Identification of Maturity Stage of Cacao using Visible Near Infrared (Vis-NIR) and Shortwave Near Infrared (SW-NIR) Reflectance Spectroscopy

Riana Listanti[1], *Rudiati* Evi Masithoh[1*], *Arifin* Dwi Saputro[1], *Hanim* Zuhrotul Amanah[1]

[1] Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Jl. Flora No. 1, Bulaksumur, Yogyakarta 55281, Indonesia

Abstract. Choosing the cacao maturity stage is essential for producing high-quality cacao beans. Identifying indicators of the maturity level of cacao is a complex task because these fruits do not exhibit the characteristics of other fruits during the ripening period. Generally, cacao maturity is determined manually based on the estimated daily harvest date using sensory observation, which is marked by changes in the color of the cacao skin. This is certainly inaccurate because visual assessment is only performed subjectively. This is inaccurate because visual assessment is only performed subjectively, which is not in line with the demands of Industrial Revolution 4.0, which is a fast and accurate technology for sorting cacao. In this study, cacao maturity was identified using visible (350-1000 nm) and shortwave near-infrared spectra (SW-NIR) spectroscopy (1000-1600 nm). Chemometric analysis using principal component analysis-linear discriminant analysis (PCA-LDA) was used to classify cacao maturity. The results showed that SW-NIR spectroscopy yielded better performances with calibration and prediction accuracy of 92,50% and 85% using Savitzky–s 1st derivative (SGD1) spectra compared to Vis-NIR spectroscopy had calibration and prediction accuracies of 90% and 86% using raw spectra for PCA-LDA model.

Keywords: cacao, maturity, PCA-LDA, spectroscopy, SW-NIR, Vis-NIR

1 Introduction

Choosing the cacao maturity stage is essential for producing the highest quality cacao beans because, as the fruit ripens, compounds such as sugars, organic acids, methylxanthines, polyphenols, and proteins, all of which are crucial for the production of aroma precursors, are formed or altered [1]. Therefore, the correct level of maturity of cocoa pods and their homogeneity will contribute to good processing and, ultimately, to the quality of cocoa beans.

Ripe cacao is characterized by a change in the color of the cacao skin. In a comprehensive review article on chocolate manufacture [3], the authors stated that ripe cacao can be identified by its outer color, which can be yellow, red, or purple, depending on the variety. Furthermore, cacao should be harvested as it reaches optimum ripeness to prevent disease attacks and the possibility of germination of cacao beans [3].

Identifying indicators of the maturity level of cacao is a complex task because these fruits do not exhibit the characteristics of other fruits during the ripening period. Generally, cocoa fruit maturity is determined manually based on the estimated daily harvest date using sensory observation, which is marked by changes in the color of the cacao skin. This is certainly inaccurate because visual assessment is only performed subjectively. The demands of Industrial Revolution 4.0 require fast and accurate technology for sorting cacao. Therefore, a quick, non-destructive, and accurate evaluation is required to classify the maturity level of cocoa pods during the harvesting period. A comprehensive review [4] concluded that all reviewed applications of NIR spectroscopy in combination with chemometric analysis can play a valuable role in the food industry.

Applications of NIR spectroscopy for on-line and in-line food and beverage quality monitoring have been extensively reviewed by [5] and according to the authors, NIR spectroscopy is one of the most promising and widely accepted as one of on-line or in-line process control technology for non-destructive, reliable and accurate monitoring of chemical and physical parameters in food processing[6]. NIR spectroscopy provides analytical profiles of food products, which can be used to characterize and assess their authenticity [7]. In other studies, NIR spectroscopy has been widely used to detect the presence of fats from sources [8], identify bioactive compounds [9], [10], identify specific content for chocolate quality control or authentication [11], investigate chocolate quality parameters and predict cacao content in chocolate [12], and predict cacao solid content [13].

Accordingly, NIR spectroscopy offers a credible alternative for assessing the integrity and quality of cacao and their derivatives [14]. García-Muñoz et al. [15] recently developed a color guide to evaluate the maturity of cocoa clones using digital image processing methods. However, no studies have investigated the

* Corresponding author: evi@ugm.ac.id

application of NIR spectroscopy to nondestructively screen and classify the maturity levels of cacao. Therefore, the main objective of this research is to classify the maturity level of cacao using non-destructive methods, especially Vis-NIR and SW-NIR spectroscopy, combined with principal component analysis linear discriminant analysis (PCA-LDA). Principal component analysis (PCA), an unsupervised technique frequently used for dimension reduction and data visualization [16] and PCA combined with linear DA (PCA-LDA), a supervised technique used to classify data with a predetermined class [17] were the discrimination methods used in this study. The main objective of classifying the maturity level of cacao using PCA-LDA is to perform rapid automated sorting using NIR spectroscopy. With NIR analysis, information about the chemical attributes of cacao beans can be obtained quickly without damaging the beans, thus enabling a more efficient and accurate sorting process. When cocoa passes through an automatic sorting machine, it categorizes cocoa beans based on their maturity level. Immature cocoa beans may not be of good quality and can be directly processed into the final product. By separating immature cocoa beans, they can be fermented to improve the quality and flavor characteristics of the resulting chocolate. The long-term goal is that by using NIR spectroscopy technology and automated sorting machines, chocolate manufacturers and the cocoa industry can improve production efficiency, reduce wastage, and produce better quality and more consistent products.

2 Material and methods

2.1 Samples

The main material used for this research was 20 ripe lindak pods harvested from a garden in Gunung Kidul, Yogyakarta, Indonesia. The cacao used in this study had an average fruit length of 25 cm, average diameter of 9 cm, and average number of seeds per cacao of 40 cacao beans, with an average shell thickness of 1.40 cm. It is also known that the cacao beans produced from this cacao are generally included in the criteria for cacao beans with quality A to AA if measured by the number of beans per 100 g.

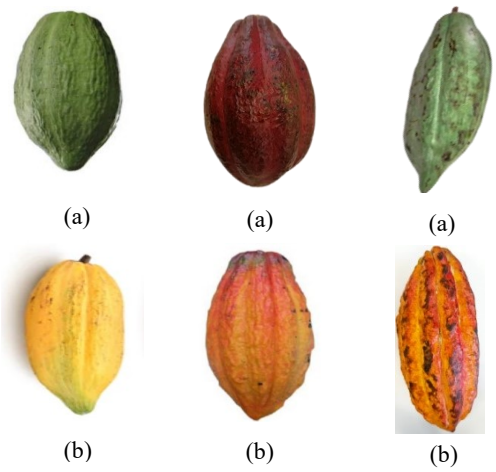


Figure 1. Cacao of different maturity levels used in this research (a) immature, (b) mature. The ripe cacao are greenish yellow or reddish orange and purple in color, with an average harvest age of around 5-6 months from the time the plants flower

2.2 Spectra Measurements

A modular spectrometer was used to measure the Vis-NIR and SW-NIR spectra of all cacao samples in reflectance mode. A visible/near-infrared spectrometer with a wavelength range of 400–1000 nm (Ocean Optics, Flame-T (Toshiba TCD1304AP)-Vis-NIR, USA) and a shortwave near-infrared spectrometer with a wavelength range of 1000–1700 nm (Ocean Optics, Flame-NIR+, USA) were used. Reflectance measurements were performed for both spectra. Fig. 2 shows the setup for combining SW-NIR and Vis-NIR spectroscopy acquisitions.

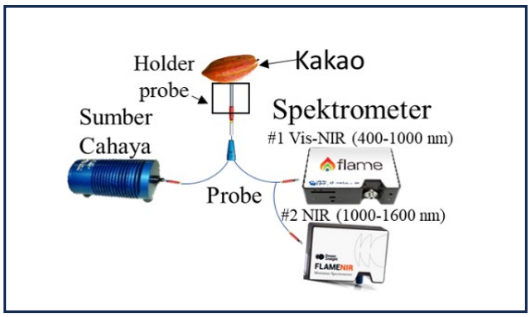


Figure 2. Spectra Measurement Schematic

A modular spectrometer system, reflection probes, tungsten halogen light, and computers (PCs/Laptops) with Ocean View 2.0.12 software were used in the experiment. The NIR instrument uses a semiconductor detector as a sensor and a halogen tungsten lamp as the radiation source, generating high-intensity radiation in the visible and IR spectra. Each cacao sample was placed on the optical probe. Ocean View 2.0.12 software was used to capture the reflectance spectra for the Vis-NIR and SW-NIR wavelengths, with integration times of 470 ms and 800 ms, respectively, scan to averages of 20 and 12, and a boxcar width of 1. Before collecting sample spectra, a calibration procedure was carried out, using a white reference using a ceramic diffuse reflectance standard (WS-1, Ocean Optics, USA) and a dark reference by switching off the light source.A total of 400 spectra were collected from 20 cacao samples, yielding 20 spectra from each sample. The same reflection probe could be used by both the Vis-NIR and SW-NIR spectrometers to read simultaneously, and the data collected could be automatically saved in different files.

(a) (b)

Figure 3. (a) Flame-T-Vis-NIR Ocean Optics, (b) Flame-NIR+ Ocean Optics

2.3. Spectra Analysis

Ocean View 2.0.12, was used to obtain the Vis-NIR reflectance spectra (400-1000 nm) and SW-NIR reflectance spectra (1000-1600 nm), which were then separately compiled in Microsoft Excel®. Chemometric studies on 400 spectra were performed using the Unscrambler® X Camo software (version 10.4). To create and validate a multivariate model, the dataset was separated into calibration sets (75% samples) and prediction sets (25% samples). Six spectrum pre-processing techniques, including SNV, MSC, area normalization, Savitzky Smoothing, Savitsky-Golay 1st Derivative, and Savitsky-Golay 2nd Derivative, were utilized to reduce the sample-size effect. Pre-processing spectra are also employed to reduce problems that can impair the accuracy of chemical composition measurements, such as baseline variation, light scattering, and path length inconsistencies [18]. The pretreatment of spectra is an important stage in the interpretation of spectral data. When collecting spectral data, the data often noise, scattering spectra, and various other disturbances are also detected. Therefore, spectral pretreatment is needed to minimize the effect of disturbances, and the intention of interpreting spectral data can be enhanced [19].

Principal component analysis (PCA) is an unsupervised method used for dimensionality reduction by converting a number of correlated variables into fewer variables called principal components (PCs) [20]. In this study, PCA was used to cluster samples and detect the occurrence of outliers. A classification model to discriminate between cacao maturities was developed using principal component analysis linear discriminant analysis (PCA-LDA).

Principal component analysis (PCA), a supervised analysis method, and a method that combines PCA and linear discriminant analysis (PCA-LDA), both of which are chemometric tools, were used to evaluate the pre-processed spectra [21]. The PCA-LDA method is a qualitative analysis consisting of 2 steps [22]. First, the dimension of the spectral variable is reduced to become a new variable called a PC. Second, new variables are classified into two maturity levels: mature (M) and immature (T). PCA-LDA was performed using the spectrum as the independent variable (X) and maturity level as the dependent variable (Y).

The qualitative performance of the model was assessed using the resulting accuracy (Acc) value. Accuracy was calculated using Equation 1. The search equation for the model accuracy value was modified from [23], [24]. The accuracy value shows a comparison of the number of successful classifications of each class with the number of samples. A higher accuracy value (maximum 100%) indicated that the model had better predictions.

$$\text{Accuracy}=\frac{n^{P}T+n^{P}M}{nT+nM}x100\% \quad (1)$$

Descriptions:

$n^{P}T$ = number of correct predictions of the immature sample
$n^{P}M$ = number of correct predictions of the mature sample
nT = total sample of immature individuals
nM = total number of maturecells

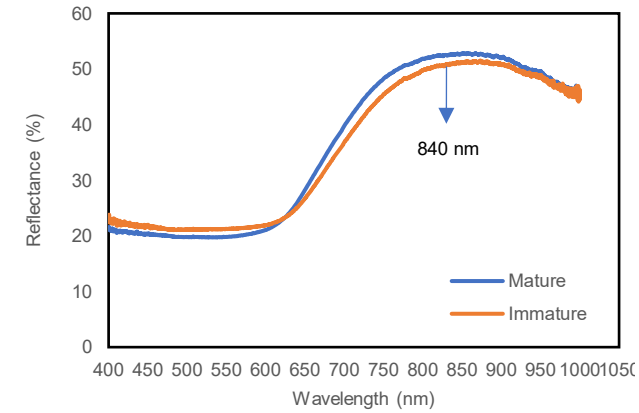
3 Results and discussion

3.1.Visible near Infrared Spectra (Vis-NIR)

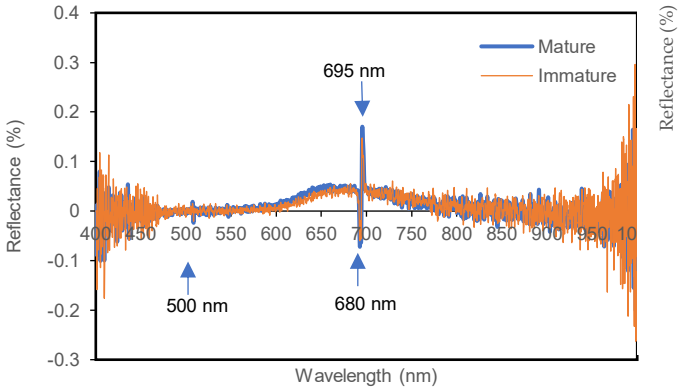
The Vis-NIR spectrometer recorded the spectral reflectance in the wavelength range of 345-1033 nm. However, the spectra used for the analysis were only in the range of 400–1000 nm because of the presence of severe noise at wavelengths below 400 nm and above 1000 nm. The number of variables in the wavelength range of 400-1000 nm was 3392 variables. This is due to interference at wavelengths below 400 nm and above 1000 nm measured from Vis-NIR Spectroscopy. This wavelength truncation avoids noise in the spectra, which could potentially interfere with the important information contained in the spectra. The reduction in resolution contributed minimally to the accuracy of the model. A reduction in the number of wavelength points can increase detection speed [25]. The truncation of the noise data at the beginning and end of the wavelength also aims to optimize the information. In addition to truncating the wavelength range, disturbances in the spectra due to light in addition to truncating the wavelength range, disturbances in the spectra due to light scattering can also be removed using preprocessing techniques that are applied before forming the calibration model so that it can be easier to analyzing the data [26].

The Vis-NIR spectral data obtained from the original spectra of the pure samples before spectral pre-processing are shown in Figure 4 (a). Spectral samples that react to light radiation determine these spectral characteristics. When light strikes cacao samples, it may be transmitted, absorbed, or reflected [27]. Using the original Vis–NIR spectra (Figure 4 (a)), the results showed that low reflectance or high absorbance at 450-500 nm wavelength is influenced by carotenoid pigments. According to [24] carotenoids are pigments that cause a bright yellow color in the fruit. Fruits with

a dominant yellow color produce spectra with low reflectance values at wavelengths of 450-500 nm. Low reflectance or high absorbance at a wavelength of 500-550 nm contains some information related to the pigment content that gives rise to the red color of the fruit. The peak at 840 nm is correlated with oxalic acid. The acid content of cocoa beans is strongly related to the formation of chocolate flavor [28].



(a)



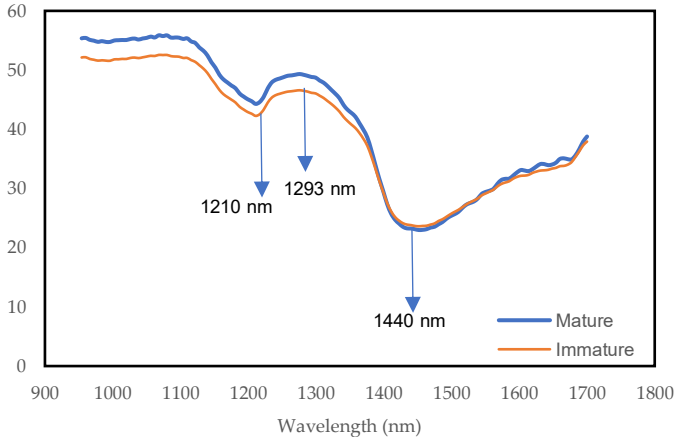
(b)

Fig. 4 (a) Original and **(b)** Savitzky–Golay 1st Derivative spectra mature and non mature of cacao observed with Vis-NIR Spectroscopy

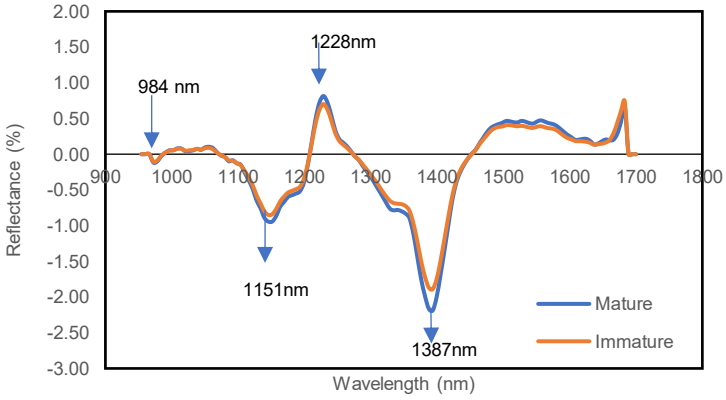
According to [29], the absorbance at approximately 550 nm is influenced by anthocyanins. The absorption peak at 620 nm corresponds to respiratory pigments, which may be correlated with anthocyanin or red pigments [30]. Wavelength of 950-1000 nm indicated information related to the sugar groups and water molecules. At these wavelengths, the absorption bands of functional groups containing hydrogen atoms, such as O-H, C-H, and water molecules. such as O-H, C-H, and N-H bonds [31]. Fig. 4b. shows that some peaks and valleys are more distinct following pre-processing using the Savitsky Golay 1st derivative. The peak at 680 nm correlated with chlorophylls. Chlorophyll pigments contain C-H-N-O bonds that can be detected at Vis-NIR wavelengths [32]. The spectrum observed at a wavelength of 695 nm showed a high absorbance, which can be attributed to the presence of polyphenols in the sample. An increase in absorbance indicated an increased percentage of the total antioxidant activity of the polyphenolic fraction. [33].

3.2.Shortwave near-infrared spectra (SW-NIR)

The wavelength used in data analysis was 1000-1600 nm, which can be performed by SW-NIR Spectroscopy at 954-1700 nm. This is due to interference at wavelengths below 1000 nm and above 1600 nm measured by SW-NIR Spectroscopy. The SW-NIR spectral data obtained from the original spectra of the pure samples before spectral pre-processing are shown in Figure 5a. The original absorbance SW-NIR spectra of the cacao samples at various maturity stages are shown in Figure 5a. between 1000 nm and 1600 nm, respectively. Numerous factors, including the principal chemical concentrations (water, carbohydrates, etc.) and particle size, affect the shape of the spectra.



(a)

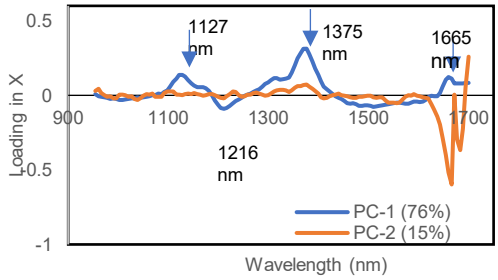


(b)

Fig. 5 (a) Original and **(b)** Savitzky–Golay 1st Derivative spectra mature and non mature of cacao observed with SW-NIR Spectroscopy

The lipid content and significant water band absorption at 1450 nm caused the spectra to peak at roughly 1200 nm. Peak absorption at wavelengths of 1210, 1161, 1188, 1212, and 1323 nm reveals the

presence of lipids [34]. The bands in the 1150–1200 nm range are attributed to the second overtone of the C–H stretching modes, those in the 1400-1450 nm region are attributed to the first overtone of the O–H stretching modes, and the combination of the C–H stretching and C–H deformation modes is responsible for the bands in the region between 1450 and 1550 nm [34]. Another absorbance peak, a broad reflectance valley at around 1180 nm, which extends up to 1200 nm, is affected by proteins as well as the second overtone of C-H stretching in lipids [34]. Fig. 5b shows that some peaks and valleys



are more distinct following pre-processing using the Savitsky Golay 1st derivative. The absorbance peaks at 1380 nm and SW-NIR band in the 1355-1400 nm region represent the second overtone of the CH (carbohydrate, lipid) combination [35]. According to a previous study, the region in the window between 1200 and 1400 nm is related to several important chemical compounds such as lipids and proteins [31]. The NIR absorption coefficient increases with an increase in glucose concentration at wavelengths of 1630–1700 nm [36].

3.3. PCA Modeling

The spectroscopic data of mature and immature cacao were analyzed using a non-supervised multivariate Principal Component Analysis (PCA) approach. The main goal of this investigation was to evaluate the ability of samples to discriminate based on their quality [5]. PCA makes it possible to identify the key properties of spectra by reducing the dimensionality of the data in an unsupervised manner. Principal component analysis (PCA), a multivariate analysis, is used to produce new variables called principal components (PCs) and to minimize the spectral dimension without sacrificing significant data [33,34]. A typical approach involves applying PCA to reduce the dimensionality of the dataset while retaining most of the variance. This step helps remove redundant or irrelevant features and capture the major sources of variation in the data. In this investigation, spectra from 3648 variables for the Vis-NIR were used to construct 14 PCs with the Mahalanobis method and 128 variables for the SW-NIR were used to construct nine PCs using the Mahalanobis method. Principal component analysis (PCA) is an unsupervised technique.

3.3.1. PCA Modeling using Vis-NIR spectra

Fig. 6a displays a score plot of PC-1 and PC-2 constructed to detect principal groupings among observations using Savitzky-Golay 1st Derivative pre-processing spectra. The PCA model contained 96% of data variance, which explained variance of 90% from PC-1 and 6% in PC-2. According to Fig. 6b, the first principal component (PC1) explained 90% of the total explained variance (X-variance on the y-axis) in the data. Fourteen PCs accounted for more than 90% of the total explained variance in the calibration and validation processes (89,58% and 89,51%, respectively). Consequently, we decided to use 14 PCs as input variables for the PCA-LDA model. The portions of the spectra where the changes among the cacao samples are clearly evident are shown in Fig. 6c, which presents the PC1 loading values of the spectral data. When interpreting the interactions between variables and clusters in the score plot, the loading plot (Fig. 6c) was used to show the significance of each variable in the model.

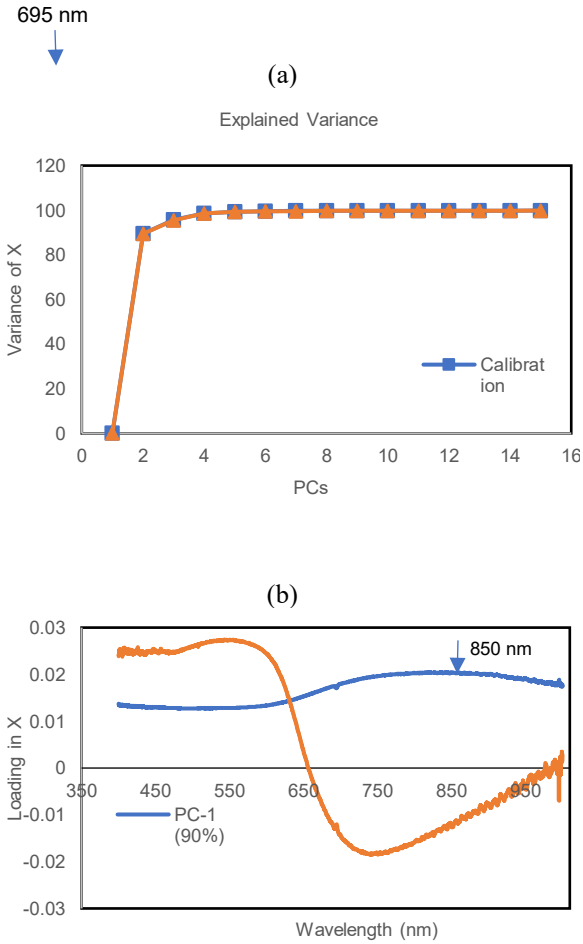
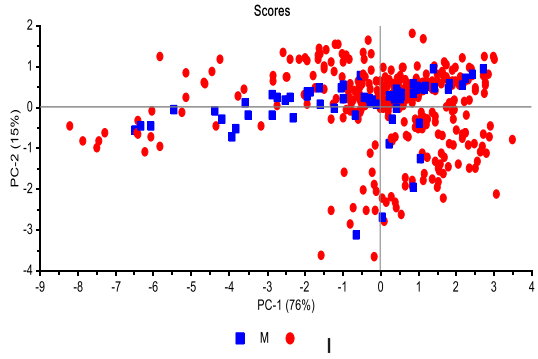


Fig. 6 Vis-NIR Analysis: **(a)** PC1/PC2 Score plot by PCA analysis on the calibration set : red (immature) ; blue (mature) ; **(b)** Plot of explained variance (%) of PCs used in the PCA model; **(c)** PC-1 and PC-2 loadings plot

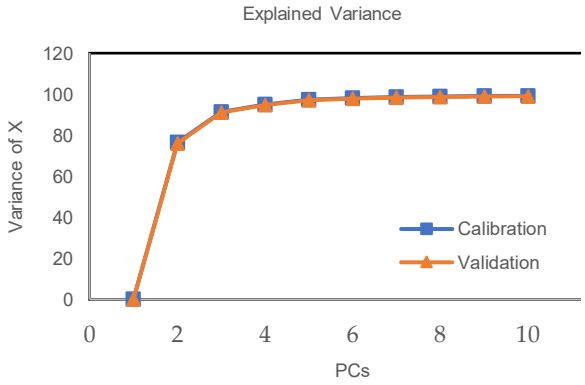
3.3.2. PCA Modeling using SW-NIR spectra

Fig. 7a displays a score plot of PC-1 and PC-2 constructed to detect principal groupings among

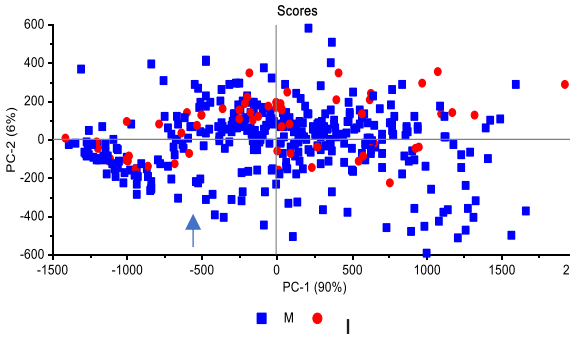
observations using Savitzky-Golay 1st Derivative pre-processing spectra. The PCA model contained 91% of data variance, which explained variance of 76% from PC-1 and 15% from PC-2. According to Fig. 7b, the first principal component (PC1) explained 76% of the total explained variance (X-variance on the y-axis) in the data. Nine PCs accounted for more than 76% of the total explained variance in the calibration and validation processes (76,40% and 76,06%, respectively). Consequently, we decided to use nine PCs as input variables for the PCA-LDA model. The portions of the spectra where the changes among the cacao samples are clearly evident are shown in Fig. 7c, which presents the PC1 loading values of the spectral data. When interpreting the interactions between variables and clusters seen in the score plot, a loading plot (Fig. 7c) was used to show the significance of each variable in the model.



(a)



(b)



(c)

Fig. 7SW-NIR Analysis: **(a)** PC1/PC2 Score plot by PCA analysis on the calibration set : red (immature) ; blue (mature) ; **(b)** Plot of explained variance (%) of PCs used in the PCA model; **(c)** PC-1 and PC-2 loadings plot

3.4. PCA-LDA Modeling

After PCA, LDA was applied to the transformed data. LDA operates by finding a linear projection that maximizes between-class scatter while minimizing within-class scatter. It aims to project the data into a lower-dimensional subspace, where the classes are well separated and the data points within each class are tightly clustered. A supervised technique called linear discriminant analysis (LDA) was utilized to categorize category variables, such as the cacao maturity level.

3.4.1.PCA-LDA using Vis-NIR Spectra

The selection of the ideal number of principal components (PCs) to be used in the model is a critical step in the creation of the PCA-LDA framework. The total number of PCs is the same as the total number of initial variables. For many datasets, the top few PCs typically accounted for the majority of the variance, making it possible to ignore the remaining PCs with little information loss. The PCA-LDA model was developed using preprocessed spectra. The results showed that the cacao maturity level was perfectly classified using raw spectra for the cacao maturity classification Vis-NIR spectrometer using PCA-LDA methods.

The accuracy of a model can be determined by its sensitivity, specificity, and misclassification rate. Sensitivity allows the assessment of how well the model can identify samples that belong to a particular class, and specificity measures the capacity of the model to reject non-belonging samples. The misclassification rate was the ratio of false positives to the total number of samples. In this study, the misclassification rate was considered to evaluate model performance. The classification of mature and immature cacao using the PCA-LDA model performance with Vis-NIR spectroscopy is shown in Table 1. After spectral preprocessing using Area Normalization (AN) resulted that the method most inaccurate with the highest misclassification was 86%. The best model for Vis-NIR Spectroscopy was obtained using raw spectra, with an accuracy of 90% for calibration and 86% for prediction.

Table 1. PCA-LDA performance of Vis-NIR spectroscopy.

Performance	Calibration						
	Raw	Norm	SNV	MSC	SG	SGD1	SGD2
Acc(%)	**90**	86	88	89	88	85	85
Misclasified rate	10	**86**	11	11	11	11	11
Performance	**Prediction**						
	Raw	Norm	SNV	MSC	SG	SGD1	SGD2
Acc(%)	**86**	86	85	87	86	85	85
Misclasified rate	13	13	15	12	13	15	15

Note: Acc = accuracy; Raw = raw data; Norm = normalized data; MSC = multiple scatter correction; SNV = standard normal variate; SG= Savitzky Golay–s; SGD1 =Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative.

The model's prediction performance also showed that it could accurately classify the cacao maturity level into each category (Table 1). Because the accuracy was greater than 90%, the model can be deemed to be a good one. In light of the overall amount of data employed, the accuracy demonstrates the model's capacity to properly categorize the cacao maturity level [37]. The outcomes of model generation using a set of samples for calibration (n=320). It seen that all samples were not correctly classified, resulting in an accuracy of 90 %. In Table 2, a confusion matrix is presented that exhibits the outcomes of the classification of prediction samples that utilized the Raw PCA-LDA model. Notably, the highest occurrence of false predictions (n=48) was observed for immature individuals, which were mistakenly classified as mature [34].

Table 2.The Result of a Classification of Cocoa Pods Samples using the PCA-LDA model from Vis NIR Spectra

True Class	**Predicted Class**	
	Calibration	Prediction
Non Mature	Non Mature (16)	Non Mature (1)
Non Mature	Mature (32)	Mature (11)
Mature	Non Mature (0)	Non Mature (0)
Mature	Mature (272)	Mature (68)
Accuracy (%)	90	86
Misclasification Rate (%)	10	14

3.4.2.PCA-LDA using SW-NIR spectra

The PCA-LDA model was developed using preprocessed spectra. The performance of the PCA-LDA model using the SW-NIR spectra is shown in Table 3. The results showed that the cacao maturity level was perfectly classified using SGD1 spectra for the cacao maturity classification SW-NIR spectrometer using PCA-LDA methods. The classification accuracies observed for differentiating the cacao samples through utilization the Vis-NIR original spectra were 91,56% and 86,25% during the calibration and prediction stages, respectively. After spectral preprocessing using Multiplicative Scatter Correction (MSC) resulted that the method most inaccurate with the highest misclassification was 9,69%. The best model for Vis-NIR Spectroscopy was obtained using Savitzky'-Golay 1st derivative pre-processing spectra, with an accuracy of 92,50% for calibration and 85% for prediction. The results show that PCA-LDA using SW-NIR spectra yielded a higher model than PCA-LDA using Vis-NIR spectra.

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

Performance	Calibration						
	Raw	Norm	SNV	MSC	SG	**SGD1**	SGD2
Acc(%)	91,56	91,56	90,63	90,31	92,05	**92,50**	90,83
Misclasified rate	8.44	8.44	9.38	**9.69**	7.95	7.50	9.17
Performance	**Prediction**						
	Raw	Norm	SNV	MSC	SG	**SGD1**	SGD2
Acc(%)	86,25	85,00	85,00	85,00	86,25	**85,00**	85,00
Misclasified rate	13.75	15.00	15.00	15.00	13.75	15.00	15.00

Note: Acc = accuracy; Raw = raw data; Norm = normalized data; MSC = multiple scatter correction; SNV = standard normal variate; SG= Savitzky Golay–s; SGD1 =Savitsky-Golay's 1st derivative, SGD2 = Savitsky-Golay's 2nd derivative.

The model's prediction performance also showed that it could accurately classify the cacao maturity level into each category (Table 4). The ability of the model to accurately categorize cocoa pod maturity was further demonstrated by its strong prediction performance. The model's prediction performance also showed that it could accurately classify the cacao maturity level into each category (Table 3). Because the accuracy is greater than 92,50%, the model can be deemed to be a good one. In light of the overall amount of data employed, the accuracy demonstrates the model's capacity to properly categorize the cacao maturity level [34]. The outcomes of model generation using a set of samples for calibration (n=320). It seen that all samples were not correctly classified, resulting in an accuracy of 92,50 %. In Table 4, a confusion matrix is presented that exhibits the outcomes of the classification of prediction samples that utilized the Raw PCA-LDA model. Notably, the highest occurrence of false predictions (n=51) was observed for immature individuals, which were mistakenly classified as mature [27].

Table 4.The Result of a Classification of Cocoa Pods Samples using the PCA-LDA model from Vis NIR Spectra

True Class	Predicted Class	
	Calibration	Prediction
Non Mature	Non Mature (27)	Non Mature (0)
Non Mature	Mature (24)	Mature (12)
Mature	Non Mature (0)	Non Mature (0)
Mature	Mature (269)	Mature (68)
Accuracy (%)	92,50	85
Misclasification Rate (%)	7,50	15

4. Conclusion

The results show that the availability of Vis-NIR and SW-NIR spectroscopy for cocoa pod maturity can be distinguished using spectroscopy combined with chemometrics analysis. The results showed that SW-NIR spectroscopy yielded a better performance than SW-NIR spectroscopy. The best model for classifying cocoa pod maturity obtained by SW-NIR spectroscopy had calibration performance of 92% accuracy for the PCA-LDA model. The best model for classifying cocoa pod maturity obtained by Vis-NIR spectroscopy had calibration performance of 90% accuracy for the PCA-LDA model. As a result, it can be said that Vis-NIR and SW-NIR spectroscopy are promising techniques for identifying cocoa pods maturity.

Acknowledgment. This research was supported by the Center for Higher Education Funding (BPPT) and the Indonesian Endowment Fund for Education (LPDP).

References

[1] Y. K. T. Dang and H. V. H. Nguyen, "Effects of Maturity at Harvest and Fermentation Conditions on Bioactive Compounds of Cocoa Beans," *Plant Foods Hum. Nutr.*, vol. 74, no. 1, pp. 54–60, 2019, doi: 10.1007/s11130-018-0700-3.

[2] N. Elkhoshkhany, "UV–Vis-NIR spectroscopy, structural and thermal properties of novel oxyhalide tellurite glasses with composition TeO_2 -B_2O_3-$SrCl_2$-LiF-Bi_2O_3 for optical application," *Results Phys.*, vol. 13, 2019, doi: 10.1016/j.rinp.2019.102222.

[3] D. A. S. Saputri, M. Fahri Reza Pahlawan, B. M. A. Murti, and R. E. Masithoh, "Vis/NIR spectroscopy for non-destructive method in detecting soybean seeds viability," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1038, no. 1, 2022, doi: 10.1088/1755-1315/1038/1/012043.

[4] K. B. Beć, J. Grabska, and C. W. Huck, "In silico NIR spectroscopy – A review. Molecular fingerprint, interpretation of calibration models, understanding of matrix effects and instrumental difference," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 279, no. March, 2022, doi: 10.1016/j.saa.2022.121438.

[5] W. Terouzi *et al.*, "Characterization and rapid detection of adulterations in sesame oil using FT-MIR and PCA-LDA," *Moroccan J. Chem.*, vol. 4, no. 4, pp. 4-4 (2016)1052-1060, 2016, [Online]. Available: https://revues.imist.ma/index.php/morjchem/article/view/5167

[6] J. Grabska and C. W. Huck, "Near-Infrared Spectroscopy in Bio-Applications," 2020.

[7] J. Van De Steene *et al.*, "Authenticity analysis of oregano: development, validation and fitness for use of several food fingerprinting techniques," *Food Res. Int.*, vol. 162, no. PA, p. 111962, 2022, doi: 10.1016/j.foodres.2022.111962.

[8] Y. B. Che Man, Z. A. Syahariza, M. E. S. Mirghani, S. Jinap, and J. Bakar, "Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy," *Food Chem.*, vol. 90, no. 4, pp. 815–819, 2005, doi: 10.1016/j.foodchem.2004.05.029.

[9] N. N. Batista, D. P. de Andrade, C. L. Ramos, D. R. Dias, and R. F. Schwan, "Antioxidant capacity of cocoa beans and chocolate assessed by FTIR," *Food Res. Int.*, vol. 90, pp. 313–319, 2016, doi: 10.1016/j.foodres.2016.10.028.

[10] Y. Hu *et al.*, "Determination of antioxidant capacity and phenolic content of chocolate by attenuated total reflectance-Fourier transformed-infrared spectroscopy," *Food Chem.*, vol. 202, pp. 254–261, 2016, doi: 10.1016/j.foodchem.2016.01.130.

[11] I. A. Santos, D. G. Conceição, M. B. Viana, G. de J. Silva, L. S. Santos, and S. P. B. Ferrão, "NIR and MIR spectroscopy for quick detection of the adulteration of cocoa content in chocolates," *Food Chem.*, vol. 349, no. June 2020, pp. 1–6, 2021, doi: 10.1016/j.foodchem.2021.129095.

[12] J. Moros, F. A. Iñón, S. Garrigues, and M. de la Guardia, "Near-infrared diffuse reflectance spectroscopy and neural networks for measuring nutritional parameters in chocolate samples," *Anal. Chim. Acta*, vol. 584, no. 1, pp. 215–222, 2007, doi: 10.1016/j.aca.2006.11.020.

[13] D. N. de Oliveira, A. C. B. Camargo, C. F. O. R. Melo, and R. R. Catharino, "A fast semi-quantitative screening for cocoa content in chocolates using MALDI-MSI," *Food Res. Int.*, vol. 103, no. October 2017, pp. 8–11, 2018, doi: 10.1016/j.foodres.2017.10.035.

[14] E. K. Anyidoho, "Differentiation of Organic Cocoa Beans and Conventional Ones by Using Handheld NIR Spectroscopy and Multivariate Classification Techniques," *Int. J. Food Sci.*, vol. 2021, 2021, doi: 10.1155/2021/1844675.

[15] M. C. García-Muñoz, M. P. Tarazona-Díaz, N. A. Meneses-Marentes, G. González-Sarmiento, A. S. Pineda-Guerrero, and G. E. Gómez-Uribe, "Development of color guides to

evaluate the maturity of cacao clones by digital image processing," *Pesqui. Agropecu. Trop.*, vol. 51, no. November, 2021, doi: 10.1590/1983-40632021v5169621.

[16] R. E. Masithoh, S. Lohumi, W. S. Yoon, H. Z. Amanah, and B. K. Cho, "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*, vol. 6, no. 10, p. e05099, 2020, doi: 10.1016/j.heliyon.2020.e05099.

[17] P. Khuwijitjaru, K. Boonyapisomparn, and C. W. Huck, "Near-infrared spectroscopy with linear discriminant analysis for green 'Robusta' coffee bean sorting," *Int. Food Res. J.*, vol. 27, no. 2, pp. 287–294, 2020.

[18] M. Ferreiro-González, E. Espada-Bellido, L. Guillén-Cueto, M. Palma, C. G. Barroso, and G. F. Barbero, "Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics," *Talanta*, vol. 188, pp. 288–292, 2018, doi: 10.1016/j.talanta.2018.05.095.

[19] M. Caredda *et al.*, "Building of prediction models by using Mid-Infrared spectroscopy and fatty acid profile to discriminate the geographical origin of sheep milk," *Lwt*, vol. 75, pp. 131–136, 2017, doi: 10.1016/j.lwt.2016.08.053.

[20] D. Suhandy and M. Yulia, "Classification of lampung robusta Specialty coffee according to differences in cherry processing methods using UV spectroscopy and chemometrics," *Agric.*, vol. 11, no. 2, pp. 1–11, 2021, doi: 10.3390/agriculture11020109.

[21] M. E. Raypah, L. J. Zhi, L. Z. Loon, and A. F. Omar, "Near-infrared spectroscopy with chemometrics for identification and quantification of adulteration in high-quality stingless bee honey," *Chemom. Intell. Lab. Syst.*, vol. 224, no. March, p. 104540, 2022, doi: 10.1016/j.chemolab.2022.104540.

[22] P. Mishra *et al.*, "MBA-GUI: A chemometric graphical user interface for multi-block data visualisation, regression, classification, variable selection and automated pre-processing," *Chemom. Intell. Lab. Syst.*, vol. 205, no. August, p. 104139, 2020, doi: 10.1016/j.chemolab.2020.104139.

[23] E. Trullols, I. Ruisánchez, and F. X. Rius, "Validation of qualitative analytical methods," *TrAC - Trends Anal. Chem.*, vol. 23, no. 2, pp. 137–145, 2004, doi: 10.1016/S0165-9936(04)00201-8.

[24] L. S. Vieira, C. Assis, M. E. L. R. de Queiroz, A. A. Neves, and A. F. de Oliveira, "Building robust models for identification of adulteration in olive oil using FT-NIR, PLS-DA and variable selection," *Food Chem.*, vol. 345, no. August 2020, 2021, doi: 10.1016/j.foodchem.2020.128866.

[25] K. Ncama, "Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' orange (Citrus sinensis) and 'Star Ruby' grapefruit (Citrus x paradisi Macfad)," *J. Food Eng.*, vol. 193, pp. 86–94, 2017, doi: 10.1016/j.jfoodeng.2016.08.015.

[26] B. N. Peshlov, F. E. Dowelt, F. A. Drummond, and D. W. Donahue, "Comparison of three near infrared spectrophotometers for infestation detection in wild blueberries using multivariate calibration models," *J. Near Infrared Spectrosc.*, vol. 17, no. 4, pp. 203–212, 2009, doi: 10.1255/jnirs.842.

[27] X. Lin and D. W. Sun, "Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses," *Trends Food Sci. Technol.*, vol. 104, no. June, pp. 163–176, 2020, doi: 10.1016/j.tifs.2020.06.009.

[28] A. F. Omar, H. Atan, and M. Z. MatJafri, "NIR spectroscopic properties of aqueous acids solutions," *Molecules*, vol. 17, no. 6, pp. 7440–7450, 2012, doi: 10.3390/molecules17067440.

[29] V. Cortés, C. Ortiz, N. Aleixos, J. Blasco, S. Cubero, and P. Talens, "A new internal quality index for mango and its prediction by external visible and near-infrared reflection spectroscopy," *Postharvest Biol. Technol.*, vol. 118, pp. 148–158, 2016, doi: 10.1016/j.postharvbio.2016.04.011.

[30] M. N. Merzlyak, A. E. Solovchenko, and A. A. Gitelson, "Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit," *Postharvest Biol. Technol.*, vol. 27, no. 2, pp. 197–211, 2003, doi: 10.1016/S0925-5214(02)00066-2.

[31] J. U. Porep, D. R. Kammerer, and R. Carle, "On-line application of near infrared (NIR) spectroscopy in food production," *Trends Food Sci. Technol.*, vol. 46, no. 2, pp. 211–230, 2015, doi: 10.1016/j.tifs.2015.10.002.

[32] J. H. Choi, "Portable, non-destructive tester integrating VIS/NIR reflectance spectroscopy for the detection of sugar content in Asian pears," *Sci. Hortic. (Amsterdam).*, vol. 220, pp. 147–153, 2017, doi: 10.1016/j.scienta.2017.03.050.

[33] R. Sahukari, S. Bhasha, H. Singamala, and V. Subbaiah, "Assessment of Potential Antioxidant Activity of Polyphenolic Fraction Separated from Acalypha Indica Leaves: An In vitro Approach Ethnopharmacology and Molecular Biology View project Transcriptome analysis View project," no. October 2017, 2015, [Online]. Available: https://www.researchgate.net/publication/320258303

[34] I. Novianty, K. B. Seminar, Irzaman, and I. W. Budiastra, "Improving the accuracy of near-infrared (NIR) spectroscopy method to predict the oil content of oil palm fresh fruits," *IOP*

Conf. Ser. Earth Environ. Sci., vol. 460, no. 1, 2020, doi: 10.1088/1755-1315/460/1/012004.

[35] T. J. Gutiérrez, "State-of-the-Art Chocolate Manufacture: A Review," *Compr. Rev. Food Sci. Food Saf.*, vol. 16, no. 6, pp. 1313–1344, 2017, doi: 10.1111/1541-4337.12301.

[36] J. Y. Tang, N. Y. Chen, M. K. Chen, M. H. Wang, and L. S. Jang, "Dual-wavelength optical fluidic glucose sensor using time series analysis of D(+)-glucose measurement," *Jpn. J. Appl. Phys.*, vol. 55, no. 10, 2016, doi: 10.7567/JJAP.55.106601.

[37] M. Rocha Baqueta, A. Coqueiro, P. Henrique Março, M. Mandrone, F. Poli, and P. Valderrama, "Integrated 1H NMR fingerprint with NIR spectroscopy, sensory properties, and quality parameters in a multi-block data analysis using ComDim to evaluate coffee blends," *Food Chem.*, vol. 355, no. March, p. 129618, 2021, doi: 10.1016/j.foodchem.2021.129618.