Advances in Biological Sciences Research, volume 19
Proceedings of the 2nd International Conference on Smart and Innovative Agriculture (ICoSIA 2021)

Classification of Cocoa Beans Based on Fermentation Level Using PLS-DA Combined with Visible Near-Infrared (VIS-NIR) Spectroscopy

Deny Saputro[1], Dede Priambodo[1], Muhammad Pahlawan[1], Rudiati Masithoh[1*]

[1] *Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[*] *Corresponding author. Email: evi@ugm.ac.id*

ABSTRACT
Fermentation is an important process in determining the quality of cocoa beans. Therefore, studies to determine the rate of fermentation of cocoa beans non-destructively and rapidly are needed. This study aimed to develop a robust model using Vis-NIR and PLS-DA to differentiate cocoa bean samples with different fermentation levels. The PLS-DA calibration and prediction classification model showed reliability (Rel) values above 75% and 73%. Accuracy (Acc) of calibration and prediction were higher than 90% and 87%, respectively. A reliable and robust model has been created, which allows determining the rate of fermentation of cocoa beans non-destructively and rapidly.

Keywords: *cocoa bean, fermentation, PLS-DA, Vis/NIR.*

1. INTRODUCTION

Chocolate is one of the most popular food in the world. Chocolate is a processed product with cocoa beans as the main ingredient and play an important role in creating flavor of the chocolate products. The variabilities of cocoa beans are influenced by growth conditions, genetics, and drying processes. One of the process that greatly affects cocoa flavor is fermentation [1]. Fermentation promotes biochemical changes in flavor precursors in cocoa beans to alter an astringent and unpleasant taste of raw cocoa into a characteristic cocoa taste and flavor [2].

Fermentation involve yeasts, acetic acid bacteria, and lactic acid bacteria which stimulate enzymatic activity to hydrolyze sugars, organic acids, proteins, and polyphenols in cocoa beans [3]. Those microbial activity causes physical and chemical changes in cocoa beans. Fermentation of cocoa beans generally lasts for 5-7 days. During the fermentation process, cocoa beans experience a decrease in pH and change in seed color from purple to brown [4]. In addition to discoloration and taste, fermentation also create cocoa beans aroma and protein degradation [5].

Fermentation of cocoa are assessed based on mass temperature, total soluble solids (TSS), pH, and acidity [3] or fermentation index and cut test [6]. The quality of cocoa bean fermentation rate of cocoa beans can also be known through chemical parameters such as pH, titration acid, amount of fat, phenol components, antioxidants, and nitrogen ammonia (NH3) [7]. Those methods involve sample destruction for preparation or analysis as well as require long-time analysis. Therefore, a rapid, non-destructive, and accurate evaluation is needed for routine and large sample analysis.

Near Infrared (NIR) spectroscopy is one of the non-destructive, inexpensive, and accurate method which have been used for detecting solid [8], liquid [9], or powder form [10] of food. NIR spectroscopy has been reported to be effective for cocoa bean quality control [11]. NIR spectroscopy was reported successfully in determining fermentation levels in cocoa beans using FT-NIR spectroscopy [12].

NIR spectroscopy measures the response of chemical components composed of C-H-N-O bonds when exposed to light at NIR wavelength ranges. Unlike NIR spectroscopy, visible and NIR (Vis/NIR) spectroscopy at 400-1000 nm has a weak absorbance to C-H-N-O bonds.

A weak absorption of Vis/NIR spectroscopy to moisture content (O-H bonds) in food provide a proper tool for detection of low chemical concentration in food. In Vis/NIR region, color of food due to various pigments exhibit different and distinct absorption values [13]. Therefore, Vis/NIR spectroscopy can be used to measure food quality parameters that correlate with color, such as cocoa bean fermentation. Due to limited studies which reported the use of Vis/NIR spectroscopy in cocoa, this research aimed to classify cocoa beans based on fermentation level by using partial least square discriminant analysis (PLS-DA).

2. MATERIALS AND METHODS

2.1. Samples

The samples used in the study were cocoa beans as many as 315 samples with the origin of Kulon Progo, Makasar, and Lampung (105 cocoa beans for each region). Cocoa beans from each region have three levels of fermentation: non-fermented, half fermented, and full fermented (35 cocoa beans for each level of fermentation). During sampling, samples were stored in airtight room conditions at room temperatures of 25-28°C.

2.2. Spectra measurements

Spectra reflectance with wavelengths of 350-1000 nm with a resolution of 0.22 nm was taken with the Vis-NIR miniature spectrometer (Flame-T-VIS-NIR Ocean Optics, 350-1000 nm) with tungsten halogen lamps (360-2400 nm, HL-2000-HP-FHSA Ocean Optics) and reflectance probes (QR400-7 VIS-NIR Ocean Optics). Spectra retrieval was done inside a black box to minimize outside light. For each sample was done 3 (three) scans on different sides. A total of 915 spectra were used for multivariate analysis without being averaged.

2.3. Physicochemical Analysis

L*, a*, and b* color parameters were measured using the Chromameter (CR-400, Konica Minolta). Each sample was measured 3 times on the different sides. Cocoa beans were then split into 2, half parts used for moisture content measurement using the thermogravimetric method. The other half of cocoa was crushed and then weighed for acidity (pH) measurement. pH measurements were done using 3-4 samples of cocoa beans with the same origin and fermentation level with a weight of around 4 grams. The samples were mixed with distilled water and the pH of the resulting aqueous solutions were measured using pH meters (Mettler Toledo, USA). Color, pH, and moisture content were then analyzed using One Way Analysis of Variance (ANOVA) (Significance level 5%) with variations in fermentation level using SPSS.

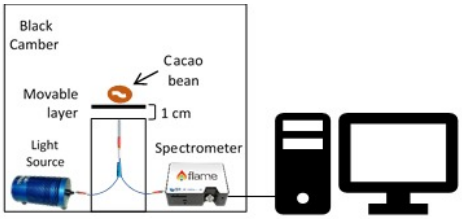

Figure 1 Spectra capture scheme

2.4. Chemometrics analysis

All spectra collected were compiled using Microsoft-Excel software and imported into The Unscrambler® X version (10.5.1, CAMO, OLSO, Norway) software for multivariate analysis. Before the multivariate analysis, all data were divided into calibration sets consisting of 2/3 of the total sample samples and a prediction set consisting of 1/3 of the total sample.

At each fermentation level, a model was built to classified samples that have the same fermentation level and those that were not. PLS models were built using spectra with wavelengths of 500-950 nm as predictors (X) and dummy variables as predictant (Y) which were a binary label on each sample. Sample with a label of 1 indicated that sample belonged to the group, while 0 indicated that the sample did not belong to the group. For example, in the making of a full-fermented model, full fermented samples will be labeled 1, while the half fermented, and non-fermented samples labeled 0. Sample labels (Non, half, full) on the model of half fermented and full fermented respectively were (0,1,0) and (0,0,1).

Calibration and prediction models with PLSR produced Y Predicted. Y Predicted results from calibration and prediction models with a value of more than 0.5 were classified in class 1, while prediction values less than 0.5 were included in group 0. An example of class classification based on PLS-DA results can be seen in Table 1.

Based on [14] and [15] some parameters that can be used to assess the qualitative performance of the model are as follows:

2.4.1. False negative and positive rate

False negative rate (FNR) is a comparison of the number of samples that truly 1 but were incorrectly classified as 0 (FN) with the total of samples that truly 1.

$$FNR = \frac{FN}{TP+FN} \cdot 100\% \quad (1)$$

False positive rate (FPR) is a comparison of the number of samples that truly 0 but were incorrectly classified as 1 (FP) with the total of samples that truly 0.

$$FPR = \frac{FP}{TN+FP} \cdot 100\% \quad (2)$$

A good model is expected to have FNR and FPR values that are closer to 0.

Table 1. Classification based on PLS-DA result

	Y Reference	Y Predicted	Class	Parameter
A	1	0.55	1	True Positive (TP)
B	1	0.49	0	False Negative (FN)
C	0	0.37	0	True Negative (TN)
D	0	0.67	1	False Positive (FP)

2.4.3. Accuracy and Reliability

Accuracy (Acc) is a comparison of the number of samples that were correctly classified with the total number of overall samples. Accuracy assesses the successful classification by considering the number of existing samples.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100\% \quad (5)$$

Reliability (Rel) is a comparison of the number of samples that were correctly classified with the total number of samples in each class. Reliability demonstrates the performance of the model in the sample classification, taking into account the correct predictions of each class independently.

$$Rel = Sen + Spe - 100 \quad (6)$$

To obtain the perfect model Acc and Rel must have the highest possible value (max 100%).

3. RESULTS

3.1. Data exploration

Table 2. shows the average values and standard deviations of the cocoa bean parameters used. Color is an important parameter that can indicate the quality of cocoa. Based on Table 2. it is known that L* (Lightness) and a* (redness) have inverse patterns. The average value of lightness of non-fermented beans is lower than half fermented beans, but full fermented cocoa beans have the lowest average compared to others. Analysis of Variance (ANOVA) shows that lightness values differ significantly. The average redness of non-fermented cocoa beans is higher than half fermented, but full fermented cocoa beans have the lowest average redness compared to others. However, redness does not differ significantly with a confidence level of 95%.

The b* (yellowness) value obtained was significantly different with a decreased pattern as the fermentation process takes place. Non-fermented cocoa beans have the highest value of yellowness which then decreases until the cocoa beans were fully fermented. A similar phenomenon was also found in [16]. During the fermentation process, the yellowish value will decrease due to the darkening of the cocoa beans.

The moisture content obtained decreased from non-fermented to half fermented beans but then increased in full fermented beans. This phenomenon is different from previous studies where full fermentation has the lowest water content [17]. This is likely because the drying technique of cocoa beans from each fermentation level is done using traditional drying methods that use sunlight as the main heat source. Traditional drying methods have a non-uniform temperature so the drying results will vary [18] .

Table 2. Average values and standard derivations of cocoa beans parameters used

Parameter	Non	Half	Full
L	42.66 ± 3.37ab	43.99 ± 7.23b	41.50± 3.14a
a	17.57 ± 2.85a	17.28 ± 3.66a	17.98 ± 0.96a
b	26.72 ± 5.55b	24.30 ± 7.64a	23.68 ± 2.87a
KA (%)	9.44 ± 2,22b	7.79 ± 0.70a	9.12 ± 3.98b
pH	6.00 ± 0.23c	5.67 ± 0.63b	5.28 ± 0.24a

Letters behind numbers across the row indicate significant differences between classes ($p < 0.05$).

The acidity level (pH) of cocoa beans decreases during the fermentation process. Based on a one-way ANOVA test, it is known that the difference in the pH value of cocoa beans differs significantly. During fermentation, there will be degradation of sugar by the microbes in the pulp, resulting in alcohol and organic acids which diffuse into the beans[19]. Figure 2. shows the reflectance spectra of cocoa beans from all regions and three levels of fermentation. The spectra obtained were similar to those obtained by [20]. Cocoa beans spectra from all regions and all maturity levels have a similar pattern. Non-fermented cocoa beans tend to have the lowest reflectance and full fermented cocoa beans have the highest reflectance. Spectra at wavelengths of

500-600 nm have low reflectance values that continue to increase until the end of the spectra at 950 nm. At wavelengths of 500- 600 nm, there is an absorbance peak characterized by a low reflectance value that is likely to be affected by anthocyanins [21]. Spectra at around 670 nm were correlated with chlorophyll [8], while wavelengths of 700-950 nm were mostly affected by water [22].

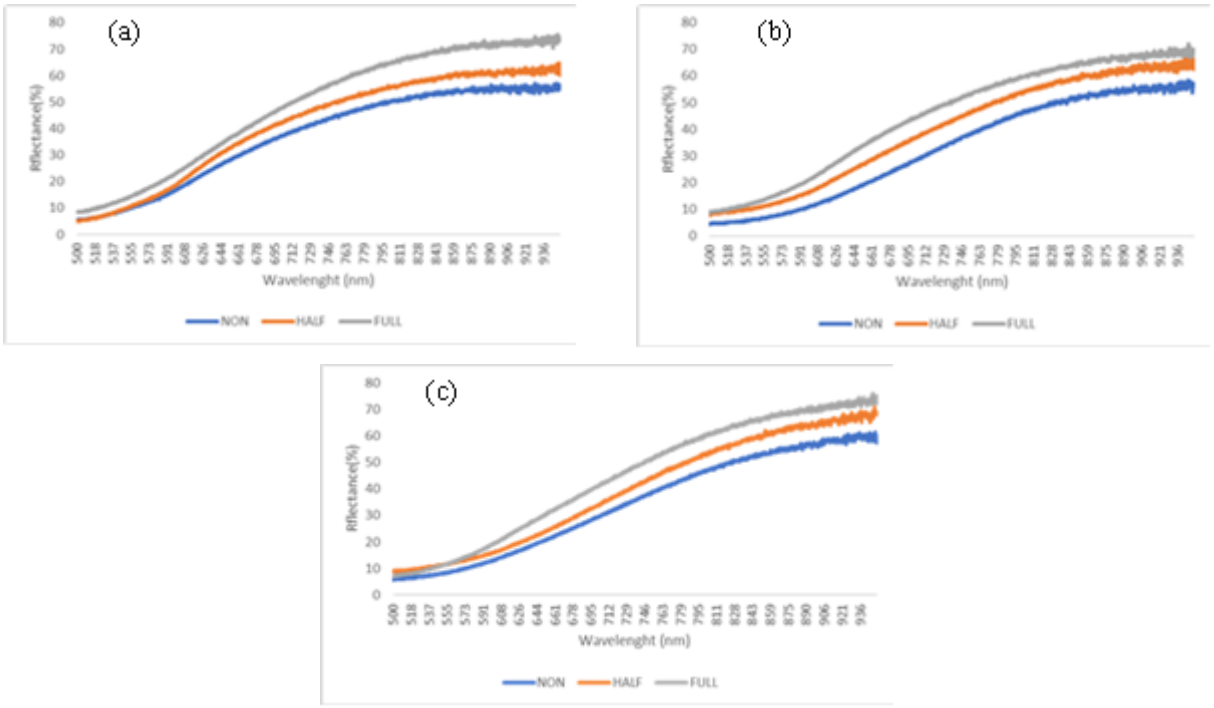

Figure 2. Cacao beans reflectance (a) Kulon Progo (b) Lampung (c) Makasar

Table 3. PLS-DA performance parameters

	Data Set	Fermentation Level		
		Non	Half	Full
Sen	C	79%	88%	81%
	P	85%	81%	84%
FNR	C	21%	12%	19%
	P	15%	19%	16%
Spe	C	97%	95%	94%
	P	88%	94%	89%
FPR	C	3%	5%	6%
	P	12%	6%	11%
Acc	C	91%	93%	90%
	P	87%	90%	87%
Rel	C	76%	84%	75%
	P	73%	76%	73%

3.2. PLS-DA model performance

Three PLS-DA models were built at each maturity level to distinguish samples that belonged to the group and those that were not. Table 3. shows the The Sen value of each model indicates the success rate of the model in distinguishing the samples that belong to the group. The highest calibration Sen value was obtained from the Half-Fermented model with a value of 88%, while the lowest was non-Fermented with a value of 79%.

The PLS-DA model produces a fairly good model with an Acc value above 90% for each model. However, the rel value for calibration and prediction of each model is still below 85%. Acc value is a percentage of overall classification success while Rel is the success of classification in each class (0 and 1). The low Rel values were likely due to a fairly low Sen value compared to Spe value.

The Vis-NIR model combined with PLS-DA can be used to classify cocoa beans from each fermentation

level. The best model obtained can distinguish Half fermented cocoa beans from Non and fully fermented cocoa beans with Sen, Spe, Acc above 81% for calibration and prediction.

4. CONCLUSIONS

Vis-NIR spectroscopy combined with PLS-DA analysis can be used to classify cocoa beans with different levels of fermentation. This method is proven to be fast, non-destructive, inexpensive, and a little preparation. Based on sensitivity, specificity, accuracy, reliability, and false-negative and positive it is known that the half fermented model is considered accurate and reliable.

AUTHORS' CONTRIBUTIONS

Deny Saputro and Dede Priambodo conceived of the presented idea and developed the theory and performed the computation. Muhammad Pahlawan verified the analytical methods. Rudiati Masithoh helped supervise the project.

ACKNOWLEDGMENTS

This work would not have been possible without the financial support of the Departement Agricultural and Biosystem Engieneering Faculty of Agricultural Technology Gadjah Mada University. I would especially like to thank Dr. Rudiati Evi Masithoh, the correspoding author of this paper.

REFERENCES

[1] Camu N, de Winter T, Addo S K, Takrama J S, Bernaert H and de Vuyst L 2008 Fermentation of cocoa beans: Influence of microbial activities and polyphenol concentrations on the flavour of chocolate *Journal of the Science of Food and Agriculture* **88** 2288–97

[2] Kongor J E, Hinneh M, de Walle D Van, Afoakwa E O, Boeckx P and Dewettinck K 2016 Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile - A review *Food Research International* **82** 44–52

[3] Hernández-Hernández C, López-Andrade P A, Ramírez-Guillermo M A, Guerra Ramírez D and Caballero Pérez J F 2016 Evaluation of different fermentation processes for use by small cocoa growers in mexico *Food Science and Nutrition* **4** 690–5

[4] Aculey P C, Snitkjaer P, Owusu M, Bassompiere M, Takrama J, Nørgaard L, Petersen M A and Nielsen D S 2010 Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics *Journal of Food Science* **75** 300–7

[5] Hue C, Gunata Z, Bergounhou A, Assemat S, Boulanger R, Sauvage F X and Davrieux F 2014 Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification *Food Chemistry* **148** 240–5

[6] Kongor J E, Felix Takrama J, Simpson Budu A, Mensah-Brown H and Afoakwa E O 2013 Effects of Fermentation and Drying on the Fermentation Index and Cut Test of Pulp Pre-conditioned Ghanaian Cocoa (Theobroma cacao) Beans *Journal of Food Science and Engineering* **3** 625–34

[7] Teye E, Anyidoho E, Agbemafle R, Sam-Amoah L K and Elliott C 2020 Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review *Infrared Physics and Technology* **104** 103127

[8] Masithoh R E, Pahlawan M F R and 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) *IOP Conference Series: Earth and Environmental Science* (IOP Publishing Ltd.)

[9] Masithoh R E, Haff R and Kawano S 2016 Determination of soluble solids content and titratable acidity of intact fruit and juice of satsuma Mandarin using a hand-held near infrared instrument in

transmittance mode *Journal of Near Infrared Spectroscopy* **24** 83–8

[10] Masithoh R E, Amanah H Z and Cho B K 2020 Application of fourier transform near-infrared (FT-NIR) and fourier transform infrared (FT-IR) spectroscopy coupled with wavelength selection for fast discrimination of similar color of tuber flours *Indonesian Journal of Chemistry* **20** 680–7

[11] Hashimoto J C, Lima J C, Celeghini R M S, Nogueira A B, Efraim P, Poppi R J and Pallone J A L 2018 Quality Control of Commercial Cocoa Beans (Theobroma cacao L.) by Near-infrared Spectroscopy *Food Analytical Methods* **11** 1510–7

[12] Shajahan S, Igathinathane C and Rangaraju V 2016 Quantification of fermentation levels in cocoa beans using FT-NIR spectroscopy *ASABE Annual International Meeting* (Orlando, Florida)

[13] Walsh K B, Blasco J, Zude-Sasse M and Sun X 2020 Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use *Postharvest Biology and Technology* **168** 111246

[14] Trullols E, Ruisánchez I and Rius F X 2004 Validation of qualitative analytical methods *TrAC - Trends in Analytical Chemistry* **23** 137–45

[15] Vieira L S, Assis C, de Queiroz M E L R, Neves A A and 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**

[16] Ndukwu M C and Udofia M 2016 Kinetics of change in colour and some bio-chemical composition during fermentation of cocoa bean *Cogent Food & Agriculture* **2**

[17] García-Alamilla P, Salgado-Cervantes M A, Barel M, Berthomieu G, Rodríguez-Jimenes G C and García-Alvarado M A 2007 Moisture, acidity and temperature evolution during cacao drying *Journal of Food Engineering* **79** 1159–65

[18] Deus V L, de Cerqueira E Silva M B, Maciel L F, Miranda L C R, Hirooka E Y, Soares S E, de Souza Ferreira E and da Silva Bispo E 2018 Influence of drying methods on cocoa (Theobroma cacao L.): Antioxidant activity and presence of ochratoxin A *Food Science and Technology* **38** 278–85

[19] Apriyanto M, Sutardi S, Supriyanto S and Harmayani E 2018 Fermentasi Biji Kakao Kering Menggunakan Saccharomyces cerevisiae, Lactobacillus lactis, dan Acetobacter aceti *Agritech* **37** 302

[20] Meulemans C C E, Surapati U and Tjatjo A 2013 Colorimetric measurements of cocoa beans (Theobroma cacao) *Indonesian Journal of Agricultural Science* **3** 52

[21] Pahlawan M F R, Wati R K and 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* vol 644 (IOP Publishing Ltd.)

[22] Wati R K, Pahlawan M F R and Masithoh R E 2021 Development of calibration model for pH content of intact tomatoes using a low-cost Vis/NIR spectroscopy *IOP Conference Series: Earth and Environmental Science* **686**

[9] A. Pnueli, In transition from global to modular temporal reasoning about programs, in: K.R. Apt (Ed.), Logics and Models of Concurrent Systems, Springer, Berlin, Heidelberg, 1984, pp. 123–144. DOI: https://doi.org/10.1007/978-3-642-82453-1_5

[10] B. Meyer, Applying "Design by Contract", Computer 25(10) (1992) 40–51. DOI: https://doi.org/10.1109/2.161279

[11] S. Bensalem, M. Bogza, A. Legay, T.H. Nguyen, J. Sifakis, R. Yan, Incremental component-based construction and verification using invariants, in: Proceedings of the Conference on Formal Methods in Computer Aided Design (FMCAD), IEEE Press, Piscataway, NJ, 2010, pp. 257–256.

[12] H. Barringer, C.S. Pasareanu, D. Giannakopolou, Proof rules for automated compositional verification through learning, in Proc. of the 2nd International Workshop on Specification and Verification of Component Based Systems, 2003.

[13] M.G. Bobaru, C.S. Pasareanu, D. Giannakopoulou, Automated assume-guarantee reasoning by abstraction refinement, in: A. Gupta, S. Malik (Eds.), Proceedings of the Computer Aided Verification, Springer, Berlin, Heidelberg, 2008, pp. 135–148. DOI: https://doi.org/10.1007/978-3-540-70545-1_14