Determination of colorant type in yellow tofu using Vis-NIR and SW-NIR spectroscopy

Laila RAHMAWATI[1]* (iD), Slamet WIDODO[2], Deni Permana KURNIADI[2], Pamungkas DAUD[2], Agus TRIYONO[3], SRIHARTI[3], Novita Dwi SUSANTI[3], Nur Kartika Indah MAYASTI[3], Ashri INDRIATI[3], Lista Eka YULIANTI[3], Devry Pramesti PUTRI[3], Seri Intan KUALA[3], Cahya Edi Wahyu ANGGARA[3], Eko Joni PRISTIANTO[2], Erry Dwi KURNIAWAN[2], Ignatius Fajar APRIYANTO[3], Dayat KURNIAWAN[2]

Abstract

A method that can detect colorant type in food is requisite against illegal practices that used non-food-grade colorants in food. In this study colorant types of yellow tofu were identified using visible near-infrared (Vis-NIR) and shortwave near-infrared spectra (SW-NIR) spectroscopy. Chemometrics analysis using partial least square-discriminant analysis (PLS-DA) and principal component analysis-linear discriminant analysis (PCA-LDA) were used to classify natural colorants, non-food-grade colorants, and food-grade colorants in yellow tofu. The best model obtained using Vis-NIR had calibration accuracy and reliability of 100% for both PCA-LDA and PLS-DA. The best model yielded by SW-NIR had calibration accuracy of 76% and 28% reliability for PLS-DA, and accuracy of 92% and 76% reliability for PCA-LDA. The result showed that Vis-NIR spectroscopy was superior to SW-NIR spectroscopy to distinguish the colorant type used in yellow tofu. Moreover, PCA-LDA yielded better accuracy and reliability compared to PLS-DA.

Keywords: food colorant; adulteration; dyes; tofu; chemometrics analysis.

Practical Application: Visible near infrared and shortwave near infrared to determine the colorant type in tofu.

1 Introduction

Food safety has become a public concern that must be ensured by the food-related industry for their products. Food-related industry play 2 roles in 17 sustainable development (SDG) goals, which are 2nd goal "Zero Hunger" that related to food production activity, and 3rd goal "Good health and wellbeing" that related to food safety. Adulteration is one of the food safety problems faced by society and food-based industries that are considered a violation of EU food law (Kucharska-Ambrożej & Karpinska, 2020). Adulteration is fraudulent practices by adding, replacing, or subtracting certain ingredients that are inferior or contain dangerous substances without the consumer's conscious (Wang et al., 2014). Deliberate adulteration was usually driven by economic reasons. However, adulteration that is motivated to harm people can also happen (Heckman et al., 1967).

Yellow tofu is a colored product that is vulnerable to fraudulent practices. Yellow tofu is a rectangular-shaped soybean-based product with a distinctive yellow color that is popular in Indonesia. The bright yellow color of tofu is the first parameter that interacts with consumers and affects consumer acceptability (Rathee & Rajain, 2019). Therefore, some parties use colorants to increase the attractiveness of the product (Zhu et al., 2015). However, not every colorant was permitted to be contained in a food product due to the chemical content or side effects that are harmful to health (Gukowsky et al., 2018). Moreover, public

concern about food safety forces the food industries to use all-natural coloring products.

Food colorants originally came from natural ingredients, such as plants, animals, or minerals. These types of colorants are called natural colorants. Another type of colorant, that is man-made from chemical material is called synthetic colorants (Mohamad et al., 2019). Natural colorants are mostly healthier than synthetic food colorants (Kamatar, 2013). However, natural colorants are not durable, unstable, and less attractive. On the other hand, synthetic food colorants, which are reliable and more attractive colorants also available in the market at a cheaper price. (Saleem et al., 2013). Based on food law, only food-grade synthetic colorants could be used for a food product. However, adulteration practice which is the counterfeiting of food colorants with a non-food-grade colorant that is dangerous for health has been reported (Nurkanti, 2009; Saleem et al., 2013; Ullah et al., 2022). Therefore, a method that can detect colorant type in food is requisite.

A method that has the potential to identify colorant adulteration is visible near-infrared (Vis-NIR) spectroscopy. Vis-NIR spectroscopy is a branch of science that studies spectra or a set of interactions between electromagnetic waves and matter in visible light (400-700 nm) and near-infrared (700-2500 nm)

region (Van Der Meer, 2018; Pahlawan & Masithoh, 2022). Vis-NIR spectroscopy is the potential to identify colorant type because color is a perception manifested in response to electromagnetic waves in the visible region (Mohamad et al., 2019). Moreover, Vis-NIR spectroscopy also uses spectra in the near-infrared region that contained information related to C-H-O-N molecules which compose organic material (Cortés et al., 2019), including natural colorants.

Vis-NIR spectroscopy offers non-destructive measurement with minimal preparation without chemical material. Spectroscopy provides technology that can measure multiple parameters at once quickly and accurately (Pahlawan et al., 2021). Vis-NIR spectroscopy is normally combined with chemometric analysis to extract information in spectra and relate the spectra with desired variables. The chemometric analysis consists of two-step, which are spectra pre-processing and multivariate modeling. Spectra pre-processing is essential to reduce noise or unusable information that could reduce model prediction performance (Mishra et al., 2020). Multivariate modeling is a statistical method used to build a prediction model (Masithoh et al., 2020).

In this study, the type of food colorants used in yellow tofu was identified using a modular type spectrometer that consists of a Vis-NIR spectrometer and a NIR spectrometer. Three types of food colorants were applied deliberately such as natural colorant (NC), food-grade synthetic colorant (FGSC), and non-food-grade synthetic colorant (FGSC). Two multivariate modeling methods were compared in this study, i.e., partial least square combined with discriminant analysis (PLS-DA) and principal component analysis combined with linear discriminant analysis (PCA-LDA).

2 Materials and methods

2.1 Sample preparation

The tofu samples were bought from the local market in Sleman, Yogyakarta, Indonesia which has a food certificate for tofu making through soybean (Glycine max L) processing by means of protein deposition, with or without the addition of other permitted ingredients. All processes were followed by Indonesian National Standard (SNI) number 01-3142-1998 about tofu making and controlled by the local government. The critical process of tofu making consist of ten stages, include selecting the raw materials and food additives, soaking, milling, cooking, flitering, clumping, pressing, slicing/cutting, packaging, and tagging. The tofu samples were then colored with three common yellow dyes which are widely used in the market. The yellow dyes used in this research were 50 gr turmeric as natural colorant (NC), 5 gr food-grade synthetic colorant (FGSC) (Sepuhan Naga Berlian, BPOM Number MD 2777-1001-2193), and 25 gr non-food-grade synthetic colorant (FGSC) (unknown colorant). Each yellow dyes treatment has three tofu which was soaked in 0.5 L warm water for 2 hrs then dried in room temperature (25-27 °C).

2.2 Spectra acquisition

Reflectance spectra were acquired by modular spectrometer that consists of 2 instrument spectrometers which are visible near-infrared spectrometer (400-1000 nm, Flame-T-VIS-NIR, Ocean Optics) and near-infrared spectrometer (1000-1700 nm,

Flame-NIR+, ocean optics). Spectra of tofu samples were collected from below with sample perpendicular to the probe. The total number of spectra obtained was 3 types of colorant × 18 samples × 10 replication = 540 spectra.

2.3 Chemometrics

Vis-NIR reflectance spectra (400-1000 nm) and NIR reflectance spectra (1000-1600 nm) obtained were compiled separately in Ms. Excel and imported to Unscrambler® X software (CAMO, Oslo, Norway) for chemometric analysis. A total of 540 spectra of each Vis-NIR spectra and NIR spectra were obtained. Total data was divided into calibration sets (1/2 samples, 270 samples) that were used to build a multivariate model and prediction sets (1/2 samples, 270 samples) that were used to validate the model built. Pre-process methods were applied to the calibration set such as area normalization, 1st and 2nd order of Savitsky-Golay's derivatives, standard normal variate, and multiple scatter correction. Two supervised multivariate modeling methods were used in this study, i.e., partial least square combined with discriminant analysis (PLS-DA) and principal component analysis combined with linear discriminant analysis (PCA-LDA). Multivariate models were built for Vis-NIR spectra and NIR spectra separately.

PLS-DA is a qualitative analysis that is constructed with partial least squares (PLS) as the basis. To perform PLS analysis for qualitative data such as colorant type, dummy numeric variables were needed. Tofu samples with a natural colorant (NC) were labeled as 1, food-grade colorant (FGC) as 2, and non-food-grade colorant (NFGC) as 3. PLS analyses were performed with spectra as independent variables (X) and colorant labels as the desired variable (Y). The predicted values between 0.5 and 1.5 will be rounded as 1; 1.5 < Y < 2.5 will be classified as 2; 2.5 < Y < 3.5 will be classified as 3. The remaining values will be classified as "Not classified" (Yangming et al., 2021).

PCA-LDA analyses were done using 7 PC scores and the Mahalanobis method. PCA-LDA is a qualitative analysis that consists of 2 steps. First, dimension reduction of spectra variables into 7 new variables called PC. Second, classification of the new variables into three types of colorants, which are NC, FGC, and NFGC. PCA-LDA was performed using spectra as the independent variable (X) and colorant type as the dependent variable (Y).

Model performance was judged by the accuracy and reliability values. Accuracy and reliability were calculated using Equation 1 and Equation 2, respectively. Accuracy and reliability equations were modified from accuracy and reliability from (Trullols et al., 2004; Vieira et al., 2021). Accuracy shows the ratio of the number of successful classifications from every class with the total samples. Reliability shows the ability of the model to predict each class correctly considering the total sample of each class independently (Saputro et al., 2022). Higher accuracy (max 100%) shows that the model has a better prediction.

$$Accuracy = \frac{n_{NC}^P + n_{NFGC}^P + n_{FGC}^P}{n_{NC} + n_{NFGC} + n_{FGC}} \times 100\% \quad (1)$$

$$Reliabilty = \left(\frac{n_{NC}^{P}}{n_{NC}} + \frac{n_{NFGC}^{P}}{n_{NFGC}} + \frac{n_{FGC}^{P}}{n_{FGC}} - 2 \right) \times 100\% \qquad (2)$$

Description:

n_{NC}^{P} = the number of correct predictions of natural colorants sample

n_{NC} = Total sample of natural colorants

n_{NFGC}^{P} = the number of correct predictions of non-food-grade colorant sample

n_{NFGC} = Total sample of non-food-grade colorant

n_{FGC}^{P} = the number of correct predictions of food-grade colorant sample

n_{FGC} = Total sample of food-grade colorant

3 Results and discussion

3.1 Yellow tofu appearance

Yellow tofu samples colored using natural colorant (NC), non-food-grade colorant (NFGC), and food-grade colorant (FGC) are shown in Figure 1. The color parameters of tofu samples are shown in Table 1. Tofu samples in Figure 1 show that each colorant has a different yellow color result, especially tofu with FGC. Based on the redness color, tofu that was colored using FGC has the highest redness value compared to other colorants. However, FGC tofu has the lowest greenness compared to NC and NFGC.
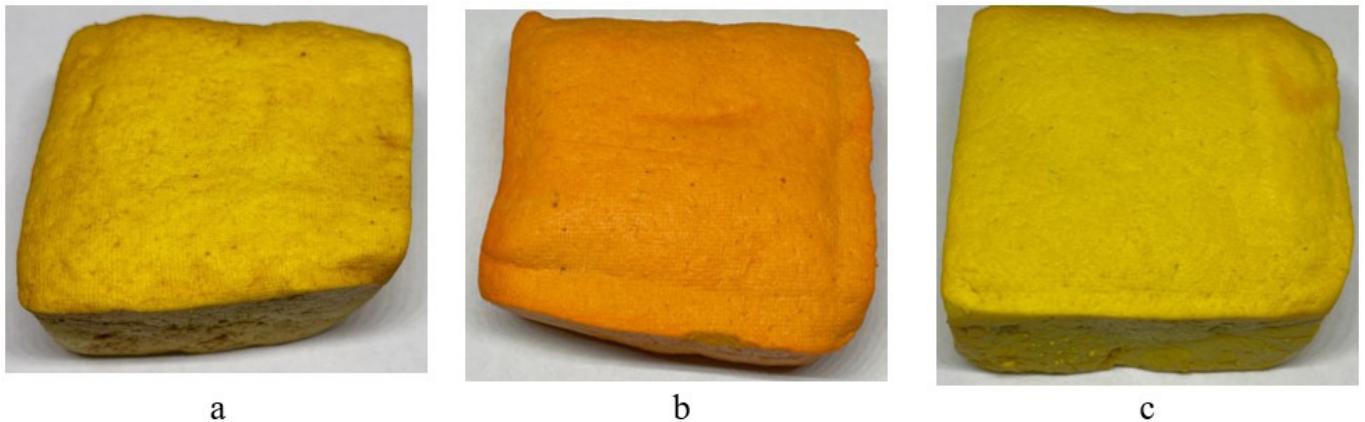
3.2 Spectra of yellow tofu

Visible near Infrared Spectra (Vis-NIR)

The Vis-NIR average raw spectra and Savitsky-Golay's preprocessed spectra of tofu were shown in Figure 2. Raw spectra of tofu with natural colorant show low reflectance at 400-500 nm, which is affected by yellow color pigment, and carotenoids (Cortés et al., 2016). Another valley was noticeable at around 970-1000 nm, which related to 2nd overtones of O-H stretching in water (Golic et al., 2003). Raw spectra of Non-food-grade

Table 1. Yellow tofu color.

Colorant type	R (Red)	G (Green)	B (Blue)
Natural colorant (NC)	207 ± 13.3	165 ± 21.1	3 ± 0.5
Food-grade colorant (FGC)	214 ± 3.5	126 ± 0.6	5 ± 2.6
Non-Food-grade colorant (NFGC)	202 ± 4	172 ± 4.2	3 ± 1

Figure 1. Yellow-colored tofu using a) natural colorant (NC), b) Food-grade colorant (FGC), c) Non-food-grade colorant (NFGC).
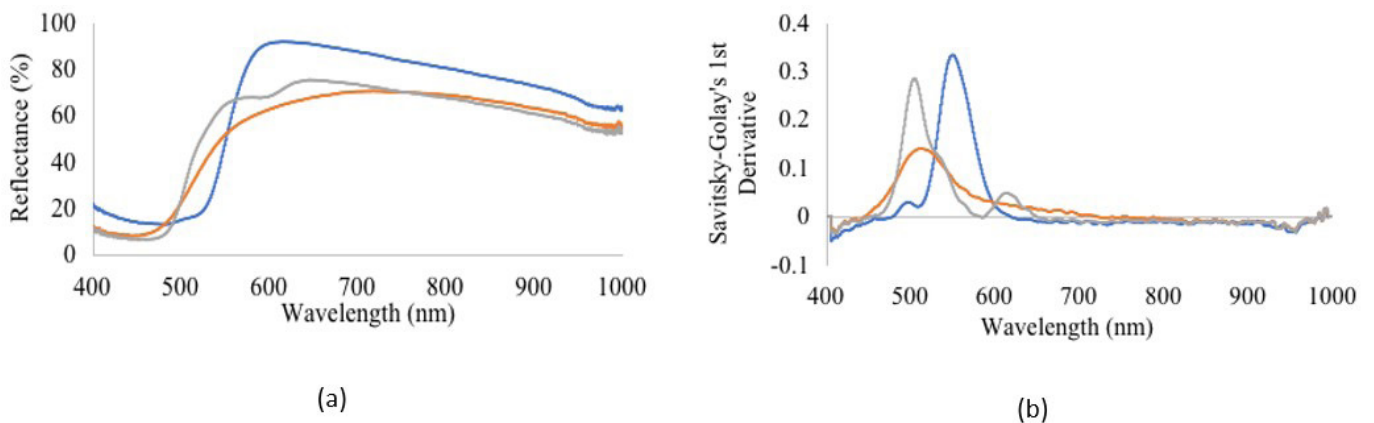
Figure 2. Visible-Near Infrared Average Spectra of Yellow Tofu, a) Raw and b) Savitsky-Golay's 1st Derivative.

colorant (NFGC) showed an almost similar trend, the difference was that the NFGC reflectance after 500 nm was still low up to 520 nm. Based on (Merzlyak et al., 2003), wavelength around 520 nm is still correlated with carotenoid pigments. Raw spectra of tofu with a food-grade colorant (FGC) are also similar to raw spectra of tofu with NC, the difference is that there is the absorbance at around 620 nm, which might be correlated with anthocyanin or red color pigments.

After raw spectra were preprocessed using Savitsky-Golay's 1st derivative technique, several strong peaks were visible at each spectra (Figure 2b). Tofu with NC shows spectra peak at around 520 nm, which correlated with carotenoid content. Tofu with FGC shows spectra peaks at around 500 nm and 620, which correlated with Tofu with NFGC shows spectra peaks around 500 and 550 nm, which correlated with carotenoids and anthocyanin contents, respectively (Merzlyak et al., 2003). Based on these, it can be concluded that despite having the same color, each colorant type has a different fingerprint Vis-NIR spectra which correlated with color pigment and color characteristic.

Shortwave near-infrared spectra (SW-NIR)

The shortwave near infrared average raw spectra of tofu with three types of colorants is shown in Figure 3. Raw spectra of yellow-colored tofu (Figure 3a) show a similar pattern to the three types of colorants. A valley in around 970 nm is corresponding to 2nd overtones of O-H stretching (Fernández-Novales et al., 2019). An absorbance at around 1150 nm is associated with 2nd overtones of C-H stretching (Golic et al., 2003). A strong absorbance that is characterized by a deep valley of reflectance spectra at around 1390 nm is correlated with 1st overtones of C-H stretching (Golic et al., 2003). Absorbance at around 1690 nm might be affected by 1st overtones of CH_3 (Wilson et al., 2015)

Pre-processed spectra of SW-NIR spectra using Savitsky-Golay's 1st derivative methods are shown in Figure 3b. Preprocessed spectra show similar peaks and valleys to raw spectra. However, a valley appears at around 1330 nm, which might be associated with O-H stretching in water and or C-H stretching in sugar (Rongtong et al., 2018). Based on the result obtained, it shows that tofu with three types of colorants has similar spectra.
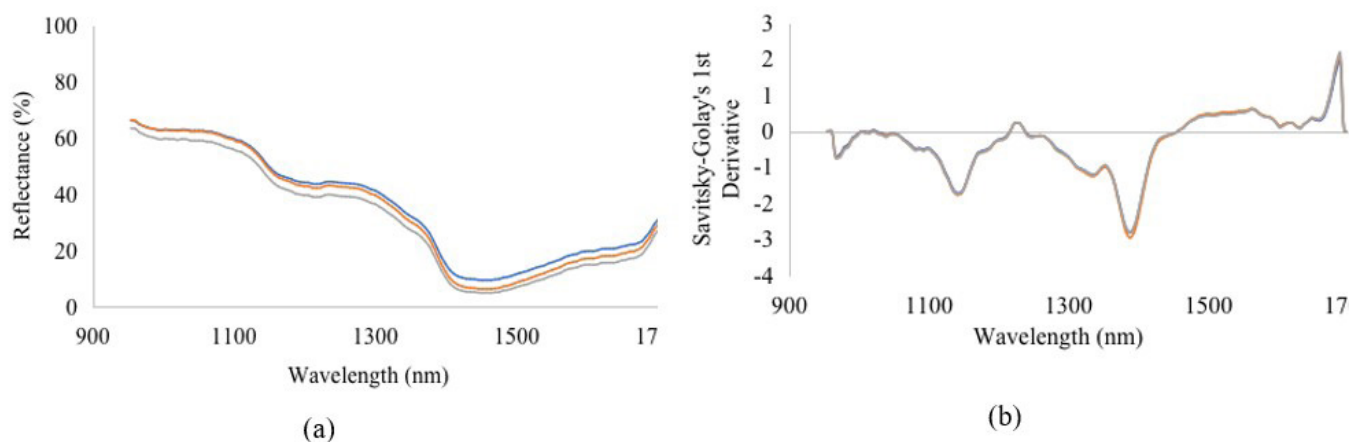
3.3 PCA-LDA modeling

Principal component analysis (PCA) is a multivariate analysis that is used to reduce spectra dimension and create new variables called principal components (PCs) without losing important information (Sabzi & Arribas, 2018; Vasques et al., 2008). In this study, 7 PCs were built from 3188 variables of spectra for Vis-NIR and 128 variables for SW-NIR. However, PCA is an unsupervised method. Therefore, to relate PCs with desired variables, LDA is used. Linear discriminant analysis (LDA) is a supervised method that was used to classify category variables such as colorant type. LDA can minimize the variance within the same colorant type while maximizing the variance between different colorant types. LDA began with building a set of orthogonal linear discriminant functions to differentiate each colorant type. Euclidean distances were calculated from the data point to each class centroid and projected to the subspace defined by a subset of the linear functions built. Then the samples were classified into the nearest class (Deng et al., 2018).

PCA-LDA using Vis-NIR spectra

Table 2 shows PCA-LDA model performance to identify the colorant type of yellow tofu. The result shows that raw spectra and all pre-processed spectra produce a model with very satisfactory performance. All models yielded a perfect model that has 100% accuracy and 100% reliability of calibration. Prediction performance also demonstrates that the model can classify each type of colorant perfectly, except for the raw spectra model. In the raw spectra model, there is only one miss classification, which identifies natural colorant tofu as food-grade colorant tofu.

The result obtained demonstrates that Vis-NIR spectroscopy can identify colorant type in tofu with a satisfying result. To explore which wavelength contributes the most to the model, the loading plot of the 7 PCs used is observed. Valleys and peaks in the loading plot show which wavelength contribute to each PC. Loading plots of the best PCA-LDA model using normalized Vis-NIR spectra were shown in Figure 4a. PCA's loadings show that colorant fingerprint wavelength also appears in the loading plots. Loading plot spectra show that carotenoids greatly contributed to the model. Valley and peaks correlated

Figure 3. Shortwave Near Infrared Average Spectra of Yellow Tofu, a) Raw and b) Savitsky-Golay's 1st Derivative.

Table 2. PCA-LDA performance of Vis-NIR spectroscopy.

True Class	Predicted Class	n	Calibration						n	Prediction					
			Predicted sample							Predicted sample					
			Raw	Norm	SGD1	SGD2	SNV	MSC		Raw	Norm	SGD1	SGD2	SNV	MSC
NC	NC	91	91	91	91	91	91	91	90	89	90	90	90	90	90
NC	FGC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NC	NFGC	0	0	0	0	0	0	0	0	1	0	0	0	0	0
FGC	NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGC	FGC	90	90	90	90	90	90	90	90	90	90	90	90	90	90
FGC	NFGC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFGC	NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFGC	FGC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFGC	NFGC	89	89	89	89	89	89	89	90	90	90	90	90	90	90
	Acc (%)		100	**100**	100	100	100	100		100	**100**	100	100	100	100
	Reliability (%)		100	**100**	100	100	100	100		99	**100**	100	100	100	100

Note: NC = natural colorant, NFGC = non-food-grade colorant, FGC = food-grade-colorant, n = number of sample, Acc = accuracy, Raw = raw data, Norm = normalized data, SGD1 = Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative, SNV = standard normal variate, MSC = multiple scatter correction.
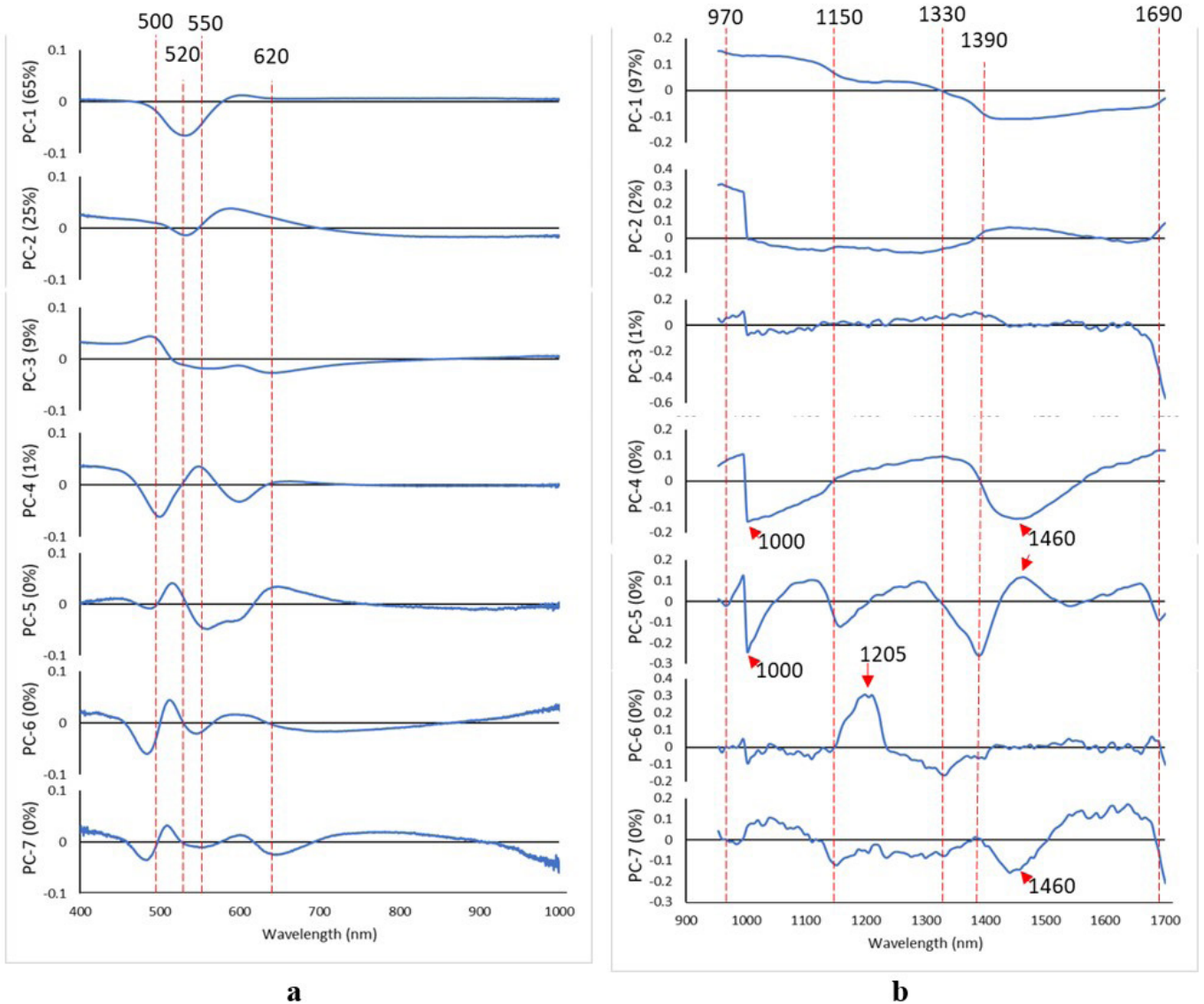
Figure 4. PCA's loading plots from the best PCA-LDA model of a) Vis-NIR using normalize spectra, b) SW-NIR using normalized spectra.

with carotenoids (480-520 nm) appear at every PC of the loading plot. In addition, anthocyanins also contribute to the model. This is evidenced by the anthocyanin wavelengths (550 & 620 nm) which can be seen in PC-4, PC-5, and PC-7.

PCA-LDA using SW-NIR spectra

PCA-LDA model performances using SW-NIR spectra are shown in Table 3. The result shows that PCA-LDA using SW-NIR spectra yielded in an inferior model compared to PCA-LDA using Vis-NIR spectra. The best model performance was acquired using normalized spectra of SW-NIR with an accuracy of calibration and prediction were 92% and 90%, respectively. Based on the accuracy, the model can be considered a good model, since it has > 90% accuracy. Accuracy shows the model's ability to correctly classify colorant type, considering the total data used (Vieira et al., 2021).

The model reliability does not show satisfactory results. Reliability indicates the ability of the model to classify each type of colorant accurately, by taking into account the total data from each class separately. To get high reliability, each class should have a minimal miss classification (Saputro et al., 2022). The best SW-NIR PCA-LDA obtained using normalized spectra has a reliability of calibration and prediction of 76% and 69%, respectively. A total of 22 missed classifications were generated from the calibration set and 28 missed classifications were generated from the prediction set. The most missed classification was produced by NC and the least missed classification was produced by NFGC, for both calibration and prediction. However, the most fatal miss classification is FGC which is classified as NC or NFGC, with a total of 12 miss classifications in the calibration and 11 miss classifications in the prediction.

Loading plots of the SW-NIR PCA-LDA model were shown in Figure 4b. Figure 4b also shows several absorbances that showed in the SW-NIR spectra. However, only a few strong peaks and valleys appear around those absorbances, such as 970 nm at PC-1 and PC-2, and 1390 nm valley at PC-5. Instead, new valleys and peaks appeared on the loading plots. Valley at 1000 nm appeared at PC-4 and PC-5, which might be related to 3rd overtone of the C-H bond (Correia et al., 2018). Strong

peaks were visible at around 1200-1210 nm on PC-6, which was affected by 2nd overtone of C-H bond stretching (Wilson et al., 2015). Peaks and valleys at around 1460 nm were associated with 1st overtone of O-H bonds (Nasir et al., 2019).

3.4 PLS-DA Modeling

Partial least square (PLS) is a well-known multivariate analysis method that is used in spectroscopy. As a supervised method, PLS can be used to relate spectra variables with desired variables. However, PLS is limited to quantitative parameters. Hence, the desired variable in a form of qualitative parameters should be converted to quantitative variables using labels. In this study, the desired variable is colorant types which is a qualitative variable. Therefore, natural colorant, non-food-grade colorant, and food-grade colorant were labeled as 1, 2, and 3, respectively.

The PLS-DA model was constructed with important 3 steps from a combination of PLS modeling and discriminant analysis. The first step was dimension reduction of spectra variables considering the desired variables. This step was performed to reduce spectra dimension, noise, and computation time (Vasques et al., 2008). The second step was the regression step where the orthogonalized PLSR algorithm were performed using the calibration set, with colorant label as the Y-variable (Abdi, 2010). The third step was rounding the predicted values into each category. The predicted values between 0.5 and 1.5 will be rounded as 1; $1.5 < Y < 2.5$ will be classified as 2; $2.5 < Y < 3.5$ will be classified as 3. The remaining values will be classified as "Not classified" (Yangming et al., 2021).

PLS-DA using Vis-NIR spectra

PLS-DA model performances using Vis-NIR spectra are shown in Table 4. Same as PCA-LDA, the PLS-DA model using Vis-NIR spectra produced perfect models with 100% accuracy and reliability. The best model. Two calibration models with maximum accuracy and reliability were yielded by raw spectra and SGD1 spectra produced the model with the highest accuracy and reliability. However, the SGD1 model was chosen as the best model because it has a better prediction performance compared to the raw spectra model. SGD1 model has 100% accuracy and 99%

Table 3. PCA-LDA performance of SW-NIR spectroscopy.

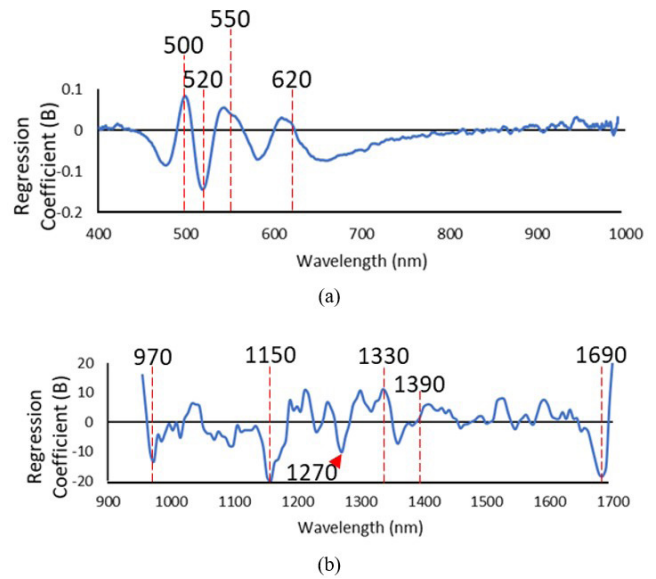| True Class | Predicted Class | Calibration | | | | | | | Prediction | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | n | Predicted sample | | | | | | n | Predicted sample | | | | | |
| | | | Raw | Norm | SGD1 | SGD2 | SNV | MSC | | Raw | Norm | SGD1 | SGD2 | SNV | MSC |
| NC | NC | 90 | 76 | 81 | 58 | 71 | 67 | 66 | 90 | 67 | 74 | 53 | 71 | 67 | 67 |
| NC | FGC | 0 | 7 | 5 | 3 | 16 | 1 | 1 | 0 | 11 | 9 | 5 | 16 | 6 | 6 |
| NC | NFGC | 0 | 7 | 4 | 29 | 3 | 22 | 23 | 0 | 12 | 7 | 32 | 3 | 17 | 17 |
| FGC | NC | 0 | 4 | 8 | 4 | 3 | 10 | 10 | 0 | 3 | 7 | 2 | 3 | 7 | 7 |
| FGC | FGC | 90 | 77 | 78 | 78 | 82 | 77 | 77 | 90 | 75 | 79 | 78 | 80 | 75 | 75 |
| FGC | NFGC | 0 | 9 | 4 | 8 | 5 | 3 | 3 | 0 | 12 | 4 | 10 | 7 | 8 | 8 |
| NFGC | NC | 0 | 0 | 0 | 2 | 0 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 4 | 4 |
| NFGC | FGC | 0 | 1 | 1 | 3 | 1 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |
| NFGC | NFGC | 90 | 89 | 89 | 85 | 89 | 85 | 85 | 90 | 90 | 89 | 89 | 90 | 84 | 84 |
| | Acc (%) | | 90 | **92** | 82 | 90 | 85 | 84 | | 86 | **90** | 81 | 89 | 84 | 84 |
| | Reliability (%) | | 69 | **76** | 46 | 69 | 54 | 53 | | 58 | **69** | 44 | 68 | 51 | 51 |

Note: NC = natural colorant, NFGC = non-food-grade colorant, FGC = food-grade-colorant, n = number of sample, Acc = accuracy, Raw = raw data, Norm = normalized data, SGD1 = Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative, SNV = standard normal variate, MSC = multiple scatter correction.

reliability with 1 misclassification, while the raw spectra model has 99% accuracy and 98% reliability with 2 misclassifications.

Important wavelengths that contribute to the model were shown in regression coefficient (B). The regression coefficient (B) or often called the beta coefficient shows the relationship between each wavelength to the desired variable (Kusumaningrum et al., 2018). The beta coefficient of the best PLS-DA model using Vis-NIR spectroscopy were shown in Figure 5. Similar to the loading plots in PCA-LDA, the beta coefficient shows that carotenoids and anthocyanins make big contributions to the model. Carotenoids' peaks and valleys that are shown in the beta coefficient are 480 nm, 500 nm, and 520 nm. Anthocyanin's peaks and valleys that appear in the beta coefficient are 550 nm, 580 nm, and 610 nm.

PLS-DA using Vis-NIR spectra

Table 5 shows PLS-DA model performance to classify colorant type in yellow tofu using SW-NIR spectroscopy. The best performance of PLS-DA using SW-NIR was yielded by SNV spectra. However, the model had the worst performance compared to the PCA-LDA model using Vis-NIR and SW-NIR, also the PLS-DA model using Vis-NIR spectra. Based

(a)

(b)

Figure 5. Regression coefficient (B) from the best PLSR model of a) Vis-NIR using Savitsky-Golay's 1st derivative spectra, and b) SW-NIR using standard normal variate spectra.

Table 4. PLS-DA performance of Vis-NIR spectroscopy.

True Class	Predicted Class	n	Calibration						n	Prediction					
			Predicted sample							Predicted sample					
			Raw	Norm	SGD1	SGD2	SNV	MSC		Raw	Norm	SGD1	SGD2	SNV	MSC
NC	NC	91	91	90	91	91	91	90	90	89	89	89	90	90	89
NC	FGC	0	0	0	0	0	0	0	0	1	0	1	0	0	0
NC	NFGC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGC	NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FGC	FGC	90	90	90	90	90	90	90	90	90	90	90	90	90	90
FGC	NFGC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFGC	NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFGC	FGC	0	0	0	0	1	0	0	0	0	0	0	1	0	0
NFGC	NFGC	89	89	89	89	88	88	88	90	89	89	90	89	90	89
Not classified		0	0	1	0	0	1	2	0	1	2	0	0	0	2
Acc (%)			100	100	**100**	100	100	99		99	99	**100**	100	100	99
Reliability (%)			100	99	**100**	99	99	98		98	98	**99**	99	100	98

Note: NC = natural colorant, NFGC = non-food-grade colorant, FGC = food-grade-colorant, n = number of sample, Acc = accuracy, Raw = raw data, Norm = normalized data , SGD1 = Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative, SNV = standard normal variate, MSC = multiple scatter correction.

Table 5. PLS-DA performance of SW-NIR spectroscopy.

True Class	Predicted Class	n	Calibration						n	Prediction					
			Predicted sample							Predicted sample					
			Raw	Norm	SGD1	SGD2	SNV	MSC		Raw	Norm	SGD1	SGD2	SNV	MSC
NC	NC	91	51	55	57	61	60	60	90	51	53	57	53	63	63
NC	FGC	0	40	35	34	30	29	29	0	38	35	31	34	24	24
NC	NFGC	0	0	0	0	0	0	0	0	0	0	1	1	1	1
FGC	NC	0	11	9	9	7	3	4	0	15	6	12	7	8	8
FGC	FGC	90	79	78	78	83	86	85	90	74	82	77	82	81	80
FGC	NFGC	0	0	3	3	0	1	1	0	1	2	1	1	1	2
NFGC	NC	0	0	1	1	0	1	1	0	1	0	0	0	0	1
NFGC	FGC	0	34	29	32	23	26	31	0	33	32	32	24	28	28
NFGC	NFGC	89	52	57	55	61	59	55	90	52	55	56	62	59	58
Not classified		0	3	3	1	5	5	4	0	5	5	3	6	5	5
Acc (%)			67	70	70	76	**76**	74		66	70	70	73	**75**	74
Reliability (%)			2	11	11	28	**28**	22		-3	11	11	19	**26**	23

Note: NC = natural colorant, NFGC = non-food-grade colorant, FGC = food-grade-colorant, n = number of sample, Acc = accuracy, Raw = raw data, Norm = normalized data, SGD1 = Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative, SNV = standard normal variate, MSC = multiple scatter correction.

on Figure 5b, several peaks and valleys were visible in the beta coefficient plot. Among all the valleys and peaks, the strongest valleys that have the lowest regression coefficient were 1150 nm and 1690 nm, which were affected by the 2nd overtone and 1st overtone of C-H stretching, respectively.

4 Conclusion

Food colorant adulteration is a dangerous practice that could harm human health. Colorant type in yellow tofu can be distinguished using spectroscopy combined with chemometrics analysis. The result showed that Vis-NIR spectroscopy yielded better performances compared to SW-NIR spectroscopy. Furthermore, the PCA-LDA model produces better performances compared to the PLS-DA model. The best model to classify colorant type in tofu obtained by Vis-NIR spectroscopy had calibration performances of 100% accuracy and 100% reliability for both the PLS-DA and PCA-LDA models. The best model obtained by SW-NIR spectroscopy had calibration accuracy of 76% and 28% reliability for PLS-DA, and 92% accuracy and 76% reliability for the PCA-LDA model.

Conflict of interest

The authors declare that we have no conflict of interest.

Acknowledgements

References

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 97-106. http://dx.doi.org/10.1002/wics.51.

Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R. T., Aquino, L. F. M., Filgueiras, P. R., Lacerda, V. Jr., & Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, 176, 59-68. http://dx.doi.org/10.1016/j.talanta.2017.08.009. PMid:28917795.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., & Talens, P. (2019). Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: a review. *Trends in Food Science & Technology*, 85, 138-148. http://dx.doi.org/10.1016/j.tifs.2019.01.015.

Cortés, V., Ortiz, C., Aleixos, N., Blasco, J., Cubero, S., & Talens, P. (2016). A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy. *Postharvest Biology and Technology*, 118, 148-158. http://dx.doi.org/10.1016/j.postharvbio.2016.04.011.

Deng, F., Chen, W., Wang, J., & Wei, Z. (2018). Fabrication of a sensor array based on quartz crystal microbalance and the application in egg shelf life evaluation. *Sensors and Actuators. B, Chemical*, 265, 394-402. http://dx.doi.org/10.1016/j.snb.2018.03.010.

Fernández-Novales, J., Garde-Cerdán, T., Tardáguila, J., Gutiérrez-Gamboa, G., Pérez-Álvarez, E. P., & Diago, M. P. (2019). Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta*, 199, 244-253. http://dx.doi.org/10.1016/j.talanta.2019.02.037. PMid:30952253.

Golic, M., Walsh, K., & Lawson, P. (2003). Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. *Applied Spectroscopy*, 57(2), 139-145. http://dx.doi.org/10.1366/000370203321535033. PMid:14610949.

Gukowsky, J. C., Xie, T., Gao, S., Qu, Y., & He, L. (2018). Rapid identification of artificial and natural food colorants with surface enhanced Raman spectroscopy. *Food Control*, 92, 267-275. http://dx.doi.org/10.1016/j.foodcont.2018.04.058.

Heckman, J. J., Pinto, R., & Savelyev, P. A. (1967). Food integrity handbook. *Angewandte Chemie International Edition*, 6(11), 951-952.

Kamatar, M. (2013). Natural Food colouring: a healthier alternative to artificial food colouring. In *Proceedings of the Global Milling Conference: Safety, Sustainability and Food Supply for 21st Centuryl*. Chennai: Perendale Publisher.

Kucharska-Ambrożej, K., & Karpinska, J. (2020). The application of spectroscopic techniques in combination with chemometrics for detection adulteration of some herbs and spices. *Microchemical Journal*, 153, 104278. http://dx.doi.org/10.1016/j.microc.2019.104278.

Kusumaningrum, D., Lee, H., Lohumi, S., Mo, C., Kim, M. S., & Cho, B. K. (2018). Non-destructive technique for determining the viability of soybean (Glycine max) seeds using FT-NIR spectroscopy. *Journal of the Science of Food and Agriculture*, 98(5), 1734-1742. http://dx.doi.org/10.1002/jsfa.8646. PMid:28858390.

Masithoh, R. E., Lohumi, S., Yoon, W. S., Amanah, H. Z., & 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), e05099. http://dx.doi.org/10.1016/j.heliyon.2020.e05099. PMid:33134571.

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. http://dx.doi.org/10.1016/S0925-5214(02)00066-2.

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *Trends in Analytical Chemistry*, 132, 116045. http://dx.doi.org/10.1016/j.trac.2020.116045.

Mohamad, M. F., Dailin, D. J., Gomaa, S. E., Nurjayadi, M., & El Enshasy, H. (2019). Natural colorant for food: alternative a healthy. *International Journal of Scientific and Technology Research*, 8(11), 3161-3166.

Nasir, V., Nourian, S., Zhou, Z., Rahimi, S., Avramidis, S., & Cool, J. (2019). Classifcation and characterization of thermally modifed timber using visible and near-infrared spectroscopy and artifcial neural networks: a comparative study on the performance of diferent NDE methods and ANNs. *Wood Science and Technology*, 53(5), 1093-1109. http://dx.doi.org/10.1007/s00226-019-01120-0.

Nurkanti, M. (2009). Analisis secara biokimia methanyl yellow pada tahu yang beredar di pasar tradisional kodya bandung. In *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA* (pp. 121-129). Yogyakarta: Universitas Negeri Yogyakarta.

Pahlawan, M. F. R., & Masithoh, R. E. (2022). Vis-NIR Spectroscopy and PLS-Da model for classification of Arabica and robusta roasted coffee bean. *Advances in Science and Technology (Owerri, Nigeria)*, 115, 45-52. http://dx.doi.org/10.4028/p-60bbc9.

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. *IOP Conference Series. Earth and Environmental Science*, 644(1), 012047. http://dx.doi.org/10.1088/1755-1315/644/1/012047.

Rathee, R., & Rajain, P. (2019). Role colour plays in influencing consumer behaviour. *International Research Journal of Business Studies*, 12(3), 209-222. http://dx.doi.org/10.21632/irjbs.12.3.209-222.

Rongtong, B., Suwonsichon, T., Ritthiruangdej, P., & Kasemsumran, S. (2018). Determination of water activity, total soluble solids and moisture, sucrose, glucose and fructose contents in osmotically dehydrated papaya using near-infrared spectroscopy. *Agriculture and Natural Resources (Bangkok)*, 52(6), 557-564. http://dx.doi.org/10.1016/j.anres.2018.11.023.

Sabzi, S., & Arribas, J. I. (2018). A visible-range computer-vision system for automated, non-intrusive assessment of the pH value in Thomson oranges. *Computers in Industry*, 99, 69-82. http://dx.doi.org/10.1016/j.compind.2018.03.016.

Saleem, N., Nasreen Umar, Z., & Ismat khan, S. (2013). Survey on the use of synthetic food colors in food samples procured from different educational institutes of Karachi city. *Journal of Tropical Life Science*, 3(1), 1-7. http://dx.doi.org/10.11594/jtls.03.01.01.

Saputro, D., Priambodo, D. C., Pahlawan, M. F. R., & Masithoh, R. E. (2022). Classification of cocoa beans based on fermentation level using PLS-DA combined with Visible Near-infrared (VIS-NIR) spectroscopy. In *2nd International Conference on Smart and Innovative Agriculture* (pp. 100-106). Yogyakarta: Atlantis Press. http://dx.doi.org/10.2991/absr.k.220305.015.

Trullols, E., Ruisánchez, I., & Rius, F. X. (2004). Validation of qualitative analytical methods. *Trends in Analytical Chemistry*, 23(2), 137-145. http://dx.doi.org/10.1016/S0165-9936(04)00201-8.

Ullah, A., Chan, M. W. H., Aslam, S., Khan, A., Abbas, Q., Ali, S., Ali, M., Hussain, A., Mirani, Z. A., Sibt-E-Hassan, S., Kazmi, M. R., Ali, S., Hussain, S., & Khan, A. M. (2022). Banned Sudan dyes in spices available at markets in Karachi, Pakistan. *Food Additives & Contaminants: Part B*. In press. PMid:35909386.

Van Der Meer, F. (2018). Near-infrared laboratory spectroscopy of mineral chemistry: a review. *International Journal of Applied Earth Observation and Geoinformation*, 65, 71-78. http://dx.doi.org/10.1016/j.jag.2017.10.004.

Vasques, G. M., Grunwald, S., & Sickman, J. O. (2008). Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma*, 146(1–2), 14-25. http://dx.doi.org/10.1016/j.geoderma.2008.04.007.

Vieira, L. S., Assis, C., Queiroz, M. E. L. R., Neves, A. A., & 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, 128866. http://dx.doi.org/10.1016/j.foodchem.2020.128866. PMid:33348130.

Wang, N., Zhang, X., Yu, Z., Li, G., & Zhou, B. (2014). Quantitative analysis of adulterations in oat flour by FT-NIR spectroscopy, incomplete unbalanced randomized block design, and partial least squares. *Journal of Analytical Methods in Chemistry*, 2014, 393596. http://dx.doi.org/10.1155/2014/393596. PMid:25143857.

Wilson, R. H., Nadeau, K. P., Jaworski, F. B., Tromberg, B. J., & Durkin, A. J. (2015). Review of short-wave infrared spectroscopy and imaging methods for biological tissue characterization. *Journal of Biomedical Optics*, 20(3), 030901. http://dx.doi.org/10.1117/1.JBO.20.3.030901. PMid:25803186.

Yangming, H., Yue, H., Xiangzhong, S., Jingxian, G., Yanmei, X., & Shungeng, M. (2021). Comparison of a novel PLS1-DA, traditional PLS2-DA and assigned PLS1-DA for classification by molecular spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 209, 104225. http://dx.doi.org/10.1016/j.chemolab.2020.104225.

Zhu, Y., Zhang, L., & Yang, L. (2015). Designing of the functional paper-based surface-enhanced Raman spectroscopy substrates for colorants detection. *Materials Research Bulletin*, 63, 199-204. http://dx.doi.org/10.1016/j.materresbull.2014.12.004.