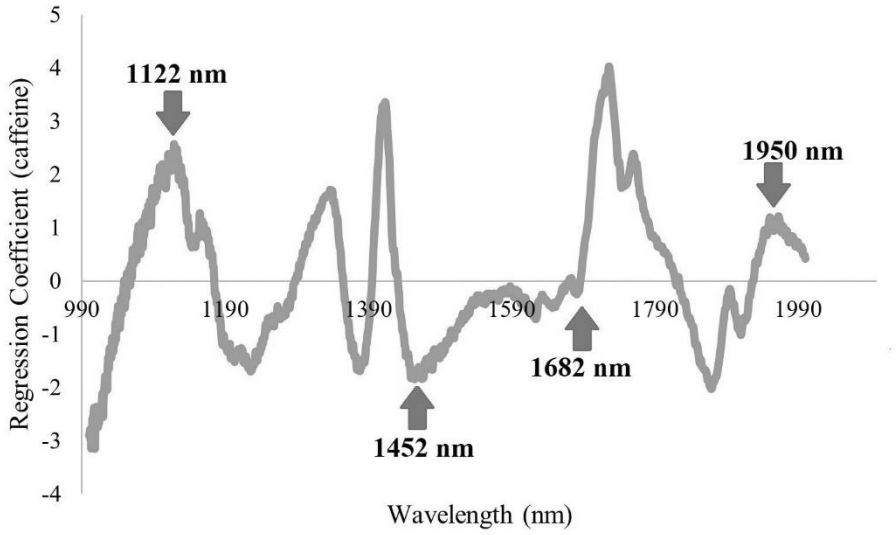


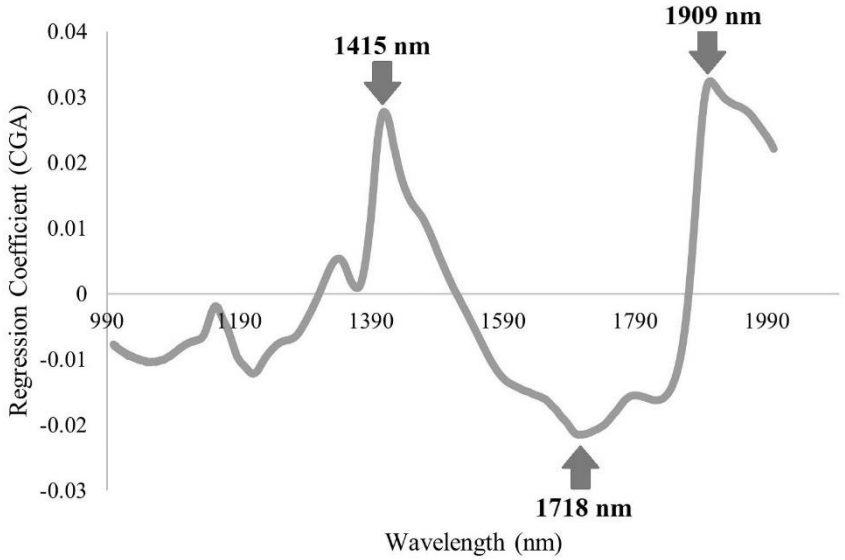
Fig. 1. Spectra characteristics of samples (Arabica and Robusta coffee beans): (a) full spectra of samples, (b) PCA scores plot extracted from the spectra data of two distinct geographical areas.

Model development

Table 2 presents the accuracy values of the PLSR and PCR methods for calibrating and cross-validating caffeine and CGA content in coffee beans. The data indicate that the PLSR method performed best for caffeine, whereas PCR yielded the optimal model for CGA. Regression coefficients (R) serve as statistical indicators describing the strength and direction of the relationship between spectral data and chemical content in regression analysis, playing a critical role in evaluating model performance. For caffeine, the highest accuracy was achieved using PLSR, with a coefficient of correlation in the calibration set (R_{cal}) of 0.85, a root mean square error of calibration (RMSEC) of 0.30, a coefficient of correlation in the cross-validation set (R_{cv}) of 0.82, and a root mean square error of cross-validation (RMSECV) of 0.31. The model also utilized seven latent variables (LVs) and achieved a ratio of prediction to deviation (RPD) of 2.21. For CGA content, PCR provided the most accurate model. It produced an R_{cal} of 0.88, an RMSEC of 0.61, an R_{cv} of 0.88, and an RMSECV of 0.65, with three latent variables and an RPD of 2.18. The reliability of both models is evident from their RPD values, which exceeded the threshold of 2. The best RPD value for caffeine was 2.21, while for CGA, it was 2.18, indicating that both models are capable of producing accurate quantitative predictions.



(a)



(b)

Fig. 2. Regression coefficient of calibration models for predicting caffeine and CGA.

Table 2. Summary of PLSR and PCR results for calibration and cross-validation of caffeine and CGA in coffee beans.

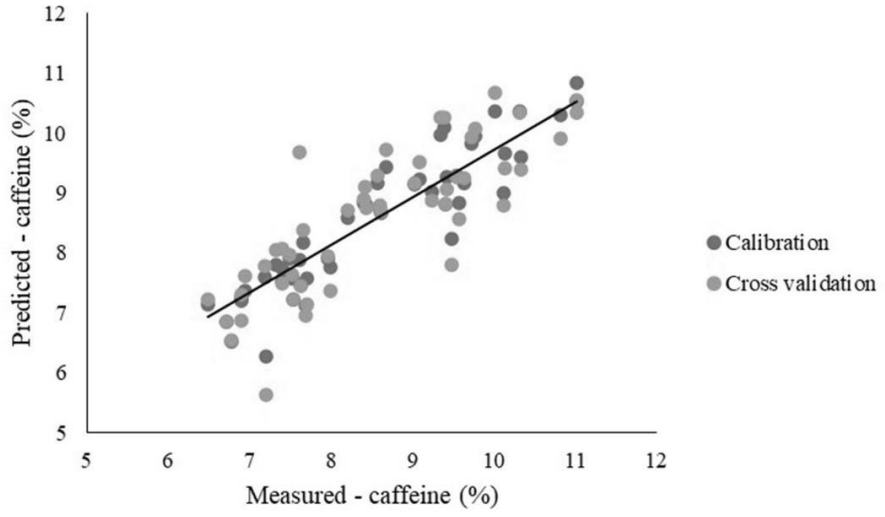
Quality attributes	Regression methods	LVs	R_{cal}	RMSEC	R_{cv}	RMSECV	RPD
Caffeine	PLSR	7	0.85	0.30	0.82	0.31	2.21
	PCR	4	0.84	0.30	0.83	0.32	2.15
CGA	PLSR	2	0.92	0.48	0.85	0.72	1.98
	PCR	3	0.88	0.61	0.88	0.65	2.18

CGA: chlorogenic acid, PLSR : partial least squares regression, PCR: principal component regression, LVs: latent variables (RMSECV), (LVs), and (RPD), R_{cal}: the coefficient of correlation in the calibration set, RMSEC: root mean square error of the calibration set, R_{cv}: the coefficient of correlation in the cross-validation, RMSECV: root mean square error of the cross-validation set, RPD: the ratio of prediction to deviation.

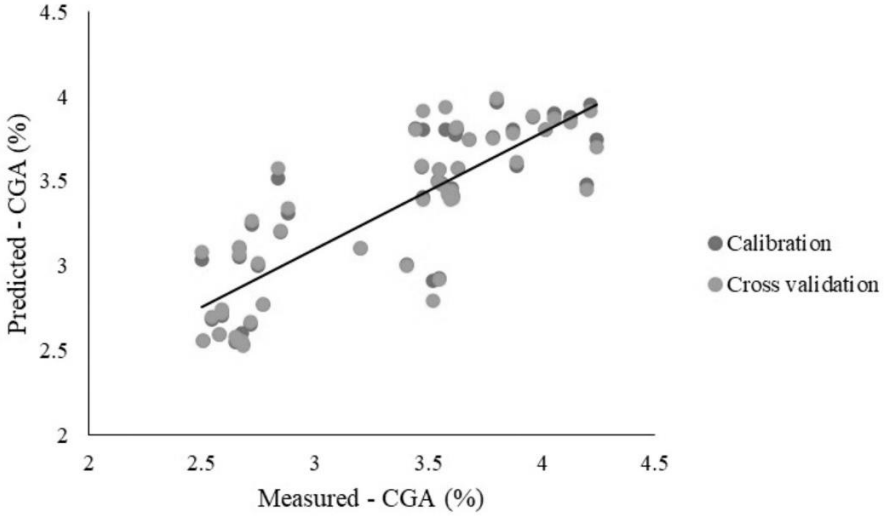
Figure 3 presents scatter plots for caffeine and CGA derived from the original spectra. The black data distribution represents the calibration set analysis, while the grey distribution corresponds

to the cross-validation set. Scatter plots visually depict the relationship between two variables and are commonly used to assess calibration and cross-validation results for caffeine and CGA content in coffee beans. The plots illustrate the outcomes of calibration and cross-validation in NIRS data analysis, demonstrating a strong correlation between reference values (obtained through laboratory analysis) and predicted values (generated by NIRS). This strong correlation indicates that the calibration and cross-validation models effectively predict caffeine and CGA content in coffee beans.

Scatter plots are valuable tools for evaluating the accuracy of prediction models. A better regression model is characterized by a closer alignment between the data distribution and the regression line. The proximity of the plotted data to the regression line reflects the accuracy of NIRS-estimated values in comparison to their corresponding laboratory reference values (Rushing et al., 2016). By facilitating such evaluations, scatter plots allow analysts to assess prediction model accuracy and compare the performance of various multivariate statistical methods (Budiastra et al., 2018). This study identified key wavelengths closely associated with caffeine and water at 1122, 1452, 1682, and 1950 nm (Fig. 2a). For CGA, the wavelengths of 1415, 1718, and 1909 nm showed strong correlations.



(a) Caffeine



(b) CGA

Fig. 3. Scatter plot for calibration model and cross-validation of caffeine and CGA in coffee beans, (a) caffeine, and (b) CGA.

Discussion

In CGA content, Arabica coffee ranged from 6.12% to 12.48%, while Robusta varied between 6.42% and 10.72%. The standard deviation (SD) measures data variation or dispersion within a sample. A large SD indicates significant dispersion from the mean, whereas a low SD suggests limited data point variation (Andrade, 2020). For CGA content in coffee beans from Java, the range was 6.12% to 12.48%, with an average of 8.46% and an SD of 0.69%. In Sumatra, CGA content ranged from 6.42% to 11.32%, with an average of 8.50% and an SD of 1.29%. The caffeine and CGA values observed in this study were higher than those reported by Maia et al. (2013).

In NIRS analysis, spectral peaks and troughs arise from the influence of chemical contents in a substance (Blanco and Villarroya, 2002). NIRS has proven effective in determining the moisture and soluble solid content in pears (Mishra et al., 2021) and evaluating the hardness of apples (Shi et al., 2011). Additionally, UV-Vis-NIRS technology combined with chemometrics has been used to differentiate civet coffee from regular coffee (Suhandy et al., 2018), successfully predicting civet content in roasted Robusta civet coffee (Suhandy and Yulia, 2017). Spectroscopy serves as a non-destructive method for assessing various quality attributes of agricultural commodities, delivering accurate and rapid results. By exposing samples to NIRS light, some wavelengths are absorbed, reflected, or transmitted. This absorption pattern reveals information about the material's chemical bonds and molecular structure, making NIRS spectra crucial for showcasing the chemical composition of a sample in a non-destructive manner. Spectral analysis enables the prediction of various chemical components, including moisture, fat, and carbohydrates (Fonseca et al., 2019; Pandiselvam et al., 2022; Sharma et al., 2017), demonstrating its value in detecting chemical contents.

NIRS technology has widespread applications across agriculture, food and beverage, pharmaceuticals, and environmental analysis. By examining peaks and valleys in NIRS spectra, analysts can enhance the precision and reliability of their methods, making NIRS a robust tool for agriculture and industrial applications. Moreover, PCA effectively classified coffee samples based on their island of origin (Java and Sumatra), underscoring the utility of this approach in distinguishing geographic origins.

The model's accuracy was assessed using parameters such as R_{cal}, RMSEC, R_{cv}, RMSECV, LVs, and RPD (Kusumiyati et al., 2021d; Kusumiyati et al., 2022b). Samples from two distinct geographical areas, Java and Sumatra, were analyzed to determine whether their spectra could be reliably distinguished based on origin. Both sample groups are represented in Figure 1b. PCA revealed that principal component (PC) 1 explained 75% of the total variation in the dataset, while PC2 accounted for an additional 16%. Together, these components described 91% of the dataset's total variance, suggesting that PC1 and PC2 provided sufficient information for effective analysis. The PCA results are visualized in a two-dimensional scatter plot (Fig. 1b), where the combined contribution of PC1 and PC2 exceeds 70% of the total variance. This highlights the effectiveness of PCA for visualizing data distribution through a two-dimensional score scatter plot (Saganowska and Wesolowski, 2017). Spectroscopy was employed to analyze and determine the chemical composition of the samples. Model development in spectroscopy analysis typically involves two key stages: calibration and cross-validation. These stages are critical for assessing model accuracy and offer an objective evaluation of the model's performance with previously unseen data. The caffeine prediction results in this study showed higher accuracy than those reported by Budiastra et al. (2018). Their original spectra yielded R_{cal} (0.582), SEC (0.045), SEP (0.042), and RPD (1.11). However, the findings closely aligned with those of Ayu et al. (2020), where original spectra produced R_{cal} (0.914), SEC (0.067), SEP (0.072), and RPD (1.965).

The prediction model for CGA yielded results that surpassed those of Budiastra et al. (2020), where original spectra obtained Rcal (0.70), SEC (0.2935), SEP (0.3182), and RPD (1.18). Additionally, Shan et al. (2014) predicted CGA content using the standard normal variate (SNV) correction method with accuracy values of R2cv (0.76) and RMSECV (1.10). A good model was predicted to provide a substantial R2 value and minimal error in calibration and validation sets (Kusumiyati et al., 2019b). The results showed that Rcal for caffeine and CGA was 0.85 and 0.88, respectively, indicating a strong positive relationship between spectral data and chemical content. NIRS could be used for predicting caffeine and CGA content in coffee beans accurately. The RPD value in this research was over 1.5. An RPD value below 1.5 showed inadequate predictive capability and unsuitability, while values between 1.5 and 2 indicated that the model was capable of

discerning between low and high response variables. RPD values within the range of 2 to 2.5 implied the potential for generating approximate quantitative predictions. When the value exceeded 2.5 and reached 3 or above, it signified excellent accuracy in predicting (Nicolaï et al., 2007; Saeys et al., 2005).

PLSR emerged as the superior regression model for predicting caffeine (Fig. 3a), whereas PCR demonstrated the highest accuracy for CGA prediction in coffee (Fig. 3b). Both PLSR and PCR create new predictor variables, referred to as components, but they employ different methodologies for their construction. The relative effectiveness of PLSR and PCR depends on the nature of the dataset. PLSR often yields better results in certain contexts, while PCR may prove more effective in others. For example, PLSR has been shown to be more suitable for predicting fat content when evaluating cocoa bean quality, whereas PCR achieves optimal accuracy for moisture content (Kamal et al., 2021). Similarly, studies on protein and hardness prediction using hyperspectral NIR spectroscopy found PLSR to outperform PCR (Mahesh et al., 2015). The comparative accuracy of PLSR and PCR depends on factors such as the dataset's characteristics, the number of predictor variables, and the interpretability of the generated components.

Both regression methods have their respective strengths and weaknesses, with their precision being influenced by additional variables, including data variability, modeling techniques, and the equipment used (Kusumiyati et al., 2022c). Wavelengths correlated with chemical composition are highlighted in Figure 2. Those critical for predicting caffeine and water content include 1128, 1450, 1672, and 1940 nm (Budiastra et al., 2018), while CGA prediction relies on wavelengths such as 1415, 1718, and 1909 nm. Peaks at 1477, 1726, and 1934 nm indicate CGA absorption (Budiastra et al., 2020; Ribeiro et al., 2011). These findings suggest that NIRS serves as an efficient and rapid method for predicting caffeine and CGA content in coffee. Additionally, NIRS has been used to predict primary alkaloids such as caffeine, theobromine, and theophylline in coffee beans (Huck et al., 2005). In previous studies, NIRS was employed to determine caffeine content in Arabica green beans across wavelengths ranging from 1000 to 2500 nm (Ayu et al., 2020). Another analysis identified 17 significant wavelengths for predicting caffeine content (Pizarro et al., 2007).

Conclusions

In conclusion, this study demonstrated that NIRS could effectively predict caffeine and CGA concentrations in coffee beans from two distinct geographical regions. Previous NIRS studies have successfully quantified various components in coffee, including CGA levels, caffeine concentration, and color. This study introduces a novel method for efficiently quantifying caffeine and chlorogenic acid in coffee beans from multiple geographical locations. Not only does this approach deliver rapid results, but it also offers a non-destructive technique, highlighting the potential of NIRS for assessing coffee quality. The results indicated that the developed model performed exceptionally well for predicting caffeine and CGA, with R values $\geq$ 0.85. Key wavelengths at 1122, 1452, 1682, and 1950 nm were strongly correlated with caffeine and water content, while wavelengths at 1415, 1718, and 1909 nm were significantly associated with CGA. The analysis confirmed that the NIRS model for predicting caffeine and CGA was satisfactory, offering a viable alternative to traditional laboratory (destructive) methods. Future studies should explore further optimization of spectral pre-treatments to enhance the precision and reliability of the models. The use of NIRS in the horticultural industry provides several advantages, including rapid analysis, minimal sample preparation, the ability to analyze multiple components simultaneously, and real-time monitoring of plant quality, nutrient content, and ripeness. Moreover, NIRS reduces environmental impact by eliminating the need for chemicals, improves quality control processes, and offers low operational costs once implemented.

Acknowledgments

The authors thank Ine Elisa Putri and Yuda Hadiwijaya for their valuable contributions as study assistants in facilitating and supporting the research.

Conflict of Interest

The authors indicate no conflict of interest in this work.

References

Adnan A, Naumann M, Mörlein D, Pawelzik E. 2020. Reliable discrimination of green coffee beans species: a comparison of UV-Vis-based determination of caffeine and chlorogenic acid with non-targeted near-infrared spectroscopy. Foods 9(6), 788.

Alamri E, Rozan M, Bayomy H. 2022. A study of chemical composition, antioxidants, and volatile compounds in roasted Arabic coffee. Saudi Journal of Biological Sciences 29(5), 3133–3139.

Andrade C. 2020. Understanding the difference between standard deviation and standard error of the mean, and knowing when to use which. Indian Journal of Psychological Medicine 42(4), 409–410.

Awaliyah S, Widiyanto SNB, Maulani RR, Hidayat A, Husyari UD, Syamsudin TS, Marwani E. 2022. Correlation of microclimate of West Java on caffeine and chlorogenic acid in coffea canephora var. robusta. 3BIO: Journal of Biological Science, Technology and Management 4(1), 54–60.

Awwad S, Issa R, Alnsour L, Albals D, Al-Momani I. 2021. Quantification of caffeine and chlorogenic acid in green and roasted coffee samples using HPLC-DAD and evaluation of the effect of degree of roasting on their levels. Molecules 26(24), 7502.

Ayelign A, Sabally K. 2013. Determination of chlorogenic acids (CGA) in coffee beans using HPLC. American Journal of Research Communication 1(2), 78–91.

Ayu P C, Budiastra I W, Rindang A. 2020. NIR spectroscopy application for determination caffeine content of Arabica green bean coffee. IOP Conference Series: Earth and Environmental Science 454(1).

Blanco M, and Villarroya I. 2002. NIR spectroscopy: A rapid-response analytical tool. TrAC - Trends in Analytical Chemistry 21(4), 240–250.

Budiastra IW, Sutrisno, Widyotomo S, Ayu PC. 2018. Prediction of caffeine content in Java Preanger coffee beans by NIR spectroscopy using PLS and MLR method. IOP Conference Series: Earth and Environmental Science 147, 012004.

Budiastra W, Sutrisno M, Widyotomo, Ayu P C. 2020. Determination of trigonelline and chlorogenic acid (CGA) concentration in intact coffee beans by NIR spectroscopy. Agricultural Engineering International: CIGR Journal 22(1), 162–168.

Ciaramelli C, Palmioli A, Airoldi C. 2019. Coffee variety, origin and extraction procedure: Implications for coffee beneficial effects on human health. Food Chemistry 278, 47–55.

Fonseca FG, Funke A, Saechua W, Sirisomboon P. 2019. Precision test for the spectral characteristic of FT-NIR for the measurement of water content of wheat straw. IOP Conference Series: Earth and Environmental Science 301(1).

Girma B, Gure A, Wedajo F. 2020. Influence of altitude on caffeine, 5-caffeoylquinic acid, and nicotinic acid contents of arabica coffee varieties. Journal of Chemistry.

Gobbi L, Maddaloni L, Prencipe SA, Vinci G. 2023. Bioactive compounds in different coffee beverages for quality and sustainability assessment. Beverages 9(1), 3.

Goisser S, Wittmann S, Fernandes M, Mempel H, Ulrichs C. 2020. Comparison of colorimeter and different portable food-scanners for non-destructive prediction of lycopene content in tomato fruit. Postharvest Biology and Technology 167, 111232.

Grassi S, Jolayemi OS, Giovenzana V, Tugnolo A, Squeo G, Conte P, De Bruno A, Flamminii F, Casiraghi E, Alamprese, C. 2021. Near infrared spectroscopy as a green technology for the quality prediction of intact olives. Foods 10(5), 1–12.

Hayati R, Zulfahrizal Z, Munawa AA. 2021. Robust prediction performance of inner quality attributes in intact cocoa beans using near infrared spectroscopy and multivariate analysis. Heliyon 7(2), 1–7.

Huck CW, Guggenbichler W, Bonn GK. 2005. Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. Analytica Chimica Acta 538(1–2), 195–203.

Kamal M, Munawar AA, Sulaiman MI. 2021. Comparison of principal component and partial least square regression method in NIRS data analysis for cocoa bean quality assessment. IOP Conference Series: Earth and Environmental Science 667(1).

Khasanov VV, Dychko KA, Kuryaeva TT, Ryzhova GL, Mal'tseva EV. 2005. A new procedure for caffeine determination. Russian Journal of Applied Chemistry 78(9), 1427–1429.

Kusumiyati K, Hadiwijaya Y, Putri IE. 2018. Determination of water content of intact sapodilla using near infrared spectroscopy. IOP Conference Series: Earth and Environmental Science 207(1), 1–7.

Kusumiyati K, Hadiwijaya Y, Putri IE, Mubarok S. 2019a. Water content prediction of "crystal" guava using visible-near infrared spectroscopy and chemometrics approach. IOP Conference

Series: Earth and Environmental Science 393(1), 1–5.

Kusumiyati K, Hadiwijaya Y, Putri IE, Mubarok S, Hamdani JS. 2020. Rapid and non-destructive prediction of total soluble solids of guava fruits at various storage periods using handheld near-infrared instrument. IOP Conference Series: Earth and Environmental Science 458, 1–7.

Kusumiyati K, Hadiwijaya Y, Suhandy D, Munawar AA. 2021a. Prediction of water content and soluble solids content of 'manalagi' apples using near infrared spectroscopy. IOP Conference Series: Earth and Environmental Science 922(1), 012062.

Kusumiyati K, Hadiwijaya Y, Putri IE, Munawar AA. 2021b. Enhanced visible/near-infrared spectroscopic data for prediction of quality attributes in Cucurbitaceae commodities. Data in Brief 39, 107458.

Kusumiyati K, Hadiwijaya Y, Sutari W, Munawar AA. 2022a. Global model for in-field monitoring of sugar content and color of melon pulp with comparative regression approach. AIMS Agriculture and Food 7(2), 312–325.

Kusumiyati K, Putri IE, Hamdani JS, Suhandy D. 2022b. Real-time detection of the nutritional compounds in green 'Ratuni UNPAD' cayenne pepper. Horticulturae 8(6), 554.

Kusumiyati K, Putri IE, Munawar AA, Suhandy D. 2022c. A data fusion model to merge the spectra data of intact and powdered cayenne pepper for the fast inspection of antioxidant properties. Sustainability 14(1), 1–11.

Kusumiyati K, Mubarok S, Sutari W, Farida, Hamdani JS, Hadiwijaya Y, Putri IE. 2019b. Non-destructive method for predicting sapodilla fruit quality using near infrared spectroscopy. IOP Conference Series: Earth and Environmental Science 334(1), 1–8.

Kusumiyati K, Mubarok S, Sutari W, Hadiwijaya Y. 2021c. Application of spectra pre-treatments on firmness assessment of intact sapodilla using vis-nir spectroscopy. IOP Conference Series: Earth and Environmental Science 644(1), 1–8.

Kusumiyati K, Munawar AA, Suhandy D. 2021d. Fast and contactless assessment of intact mango fruit quality attributes using near infrared spectroscopy (NIRS). IOP Conference Series: Earth and Environmental Science 644(1).

Levate ML, da Silva AC, Costa VW, Gherardi HPR, Pimenta CJ, Henriques SS. 2021. Evaluation of chemical properties of intact green coffee beans using near-infrared spectroscopy. Journal of the Science of Food and Agriculture 101(8), 3500–3507.

Maia SRM, Hunter T, Wright N, Lima DRA. 2013. Caffeine and chlorogenic acids in coffee and effects on selected neurodegenerative diseases. Journal of Pharmaceutical and Scientific Innovation 2(4), 9–17.

Mahesh S, Jayas DS, Paliwal J, White NDG. 2015. Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat. Food and Bioprocess Technology 8(1), 31–40.

Mishra P, Woltering E, Brouwer B, Hogeveen-van EE. 2021. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. Postharvest Biology and Technology 171.

Munawar AA, Kusumiyati K, Andasuryani,Yusmanizar, Adrizal. 2024. Near infrared technology coupled with different spectra correction approaches for fast and non-destructive prediction of chlorogenic acid on intact coffee beans. Acta Technologica Agriculturae 27(1), 23–29.

Naveen P, Lingaraju H, Deepak M, Medhin B, Prasad K. 2018. Method development and validation for the determination of caffeine: An alkaloid from coffea arabica by high-performance liquid chromatography method. Pharmacognosy Research, 10(1).

Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46(2), 99–118.

Pandiselvam R, Mahanti NK, Manikantan MR, Kothakota A, Chakraborty SK, Ramesh SV, Beegum PPS. 2022. Rapid detection of adulteration in desiccated coconut powder: vis-NIR spectroscopy and chemometric approach. Food Control 133.

Pizarro C, Esteban-Díez I, González-Sáiz JM, Forina M. 2007. Use of near-infrared spectroscopy and feature selection techniques for predicting the caffeine content and roasting color in roasted coffees. Journal of Agricultural and Food Chemistry 55(18), 7477–7488.

Purwoko T, Suranto SR, Marliyana SD. 2022. Chlorogenic acid and caffeine content of fermented robusta bean. Biodiversitas Journal of Biological Diversity 23(2), 902–906.

Ribeiro JS, Ferreira MMC, Salva TJG. 2011. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. Talanta 83(5), 1352–1358.

Ribeiro JS, Salva T de JG, Silvarolla MB. 2021. Prediction of a wide range of compounds concentration in raw coffee beans using NIRS, PLS and variable selection. Food Control 125.

Rodríguez P, Villamizar J, Londoño L, Tran T, Davrieux F. 2023. Quantification of dry matter content in hass avocado by near-infrared spectroscopy (NIRS) scanning different fruit zones. Plants 12(17).

Rushing JB, Saha UK, Lemus R, Sonon L, Baldwin BS, Rushing JB, Saha UK, Lemus R, Sonon L, Baldwin BS. 2016. Analysis of some important forage quality attributes of Southeastern wildrye (*Elymus glabriflorus*) using near-infrared reflectance spectroscopy. American Journal of Analytical Chemistry 7(9), 642–662.

Sa AA, Fa AH, Al-Kaf G, Ahmed S, Alhaidrai A. 2023. Determination of caffeine and chlorogenic acid (CGA) in the methanolic determination of caffeine and chlorogenic acid (CGA) in the methanolic extracts coffee (*C. arabica*. L) to seeds and peels (unroasted and roasted) cultivars grown in Yemen by high perfo. Bioequivalence & Bioavailability International Journal 7(1).

Saeys W, Mouazen A M, Ramon H. 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosystems Engineering 91(4), 393–402.

Saganowska P, Wesolowski M. 2017. Principal component and cluster analyses as supporting tools for co-crystals detection. Journal of Thermal Analysis and Calorimetry 130, 45–55.

Santosa KM, Supriyadi S, Anggrahini S, Rahmadia Y. 2021. Sensory analysis, caffeine, chlorogenic acid and non-volatile taste compounds of Arabica coffee (*Coffea arabica*) fermented with sugar addition for brew taste. Indonesian Food and Nutrition Progress 17(2), 37–44.

Saud S, Salamatullah AM. 2021. Relationship between the chemical composition and the biological functions of coffee. Molecules 26(24).

Saurnida ASA, Masrul HE, Hanum C, Karim A, Vinceviča-Gaile Z. 2023. The taste of Arabica coffee in several altitudes and shading conditions. E3S Web of Conferences 374, 00001.

Scholz MB dos S, Kitzberger CSG, Pereira LFP, Davrieux F, Pot D, Charmetant P, Leroy T. 2014. Application of near infrared spectroscopy for green coffee biochemical phenotyping. Journal of Near Infrared Spectroscopy 22(6), 411–421.

Shan J, Suzuki T, Suhandy D, Ogawa Y, Kondo N. 2014. Chlorogenic acid (CGA) determination in roasted coffee beans by Near Infrared (NIR) spectroscopy. Engineering in Agriculture, Environment and Food 7(4), 139–142.

Sharma S, Srivastava S, Singh R, Uttam KN. 2017. Label-free and rapid spectroscopic evaluation of ripening of syzygium cumini fruit. Spectroscopy Letters 50(2), 115–123.

Shi B, Zhao L, Wang H, Zhu D. 2011. Signal optimization approaches on the prediction of apples firmness by near infrared spectroscopy. Sensor Letters 9(3), 1062–1068.

Shokouh P, Christiansen CB, Mellbye FB, Hermansen K, Gregersen S. 2019. Efficacy of Arabica versus robusta coffee in improving weight, insulin resistance, and liver steatosis in a rat model of type-2 diabetes. Nutrients 11(9).

Suhandy D, Suhandy D, Yulia M. 2018. Luwak coffee classification using UV-Vis spectroscopy data: comparison of linear discriminant analysis and support vector machine methods. Aceh International Journal of Science and Technology 7(2), 115–121.

Suhandy D, Yulia M. 2017. The use of partial least square regression and spectral data in UV-visible region for quantification of adulteration in Indonesian palm civet coffee. International Journal of Food Science.

Tsegay G, Redi-Abshiro M, Chandravanshi BS, Ele E, Mohammed AM, Mamo H. 2020. Effect of altitude of coffee plants on the composition of fatty acids of green coffee beans. BMC Chemistry 14(1), 1–11.

Tugnolo A, Giovenzana V, Malegori C, Oliveri P, Casson A, Curatitoli M, Guidetti R, Beghi R. 2021. A reliable tool based on near-infrared spectroscopy for the monitoring of moisture content in roasted and ground coffee: a comparative study with thermogravimetric analysis. Food Control 130.

Vitale R, Bevilacqua M, Bucci R, Magrì AD, Magrì AL, Marini F. 2013. A rapid and non-invasive method for authenticating the origin of pistachio

samples by NIR spectroscopy and chemometrics. Chemometrics and Intelligent Laboratory Systems 121 90–99.

Windig W, Shaver J, Bro R. 2008. Loopy MSC: A simple way to improve multiplicative scatter correction. Applied Spectroscopy 62(10), 1153–1159.

Yusmanizar, Setiasih IS, Nurjanah S, Muhaeimin M, Nurhadi B, Rosniawaty S, Munawar AA. 2019. Fast and non-destructive prediction of moisture content and chologenic acid of intact coffee beans using near infrared reflectance spectroscopy. IOP Conference Series: Materials Science and Engineering 506, 012033.