

## Visible-Near-Infrared Spectroscopy and Chemometrics for Authentication Detection of Organic Soybean Flour

Rudiati Evi Masithoh\*, Muhammad Fahri Reza Pahlawan, Devi Alicia Surya Saputri and Farid Rakhmat Abadi

*Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia*

### ABSTRACT

Organic and non-organic soybean flours, although visually indifferent, have a significant difference in price and nutrition content. Therefore, the accurate authentication detection of organic soybean flour is necessary. Visible-near-infrared (Vis-NIR) spectroscopy coupled with chemometric methods is a non-destructive technique applied to detect authentic or adulterated organic soybean flour. The spectra of organic, adulterated organic, and non-organic soybean flours were captured using a Vis-NIR spectrometer at 350–1000 nm. The spectra were analyzed using partial least squares (PLS), principal component analysis (PCA), and the combination of these two with discriminant analysis (DA). The results showed that PCA using PC1 and PC2 could differentiate organic and non-organic soybean flours, whereas PC1 and PC4 can detect pure and adulterated organic soybean flours. The PCA–linear DA models showed 98.5% accuracy (Acc) for predicting pure organic and adulterated soybean flours and 100% Acc for predicting organic and non-organic flours. Moreover, PLS regression models resulted in a high  $R^2$  of >95% for predicting organic and non-organic flours and pure and adulterated soybean flours. In addition, the PLS-DA models can differentiate organic from non-organic soybean flour and distinguish pure and adulterated soybean flours with 100% Acc and reliability.

### ARTICLE INFO

*Article history:*

Received: 12 March 2022

Accepted: 28 July 2022

Published: 06 March 2023

DOI: <https://doi.org/10.47836/pjst.31.2.03>

*E-mail addresses:*

[evi@ugm.ac.id](mailto:evi@ugm.ac.id) (Rudiati Evi Masithoh)

[freza836@gmail.com](mailto:freza836@gmail.com) (Muhammad Fahri Reza Pahlawan)

[devialicia01@mail.ugm.ac.id](mailto:devialicia01@mail.ugm.ac.id) (Devi Alicia Surya Saputri)

[faridra1998@gmail.com](mailto:faridra1998@gmail.com) (Farid Rakhmat Abadi)

\*Corresponding author

*Keywords:* Authentication, PCA, PCA-LDA, PLS-DA, PLSR, soybean flour, Vis-NIR

## INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is a protein-rich commodity popularly used as a staple food in some Asian countries (Hartman et al., 2016). Soybean is usually processed into various foods, such as tofu, tempeh, soy sauce, soy milk, miso, and beverages, or consumed as soy grain or flour. Soybean is labeled “organic” if produced from organic farming, where cultivation is performed without synthetic fertilizers and pesticides (Gomiero, 2018). The organic soybean market has grown rapidly along with increased public awareness of the environment, human health, and food safety.

The high demand for organic soybean-based products can lead to potential fraudulent practices such as adulteration. Food adulteration is the intentional addition or substitution of inferior substances into the main food ingredients. Food adulteration can be dangerous to consumers’ health; it affects food safety and quality and reduces consumer trust (Meerza & Gustafson, 2019).

Although organic and non-organic soy flours have similar nutritional contents, they differ in nutritional composition. Organic soybeans contain more sugars, total proteins, and zinc but less fiber, saturated fat, and omega-6 fatty acids than non-organic soybeans (Bøhn et al., 2014). However, organic and non-organic soybean flours have no visible difference, thus causing difficulty in distinguishing them visually if both products are mixed. Given the higher cost of organic soybeans, the addition of non-organic soybeans into organic ones is inevitable. Several methods, ranging from analytical (Martins et al., 2019), chromatographic (Esteki, Simal-gandara, et al., 2018) to spectroscopic ones (Esteki, Shahsavari, et al., 2018), have been used to detect food additives or adulteration.

The infrared (IR) spectroscopy method has been used to detect adulteration (Sørensen et al., 2016) or food fraud (Nobari-Moghaddam et al., 2021). IR spectroscopy studies have been carried out to detect adulteration in commodities that have similar visual appearances, such as adulteration in brown sugar (using Fourier transform (FT)-NIR and FT-IR spectroscopy) (Masithoh, Roosmayanti, et al., 2021; Rismiwandira et al., 2020; Roosmayanti et al., 2021), dairy milk (Jawaid et al., 2013), or sugar (Masithoh, Rondonuwu, et al., 2020). On the other hand, limited studies have been conducted to detect adulteration using visible–near-IR (Vis-NIR) spectroscopy. Some focused on adulterating desiccated coconut powder with coconut milk residue (Pandiselvam et al., 2022) or green banana flour with wheat (Ndlovu et al., 2021). Nonetheless, studies for detecting adulteration of non-organic to organic soybean flours using Vis-NIR spectra have not been studied.

Therefore, this study aimed to investigate the feasibility of Vis-NIR spectroscopy to detect the authentication of organic soybean flour using supervised and unsupervised pattern recognition methods. In this study, the discrimination methods applied were principal component analysis (PCA), which is an unsupervised and often used for dimension reduction and data visualization (Masithoh, Lohumi, et al., 2020) and PCA combined

with linear DA (PCA-LDA), which is a supervised method used to classify data with a predetermined class (Khuwijitjaru et al., 2020). Partial least square regression (PLSR) is the most used quantitative multivariate analysis method for predicting numerical parameters in spectroscopy. For classification with qualitative parameters, PLSR is combined with discriminant analysis (PLS-DA). Therefore, four multivariate analyses were used in the present work: PCA, PCA-LDA, PLSR, and PLS-DA. This research is the first study of the quantitative assessment of authentication and adulteration in organic soybean flour based on Vis-NIR spectroscopy using various pattern recognition methods.

## MATERIALS AND METHODS

### Sample Preparation

Organic and non-organic soybean flours were obtained from local markets in Indonesia from respected online shops after studying the sellers from the reviews provided by the previous buyers. The sample price for organic soybean was significantly higher compared to non-organic ones. PCA analysis was performed to confirm organic and non-organic samples, which resulted in the clear separation of both samples. Four brands of organic flour and two brands of non-organic flour were used in this study. Soybean flours from different brands were mixed to prepare pure and pure non-organic flours. Pure flours were sieved manually using a 50-mesh sieve (0.29 mm, American Society for Testing and Materials standard) to obtain uniform-size samples. Samples that did not pass the 50-mesh sieve were re-grinded and re-sieved.

Five concentrations (5, 10, 15, 20 and 25%) of non-organic soybean flour as adulterant materials were added to organic soybean flour. Ten samples were prepared for each adulterant concentration. Each sample consisted of 80 g flour mixed manually in a closed bottle for 5 min. Another 20 samples of pure organic and pure non-organic soybean flours were prepared. In total, 70 samples were dried using a food dehydrator at 60°C for 12 h to remove excess water before spectrum acquisition.

### Spectrum Acquisition

A soybean flour sample was placed in an aluminum cup with 1 cm height and 4 cm diameter for spectrum acquisition. All samples were scanned using a Vis-NIR miniature spectrometer (Flame-T-VIS-NIR Ocean Optics, 350–1000 nm) with tungsten halogen light (360–2400 nm, HL-2000-HP-FHSA Ocean Optics) and a reflection probe (QR400-7-VIS-NIR Ocean Optics). A probe holder was purposely built to hold the probe and minimize the external light. The distance between the probe and the sample was 5 mm. Figure 1 shows the spectrum measurement setup.

Reflectance spectra were collected using OceanView 1.6.7 software with an integration time of 150 ms, scanning average of 50, and boxcar width of 1. The white and black reference spectra were measured to calibrate the spectrometer instrument. Instrument calibration was repeated every 10 scanings. Ten spectra were obtained for each sample of pure and adulterated soybean flour at room temperature (28°C), resulting in a total of 600 spectra.

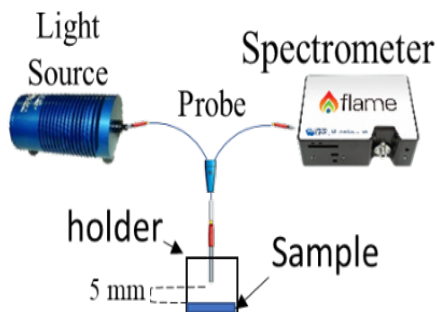


Figure 1. Schematic of the spectrum measurement set-up

### Multivariate Analysis

A total of 600 spectra were obtained from pure non-organic, purely organic, and adulterated organic soybean flours (5, 10, 15, 20 and 25% of adulterant). The obtained spectra were compiled in MS Excel® and imported to Unscrambler®X software (CAMO, Oslo, Norway) for multivariate analysis. This research used PCA, PCA-LDA, PLSR, and PLS-DA to differentiate organic and non-organic soybean flours and determine the adulteration concentration of non-organic in organic soybean.

Unsupervised techniques, such as PCA, were used for dimension reduction, pattern recognition, and outlier identification (Berrueta et al., 2007). PCA was applied to discriminate pure organic and non-organic soybean flours using 200 spectra. Moreover, PCA was used to discriminate pure and adulterated organic soybean flours using 600 spectra by omitting the spectra of pure non-organic soybean flour. Outlier identifications were performed by analyzing residuals and  $T^2$ . The supervised analysis used in this study was PCA-LDA, PLSR, and PLS-DA. For supervised analysis, the samples were randomly divided into calibration and prediction data sets. The calibration sets for organic and non-organic analysis consisted of 133 data, whereas pure and adulterated organic analysis consisted of 400 data. The prediction set for organic and non-organic analysis and pure and adulterated organic analysis consisted of 67 and 200 data, respectively.

PCA-LDA was conducted using a linear method with seven components. PCA-LDA uses PCA to reduce the spectral dimension and LDA to maximize the separation of a given class (Berrueta et al., 2007). PCA-LDA was performed using spectra as the predictor and sample types, such as organic, non-organic, and adulterated flours, as classification category variables. The classification category variables used were labeled “Pure” for pure organic soybean flour and “5, 10, 15, 20 and 25%” for adulteration concentrations of non-organic soybean flour. The classification category variables used in the organic and non-organic analysis were the “Organic and “Non-organic” label.

PLSR is a supervised quantitative analysis used to analyze the correlation between spectra and quantitative variables. PLSR analysis was performed using spectrum data as predictors and quantitative variables as responses. The PLSR method was applied to distinguish organic, and non-organic soybean flours using quantitative Y-variables, namely, purely organic and various adulteration concentrations (5, 10, 15, 20 and 25%). Moreover, PLSR analysis was used to predict the adulteration of non-organic to organic soybean flour, using Y-variables based on the binary label “0” for organic soybean flour and “1” for non-organic soybean flour.

PLS-DA is a classification analysis technique that can analyze the correlation between spectra and non-numerical variables in binary label form. Similar to PLSR, PLS-DA is performed using spectrum data as predictors and quantitative variables as responses (Vieira et al., 2021). For the determination of organic and non-organic soybean flour, the binary label was determined as “0” for organic soybean flour and “1” for non-organic soybean flour. For the determination of pure and adulterated organic soybean flour, the binary label “0” was used for the pure organic sample, and the value of “1” was used for adulterated samples in all concentrations. To obtain a clear class prediction, we classified the predicted values <0.5 as “0” and >0.5 as “1.” The predicted binary values were then classified into DA parameters (Table 1). The DA classes from each class were used to calculate the model accuracy (Acc) and reliability (Rel) using Equations 1 and 2, respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Rel = \left\{ \left( \frac{TP}{TP + TN} + \frac{TN}{TN + FP} \right) - 1 \right\} \times 100\% \quad (2)$$

Table 1

*PLS-DA binary predicted value classification*

Case	True Class	Predicted Value	Predicted Class	DA Parameter
1	1	0.55	1	True Positive (TP)
2	0	0.37	0	True Negative (TN)
3	0	0.67	1	False Positive (FP)
4	1	0.49	0	False Negative (FN)

## RESULTS AND DISCUSSION

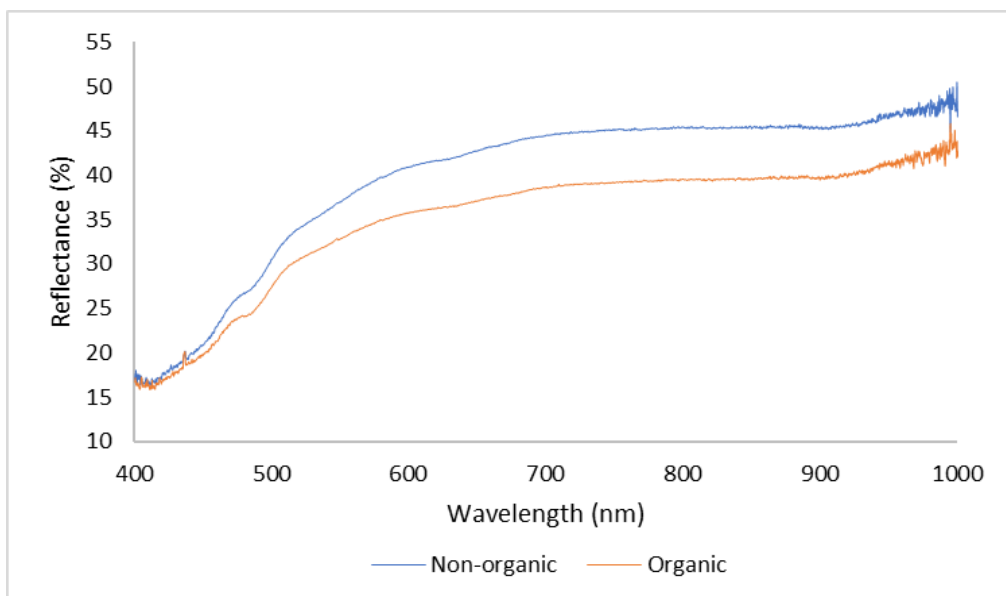
### Vis-NIR Spectra of Soybean Flour

Figure 2a shows the Vis-NIR spectra of pure organic and non-organic soybean flour. The Vis-NIR spectra showed information regarding pigments and molecular vibration (Guo et al., 2016). In reflectance spectra, a low reflectance indicates the absorption of photon energy

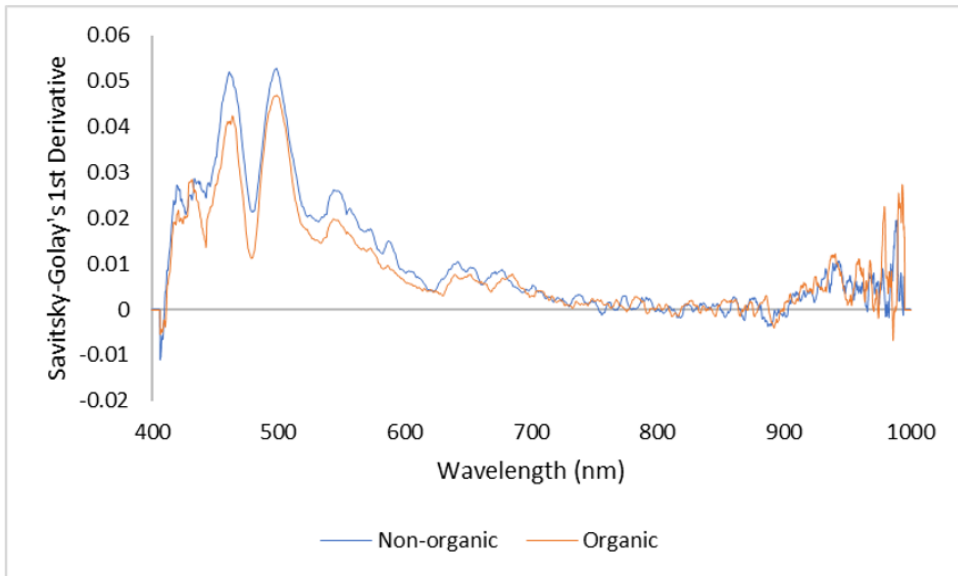
by a certain component. In general, as illustrated in Figure 2a, the absorption of organic flour to visible and NIR light is higher than that of non-organic flour but determining at which wavelength the absorption occurs is difficult. Figure 2b shows the organic and non-organic soybean flours after pre-processing via the Savitzky–Golay 1<sup>st</sup> derivative method. Unlike the original spectra shown in Figure 2a, in Figure 2b, the absorption peaks or valleys were more distinct at 430–443, 460, 480, 500, 534, 550, and 950–1000 nm.

Figure 3a shows the original Vis-NIR spectra of pure and adulterated organic soybean flours. In general, the higher the concentration of adulterant concentration, the higher the reflectance value. However, similar to Figure 2a, given the original spectra, no distinct peaks revealed the absorption of molecules. After processing the original spectra using the Savitzky–Golay 1<sup>st</sup>-derivative method, the peaks and valleys became visible (Figure 3b). The peaks and valleys appeared at the wavelengths of 420–443, 460, 480, 500, 533, 550, and 970–1000 nm.

The wavelengths between 400 and 500 nm contained information related to the absorption of carotenoid pigments (Monma et al., 1994). High absorptions at 500–600 nm due to anthocyanin (Merzlyak et al., 2003) were also found in green and yellow soybean flours (Pahlawan et al., 2022). Low values at 600–700 nm denote the weak absorptions of chlorophylls (Lichtenthaler & Buschmann, 2001). The weak water absorption at 700–1000 nm was due to the low detection of water molecules using the Vis/NIR spectrometer (Zhang et al., 2012). However, several troughs around 950–1000 nm reflected the absorption information of O-H, C-H, and N-H bonds, which build the functional components of soybean (Dixit et al., 2011).

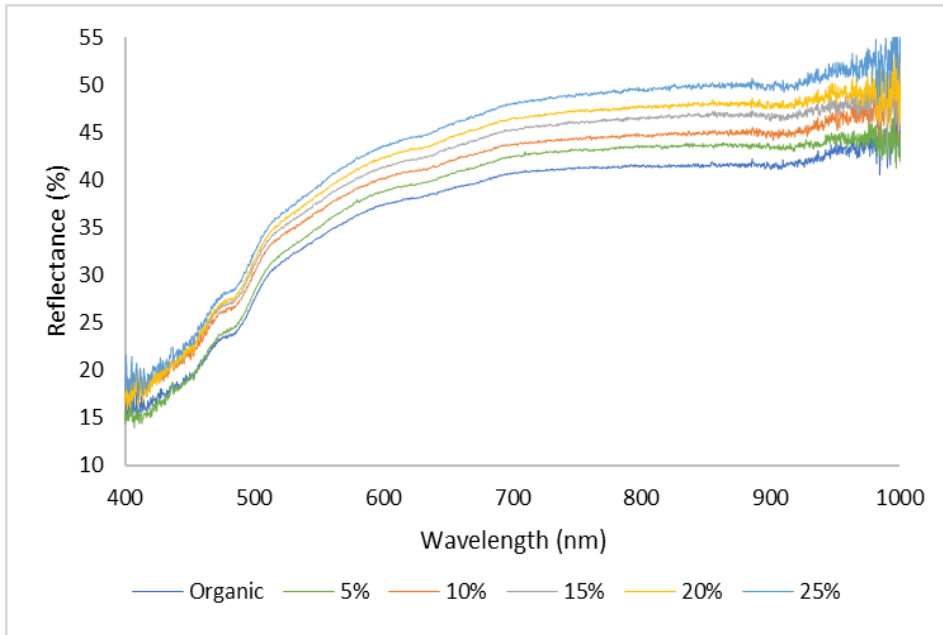


(a)

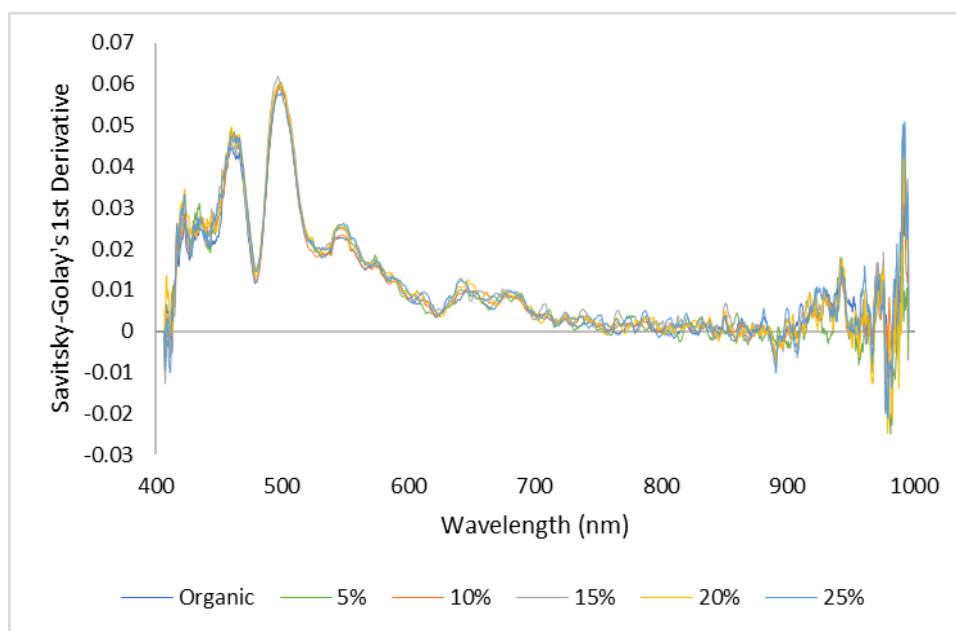


(b)

Figure 2. (a) Original and (b) Savitzky–Golay 1<sup>st</sup>-derivative spectra of organic and non-organic soybean flour observed with Vis-NIR spectroscopy



(a)



(b)

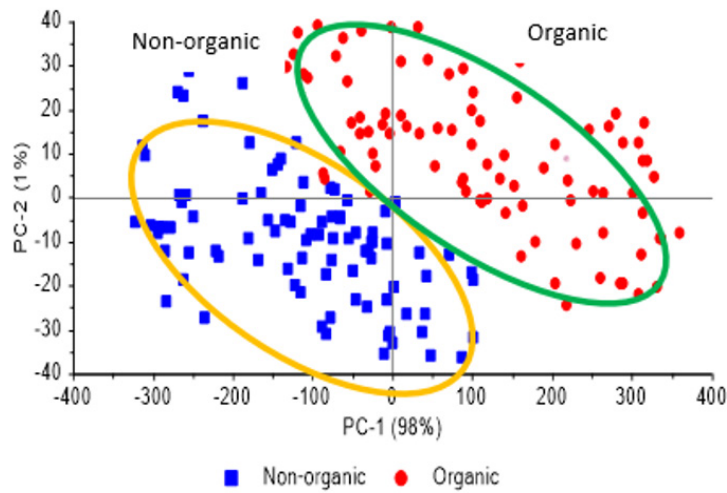
Figure 3. (a) Original and (b) Savitzky–Golay 1<sup>st</sup> derivative of the Vis-NIR spectra of pure and adulterated organic soybean flours in various concentrations

### PCA-LDA

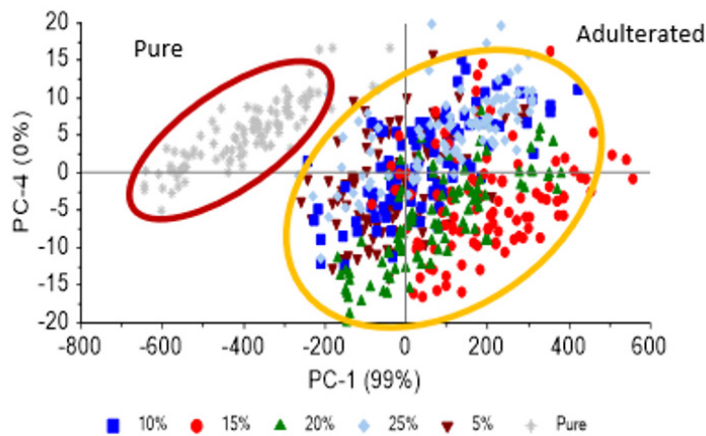
Figure 4a shows the PCA score plot of pure organic and pure non-organic soybean flours. The principal component (PC1) and PC2 explained 98% and 1% of the variance, respectively. These PCs classified organic and non-organic flours, in which non-organic flour had a negative PC1, whereas organic flour was PC1 positive. As illustrated in Figure 4b, purely organic and adulterated organic soybean flours were distinctively separated. The clear separation of pure and adulterated organic soybean flours was expressed by PC1 and PC4, which explained 99% of their variance. Pure organic flour was located at the PC1 negative axis, whereas adulterated organic samples were spread from the negative to positive axes of PC1. Samples with 5–20% adulteration concentrations can be differentiated; the higher the adulteration concentration, the higher the PC1 positive score values.

Given that PCA models explained 99% of the total variance, they were considered good. The explained variance showed the percentages of variances represented by the new variable (PC). Therefore, maximizing the explained variance will result in an ideal model to avoid overfitting.





(a)



(b)

Figure 4. PCA score plot of (a) organic and non-organic soybean flours; and (b) pure and adulterated organic soybean flours in various concentrations

Figure 5a displays the loading plot showing the wavelengths that contributed to separating organic and non-organic flours. Loadings of PC1 and PC2 at wavelengths of 400–500 nm, which correspond to carotenoid molecules, were responsible for classifying organic and non-organic flours. The result corresponded to the finding by Lakshmisha et al. (2012), who showed the different carotenoid contents of organic and non-organic soybean leaves and oils.

Based on the loading plot in Figure 5b, the classification of pure and organic adulterated organic soybeans was determined mostly from PC1, whereas the loadings of PC4 were

relatively flat. Using only two axes (PC1 and PC4), the higher the concentration of adulteration of non-organic flour, the higher the values of PC1. However, this trend only applied to 5–15% adulteration. For 20–25% adulteration concentrations, the samples overlapped with other data with different concentrations, implying that another axis or PC is required to describe data for 20–25% adulteration concentrations.

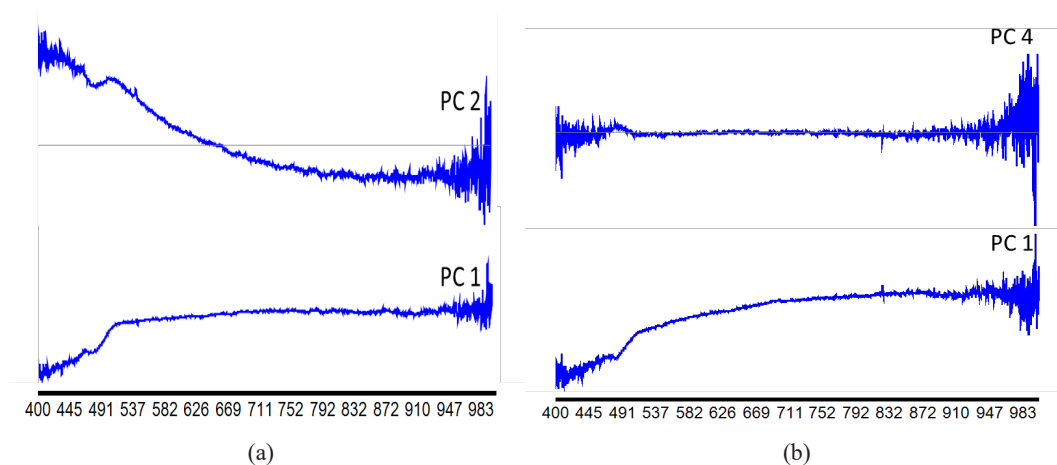


Figure 5. PCA loading plot of (a) organic and non-organic soybean flours; and (b) pure and adulterated organic soybean flours in various concentrations

Figure 6 shows the LDA classification of two organic and non-organic sample categories using a linear model and seven PCs. With the calibration data set, the LDA model can classify with 100% Acc with more distinct separation compared with the classification using PCA (Figure 4a). Table 2 shows that the classification also perfectly works using the prediction data set.

Figure 7 shows the LDA classification of pure and adulterated samples. Although not as good as the organic and non-organic (O versus N) LDA model shown in Figure 6a, the pure organic and adulterated (P versus A) LDA model exhibited a very good Acc (98.5%). The classification capability of the P versus A model was considered good, with only 4 out of 200 samples being misclassified using the prediction data set (Table 3).

The results of PCA-LDA in this research are comparable to those of studies on the classification of soybean oils based on NIR spectra (de Almeida et al., 2021) and LDA for authentication of babassu oil using Mid-IR spectra (Pereira et al., 2022). The results were also better than those of LDA in the authentication of dark-brown sugar using UV-Vis, fluorescence spectroscopy, and mass spectrometry (Chen et al., 2021). Therefore, the PCA-LDA model can classify organic and non-organic soybean flour or detect pure and adulterated organic soybean flour.

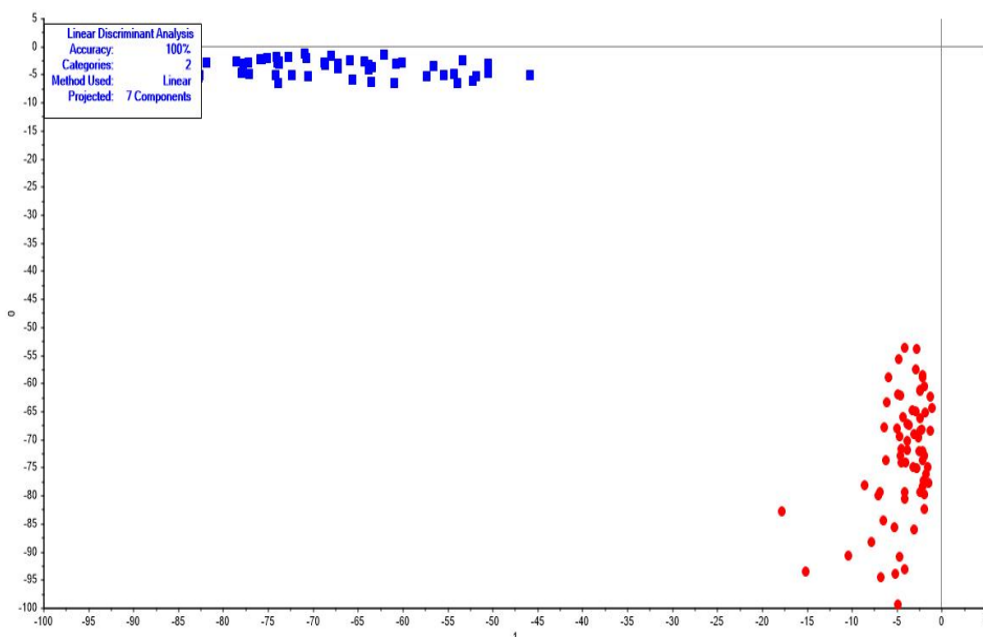


Figure 6. Classification of organic and non-organic soybean flours using PCA-LDA

Table 2

PCA-LDA model of organic and non-organic flours using the prediction data set

True class	N	Predicted Class		Classification Accuracy (%)
		Organic	Non-Organic	
Organic	35	35	0	100
Non-Organic	32	0	32	100

Table 3

PCA-LDA model of pure organic and adulterated flours using the prediction data set

True class	N	Predicted class						Classification Accuracy (%)
		Pure	5%	10%	15%	20%	25%	
Pure	32	32	0	0	0	0	0	100
5%	40	0	37	3	0	0	0	93
10%	38	0	0	38	0	0	0	100
15%	32	0	0	1	31	0	0	97
20%	33	0	0	0	0	33	0	100
25%	25	0	0	0	0	0	25	100

N = number of samples

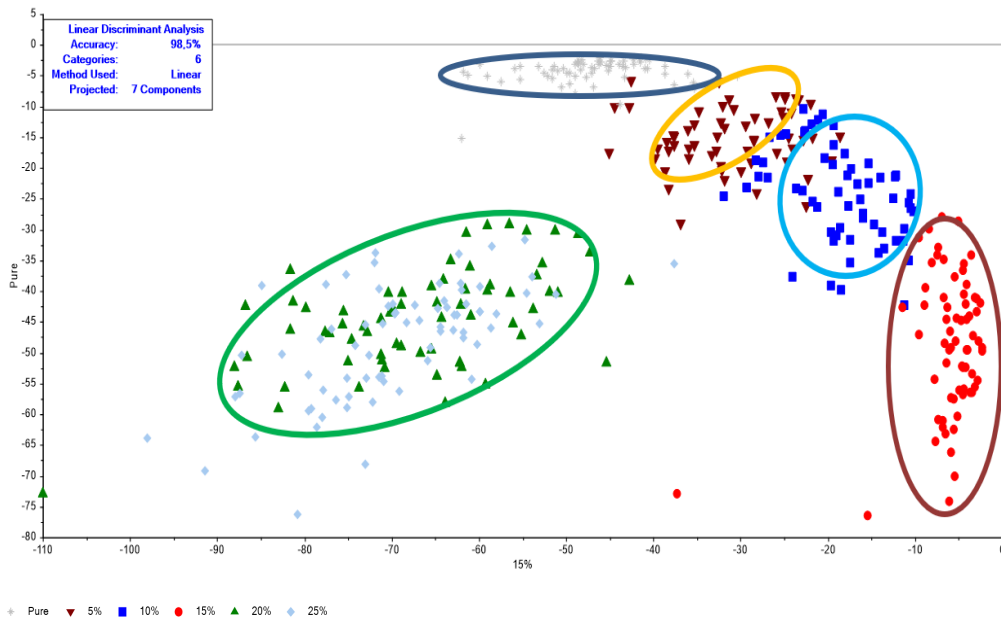


Figure 7. Classification of pure and adulterated soybean flours using PCA-LDA

### PLSR and PLS-DA

The PLSR model was used to distinguish organic and non-organic soybean flours. In the PLSR analysis, quantitative variables were based on binary labels “0” for organic and “1” for non-organic flour. Table 4 shows the PLSR results for determining organic and non-organic flour samples. Given the original and several pre-processed spectra (Table 4), the PLSR models had a high determination coefficient ( $R^2$ ) > 0.97 and low root mean square error (RMSE) < 0.06 for calibration. The results indicated that applying the pre-processing method to the original Vis-NIR spectra did not significantly affect the PLSR results. All the findings showed that the  $R^2$  and RMSE values of all calibration and validation data sets were very close, implying that the calibration PLSR models were robust. When applied to prediction data sets, the obtained PLSR models also yielded a high performance, as indicated by the high value of  $R^2$  (>0.97) and low RMSE (<0.07) of the prediction.

Figure 8a shows the PLSR plot of organic and non-organic soybean flours. The model plot was developed using reflectance spectra as X-variables and binary numbers “0” for pure organic flour and “1” for non-organic flour as Y-variables. By using Savitzky–Golay 1<sup>st</sup> derivative pre-processed spectra, the PLSR model can predict the class of organic and non-organic soybean flours with  $R^2$  of 0.98 and RMSE of 0.07.

Table 4

PLSR results for distinguishing organic and non-organic soybean flour

Pre-processing	Calibration		Cross Validation		Prediction	
	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE
Raw	0.97	0.07	0.97	0.08	0.98	0.07
SGD1	0.99	0.06	0.97	0.09	0.98	0.07
SGD2	0.98	0.07	0.97	0.08	0.97	0.08
SNV	0.99	0.06	0.98	0.07	0.98	0.08
MSC	0.99	0.06	0.98	0.07	0.98	0.07
Norm	0.99	0.05	0.98	0.06	0.98	0.06

SGD1 = Savitzky–Golay's 1<sup>st</sup> derivative, SGD2 = Savitzky–Golay's 2<sup>nd</sup> derivative, SNV = standard normal variate, MSC = multiple scatter correction, Norm = normalization, R<sup>2</sup> = determination coefficient, and RMSE = root mean square error.

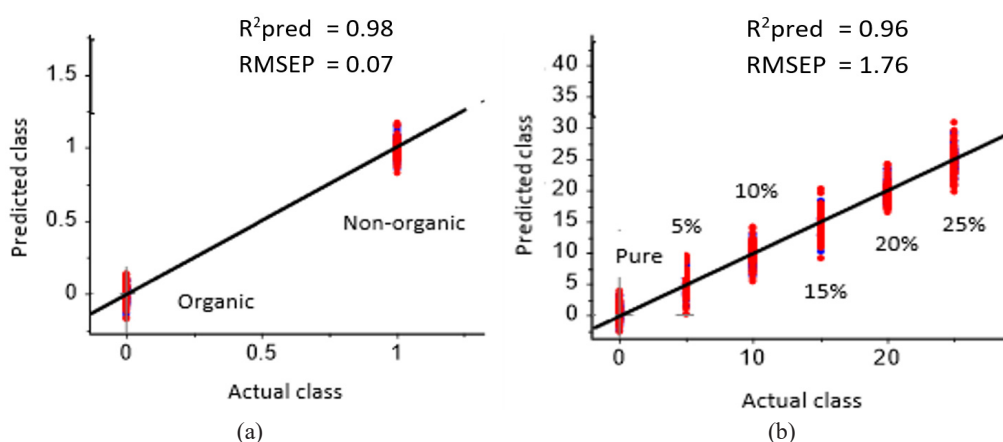


Figure 8. PLSR plots of (a) organic and non-organic soybean flour; and (b) pure and adulterated organic soybean flour

For the differentiation of pure and adulterated organic soybean flours using PLSR, quantitative variables were used as Y-variables and Vis-NIR reflectance spectra as X-variables. Non-organic soybean flour in various concentrations of 0, 5, 10, 15, 20 and 25% was added to the organic soybean flour. Table 5 shows the PLSR results of the prediction for non-organic adulteration in organic soybean flour using various original and pre-processed spectra. For calibration and validation models, all obtained R<sup>2</sup> were significantly high (>0.95), and the RMSEs were very low (<2%). Similar to the findings shown in Table 4, the effect of pre-processed spectra on PLSR was not significantly different from that without pre-processing and the original spectra. These phenomena

were characteristics of the Vis-NIR spectra; that is, the pre-processing method did not significantly improve the PLSR results, as also reported by (Masithoh, Pahlawan, et al., 2021; Pahlawan et al., 2021) when predicting solid soluble content and the pH of banana.

Figure 8b shows the PLSR plot of pure and adulterated organic soybean flours. Using Savitzky–Golay 1<sup>st</sup> derivative pre-processing of spectra, the PLSR model can predict the class of organic and non-organic soybean flours with R<sup>2</sup> of 0.96 and RMSE of 1.76. The PLSR result confirmed the capability of Vis-NIR spectroscopy in detecting low concentrations of non-organic soybean flour adulteration in organic samples.

Table 5

*PLSR results for the prediction of adulteration in non-organic to organic soybean flour*

Pre-processing	Calibration		Cross Validation		Prediction	
	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE
Original	0.96	1.74	0.95	1.97	0.95	1.99
SGD1	0.97	1.49	0.95	1.89	0.96	1.76
SGD2	0.96	1.75	0.95	1.99	0.95	1.98
SNV	0.95	1.84	0.94	2.06	0.95	0.19
MSC	0.95	1.84	0.94	2.07	0.95	1.96
Norm	0.96	1.77	0.95	1.98	0.95	1.99

SGD1 = Savitzky–Golay’s 1<sup>st</sup> derivative, SGD2 = Savitzky–Golay’s 2<sup>nd</sup> derivative, SNV = standard normal variate, MSC = multiple scatter correction, Norm = normalization, R<sup>2</sup> = determination coefficient, and RMSE = root mean square error

In this research, the determination of organic, non-organic, pure, and adulterated soybean flours was performed using PLS-DA. PLSR and PLS-DA have the same mathematical operations. However, the difference is in the response variables, in which PLS-DA uses categorical responses, whereas PLSR uses continuous sample responses. In PLS-DA, reflectance spectra were used as X-variables, and binary numbers 0 or 1 as Y-variables. The values of 0 and 1 were assigned to organic and non-organic soybean flours, respectively. The values of 0 and 1 were assigned to pure and adulterated organic soybean flours. Table 6 shows the PLS-DA performances of organic and non-organic models and pure and adulterated organic models.

The PLS-DA performances were determined based on the Acc, which showed the percentage of correct classification compared with the number of samples, and Rel, which revealed the model’s capability to predict each class correctly (Vieira et al., 2021). Table 6 shows the Acc and Rel of the PLS-DA model. The Acc and Rel showed that the model could classify samples perfectly. For organic and non-organic models, Acc and Rel of 100%

mean that the model can differentiate organic from non-organic soybean flour. For pure and adulterated organic models, Acc and Rel of 100% mean that PLS-DA can differentiate pure samples (0% of adulterant) from those containing adulterant in various concentrations. In this research, PLS-DA performance was better compared with the authentication of turmeric (Khodabakhshian et al., 2021) or *Ginkgo biloba* extract (Walkowiak et al., 2019).

Table 6

*PLS-DA performances using Vis-NIR original spectra*

Model	Calibration		Prediction	
	Acc	Rel	Acc	Rel
Organic and non-organic	100%	100%	100%	100%
Pure and adulterated organic	100%	100%	100%	100%

Acc = accuracy; Rel = reliability

## CONCLUSION

The authentication of organic soybean flour is important to guarantee product quality. Using Vis-NIR spectroscopy combined with supervised and unsupervised multivariate analysis, namely, PCA, LDA, PLSR, and PLS-DA, organic and non-organic soybean flours can be distinguished perfectly. Moreover, these multivariate analyses determine pure and adulterated organic soybean flours. With the relatively low price of Vis-NIR fiber optic spectra, the resulting models can be used for small-scale soybean industries to assure the quality of their products, especially in food authentication.

## ACKNOWLEDGEMENT

Great appreciation is delivered to Universitas Gadjah Mada for financial support under Hibah Rekognisi Tugas Akhir 2021 No: 3190/UN1/DITLIT/DIT-LIT/PT/2021.

## REFERENCES

- Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158(1-2), 196-214. <https://doi.org/10.1016/j.chroma.2007.05.024>
- Bøhn, T., Cuhra, M., Traavik, T., Sanden, M., Fagan, J., & Primicerio, R. (2014). Compositional differences in soybeans on the market: Glyphosate accumulates in Roundup Ready GM soybeans. *Food Chemistry*, 153, 207-215. <https://doi.org/10.1016/j.foodchem.2013.12.054>
- Chen, J. Y., Chen, X. W., Lin, Y. Y., Yen, G. C., & Lin, J. A. (2021). Authentication of dark brown sugars from different processing using three-dimensional fluorescence spectroscopy. *LWT*, 150, Article 111959. <https://doi.org/10.1016/j.lwt.2021.111959>

- de Almeida, V. E., de Sousa Fernandes, D. D., Diniz, P. H. G. D., de Araújo Gomes, A., Vêras, G., Galvão, R. K. H., & Araujo, M. C. U. (2021). Scores selection via Fisher's discriminant power in PCA-LDA to improve the classification of food data. *Food Chemistry*, 363, Article 130296. <https://doi.org/10.1016/j.foodchem.2021.130296>
- Dixit, A., Antony, J. I., Sharma, N. K., & Tiwari, R. K. (2011). Soybean constituents and their functional benefits. *Research Signpost*, 661(2), 367-383.
- Esteki, M., Shahsavari, Z., & Simal-Gandara, J. (2018). Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. *Food Control*, 91, 100-112. <https://doi.org/10.1016/j.foodcont.2018.03.031>
- Esteki, M., Simal-gandara, J., Shahsavari, Z., Zandbaaf, S., & Dashtaki, E. (2018). A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control*, 93(April), 165-182. <https://doi.org/10.1016/j.foodcont.2018.06.015>
- Gomiero, T. (2018). Food quality assessment in organic vs. conventional agricultural produce: Findings and issues. *Applied Soil Ecology*, 123(February), 714-728. <https://doi.org/10.1016/j.apsoil.2017.10.014>
- Guo, Z., Huang, W., Peng, Y., Chen, Q., Ouyang, Q., & Zhao, J. (2016). Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of "Fuji" apple. *Postharvest Biology and Technology*, 115, 81-90. <https://doi.org/10.1016/j.postharvbio.2015.12.027>
- Hartman, G. L., Pawlowski, M. L., Herman, T. K., & Eastburn, D. (2016). Organically grown soybean production in the USA: Constraints and management of pathogens and insect pests. *Agronomy*, 6(1), Article 16. <https://doi.org/10.3390/agronomy6010016>
- Jawaid, S., Talpur, F. N., Sherazi, S. T. H., Nizamani, S. M., & Khaskheli, A. A. (2013). Rapid detection of melamine adulteration in dairy milk by SB-ATR-Fourier transform infrared spectroscopy. *Food Chemistry*, 141(3), 3066-3071. <https://doi.org/10.1016/j.foodchem.2013.05.106>
- Khodabakhshian, R., Bayati, M. R., & Emadi, B. (2021). An evaluation of IR spectroscopy for authentication of adulterated turmeric powder using pattern recognition. *Food Chemistry*, 364, Article 130406. <https://doi.org/10.1016/j.foodchem.2021.130406>
- Khuwijitjaru, P., Boonyapisompan, K., & Huck, C. W. (2020). Near-infrared spectroscopy with linear discriminant analysis for green "Robusta" coffee bean sorting. *International Food Research Journal*, 27(2), 287-294.
- Lakshmisha, G., Singh, V. P., Shivakumar, B. G., & Arora, A. (2012). Effect of organic and inorganic nutrients on physiological traits of soybean (*Glycine max* (L.) Merr). *Indian Journal of Plant Physiology*, 17(1), 52-56.
- Lichtenthaler, H. K., & Buschmann, C. (2001). Chlorophylls and carotenoids: Measurement and characterization by UV-VIS spectroscopy. *Current Protocols in Food Analytical Chemistry*, 1, F4.3.1-F4.3.8. <https://doi.org/https://doi.org/10.1002/0471142913.faf0403s01>
- Martins, F. C. O. L., Sentanin, M. A., & De Souza, D. (2019). Analytical methods in food additives determination: Compounds with functional applications. *Food Chemistry*, 272(April 2018), 732-750. <https://doi.org/10.1016/j.foodchem.2018.08.060>



- Masithoh, R. E., Lohumi, S., Amanah, H. Z., Yoon, W. S., & Cho, B. K. (2020). Development of multi-product calibration models of various root and tuber powders by Fourier Transform Near Infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents. *Heliyon*, 6(10), Article e05099. <https://doi.org/10.1016/j.heliyon.2020.e05099>
- Masithoh, R. E., Pahlawan, M. F. R., & Wati, R. K. (2021). Non-destructive determination of SSC and pH of banana using a modular Vis/NIR spectroscopy: comparison of partial least square (PLS) and principle component regression (PCR). In *IOP Conference Series: Earth and Environmental Science* (Vol. 752, No. 1, p. 012047). IOP Publishing. <https://doi.org/10.1088/1755-1315/752/1/012047>
- Masithoh, R. E., Rondonuwu, F., Setyabudi, F. M. C. S., & Cho, B. K. (2020). Development of calibration model for determination of sweeteners additives in Indonesia rice flour-based food by FT-NIR spectroscopy. In *IOP Conference Series: Earth and Environmental Science* (Vol. 542, No. 1, p. 012017). IOP Publishing. <https://doi.org/10.1088/1755-1315/542/1/012017>
- Masithoh, R. E., Roosmayanti, F., Rismiwandira, K., & Pahlawan, M. F. R. (2021). Detection of palm sugar adulteration by fourier transform near-infrared (FT-NIR) and fourier transform infrared (FT-IR) spectroscopy. *Sugar Tech*, 24(3), 920-929. <https://doi.org/10.1007/s12355-021-01058-3>
- Meerza, S. I. A., & Gustafson, C. R. (2019). Does prior knowledge of food fraud affect consumer behavior? Evidence from an incentivized economic experiment. *PLoS ONE*, 14(12), 1-14. <https://doi.org/10.1371/journal.pone.0225113>
- Merzlyak, M. N., Solovchenko, A. E., & Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology*, 27(2), 197-211. [https://doi.org/10.1016/S0925-5214\(02\)00066-2](https://doi.org/10.1016/S0925-5214(02)00066-2)
- Monma, M., Terao, J., Lto, M., Saito, M., & Chikuni, K. (1994). Carotenoid components in soybean seeds varying with seed color and maturation stage. *Bioscience, Biotechnology, and Biochemistry*, 58(5), 926-930. <https://doi.org/10.1271/bbb.58.926>
- Ndlovu, P. F., Magwaza, L. S., Tesfay, S. Z., & Mphahlele, R. R. (2021). Vis-NIR spectroscopic and chemometric models for detecting contamination of premium green banana flour with wheat by quantifying resistant starch content. *Journal of Food Composition and Analysis*, 102, Article 104035. <https://doi.org/10.1016/j.jfca.2021.104035>
- Nobari-Moghaddam, H., Tamiji, Z., Akbari-Lakeh, M., Khoshayand, M. R., & Haji-Mahmoodi, M. (2021). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, 107, Article 104343. <https://doi.org/10.1016/j.jfca.2021.104343>
- Pahlawan, M. F. R., Murti, B. M. A., & Masithoh, R. E. (2022). The potency of Vis/NIR spectroscopy for classification of soybean based of colour. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1018, No. 1, p. 012015). IOP Publishing.
- Pahlawan, M. F. R., Wati, R. K., & Masithoh, R. E. (2021). Development of a low-cost modular VIS / NIR spectroscopy for predicting soluble solid content of banana. In *IOP Conference Series: Earth and Environmental Science* (Vol. 644, No. 1, p. 012047). IOP Publishing. <https://doi.org/10.1088/1755-1315/644/1/012047>

- Pandiselvam, R., Mahanti, N. K., Manikantan, M. R., Kothakota, A., Chakraborty, S. K., Ramesh, S. V., & Beegum, P. P. S. (2022). Rapid detection of adulteration in desiccated coconut powder: vis-NIR spectroscopy and chemometric approach. *Food Control*, 133, Article 108588. <https://doi.org/10.1016/j.foodcont.2021.108588>
- Pereira, S. N. G., De Lima, A. B. S., Oliveira, T. D. F., Batista, A. S., Jesus, J. C. De, Ferrão, S. P. B., & Santos, L. S. (2022). Non-destructive detection of soybean oil addition in babassu oil by MIR spectroscopy and chemometrics. *LWT*, 154, Article 112857. <https://doi.org/10.1016/j.lwt.2021.112857>
- Rismiwindira, K., Roosmayanti, F., Pahlawan, M. F. R., & Masithoh, R. E. (2020). Application of fourier transform near-infrared (FT-NIR) spectroscopy for detection of adulteration in palm sugar. In *IOP Conference Series: Earth and Environmental Science* (Vol. 653, No. 1, p. 012122). IOP Publishing. <https://doi.org/10.1088/1755-1315/653/1/012122>
- Roosmayanti, F., Rismiwindira, K., & Masithoh, R. E. (2021). Detection of coconut (*Cocos nucifera*) sugar adulteration in palm (*Arenga pinnata* Merrill) sugar by fourier transform infrared (FT-IR) spectroscopy. *Food Research*, 5, 31-36.
- Sørensen, K. M., Khakimov, B., & Engelsen, S. B. (2016). The use of rapid spectroscopic screening methods to detect adulteration of food raw materials and ingredients. *Current Opinion in Food Science*, 10, 45-51. <https://doi.org/10.1016/j.cofs.2016.08.001>
- Vieira, L. S., Assis, C., de Queiroz, M. E. L. R., Neves, A. A., & de Oliveira, A. F. (2021). Building robust models for identification of adulteration in olive oil using FT-NIR, PLS-DA and variable selection. *Food Chemistry*, 345, Article 128866. <https://doi.org/10.1016/j.foodchem.2020.128866>
- Walkowiak, A., Ledziński, Ł., Zapadka, M., & Kupcewicz, B. (2019). Detection of adulterants in dietary supplements with Ginkgo biloba extract by attenuated total reflectance Fourier transform infrared spectroscopy and multivariate methods PLS-DA and PCA. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 208, 222-228. <https://doi.org/10.1016/j.saa.2018.10.008>
- Zhang, Q., Li, Q., & Zhang, G. (2012). Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection. *Spectroscopy*, 27(2), 93-105. <https://doi.org/10.1155/2012/276795>