Nondestructive Authentication of Cocoa Bean Cultivars by FT-NIR Spectroscopy and Multivariate Techniques

Article *in* Focus on Sciences · August 2016

DOI: 10.21859/focsci-020347

Focus on Sciences **Research Article**
Aug 2016, Volume 2, Issue 3

Nondestructive Authentication of Cocoa Bean Cultivars by FT-NIR Spectroscopy and Multivariate Techniques

Ernest Teye [1,*], James Uhomoibhi [2], Hui Wang [3]

[1] *Department of Agricultural Engineering, School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, Ghana*
[2] *School of Engineering, Ulster University, Jordanstown Campus, Newtownabbey, Northern Ireland, UK*
[3] *School of Computing and Mathematics, Ulster University, Jordanstown Campus, Newtownabbey, Northern Ireland, UK*

** Corresponding author: Ernest Teye, Department of Agricultural Engineering, School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, E-mail: ernest.teye@ucc.edu.gh*

DOI: 10.21859/focsci-020247

Submitted: 08.05.2016
Accepted: 09.16.2016

Keywords:
Cocoa Bean
Spectroscopy, Near-Infrared
Support Vector Machine

Abstract

Introduction: Rapid identification of cocoa bean varieties is vital for the authentication in cocoa trade. This paper examines the use of Near Infrared (NIR) Spectroscopy for nondestructive identification of cocoa bean cultivars.
Methods: In this study, five cocoa bean cultivars (IMC85 x IMC47, PA7 x PA150, PA150 x Pound7, Pd10 x Pd15 and T63/967 x T65/238) were scanned in the NIR range of 10000-4000 cm^{-1}. Linear discriminant analysis (LDA) and Support vector machine (SVM) algorithms were performed comparatively to build discrimination models based on principal component analysis (PCA). The models were optimized by cross validation to ensure their stability.
Results: The performance of SVM model was superior to LDA model. The optimal SVM model was achieved with five principal components (PCs) and an identification rate of 100% in both training set and prediction set.
Conclusions: The results proved that NIR spectroscopy technology with SVM algorithm can provide quick and reliable nondestructive analytical tool for the identification of cocoa bean cultivars and this would aid in quality control of cocoa bean.

INTRODUCTION

Cocoa (*Theobroma cacao L.*) bean is among the major food commodities worldwide and its consumption has increasingly become a daily activity due to its beneficial medicinal properties. These beneficial properties are as a result of the presence of polyphenols, procyanidin and a concentrated source of antioxidants [1]. Quality of cocoa beans depends on many factors such as genotype, agronomic management, soil factor, climatic condition and most importantly the post-harvest technology employed [2]. However, the single most important factor for the determination of cocoa bean quality is its cultivar because it is known that the phenolic content and concentration in cocoa beans depends on bean variety [3, 4]. Financial motivation continues to propel prot ducers and retailers to mislabel commodities. Furthermore, cocoa producing countries through their research stations are working judiciously through several years of breeding research to come out with varieties that are superior to the existing cultivars.

Therefore, cocoa bean identification has attained critical attention as a means to control adulteration and mislabeling, and will also facilitate breeding exercise. Currently, the identification of varietal differences is done by liquid chromatography, gas chromatography, capillary electrophoresis, sensory evaluation or plasma atomic emission etc. However, in spite of the relative success, they are very expensive, time consuming & tedious, destructive and involves chemical use.

For these reasons, breeders, food industry, regulatory authorities and consumer groups are searching for rapid, nondestructive and environmentally friendly technique for the authentication of cocoa beans. Near Infrared (NIR) Spectroscopy technology has emerged as a promising tool. This technique in the last years has proven to be a powerful analytical tool for estimating wide kinds of samples, it has been used in areas such as agriculture, nutrition, petrochemical, textile and pharmaceutical industries [5]. NIR spectroscopy offers several useful advantages over the traditional chemical methods; it is nondestructive, physical, rapid, simple operation, require small samples and minimal sample preparation [6].

Application of NIR spectroscopy on the analysis of cocoa bean product have been reported by a few researchers; [6] used NIR for the determination of fat, nitrogen and moisture, characterization of chocolate and cocoa powder. [7] for the determination of sucrose, lactose, fat and moisture, [8] for prediction of procyanidins in cocoa beans. Also, Teye and co-workers used it for the prediction of fat content, quantification of pH and fermentation index [9, 10]. However, upon literature search nondestructive identification of cocoa beans varieties by NIR spectroscopy has received no attention. The closest was the integration of NIR spectroscopy and Electronic tongue for cocoa bean varietal classification [11]. However, all other researchers, including my previous work performed the studies on powdered or grinded samples. On the other hand, only recently, Sunoj and other co-workers used whole bean and FT-NIR to determined cocoa bean quality [12]. However, they did not consider varietal classie fication. This research therefore aims at using FT-NIR spectroscopy and chemometrics techniques to nondestructively identify five cocoa cultivars.

METHODS

Sample Preparation

Five cultivars of fermented and dried cocoa beans samples (shown in Table 1) were obtained from Cocoa Research Institute of Ghana (CRIG) and transported to the laboratory.

Table 1: Cocoa Bean Samples in Training Set and Prediction Set

Cultivars	Number of Samples		Total
	Training Set	Prediction Set	
IMC85 x IMC47	42	28	70
PA7 x PA150	33	22	55
PA150 x Pound7	42	28	70
Pd10 x Pd15	42	28	70
T63/967 x T65/238	33	22	55
Total Samples	192	128	320

Spectra Collection and Preprocessing

The spectra of each sample were collected in the reflectance mode using the Antaris II Near Infrared Spectrophotometer (Thermo Electron Company, USA) with an integrating sphere. Samples of raw cocoa beans were scanned five times after 70^0 rotation. The whole experiments were conducted at ambient temperature and humidity kept at steady state. Each spectrum was an average of 32 scans with a spectra range of 10000-4000 cm^{-1} and raw data sets were measured in 3.856 cm^{-1} interval, resulting in 1557 variables. Standard normal variate (SNV) method was selected for pre-processing raw spectra to remove slope variation and to correct for scatter effects [5].

Statistical Analysis

All statistical calculations and algorithms were carried out in Matlab Version 7.14 (Mathworks Inc., USA) with Windows 7 ultimate for data processing. Antaris II System (Thermo Electron Company, USA) was used for spectra acquisition.

RESULTS

Spectra Preprocessing

Fig 1a presents the raw spectral information of the cocoa bean cultivars used. The raw spectral profile from the cocoa bean samples as shown in Fig 1a were pre-processed with SNV before further analysis to remove noise and improve the NIR spectra results as seen in Fig 1b.

Spectra Investigation

Fig 1a presents the raw spectra of the cocoa bean samples and this revealed water absorption bands around 5312 cm^{-1} and 7202 cm^{-1} corresponding to first overtone region which is O-H stretching and +O-H deformation. These regions were eliminated during the analysis together with other regions

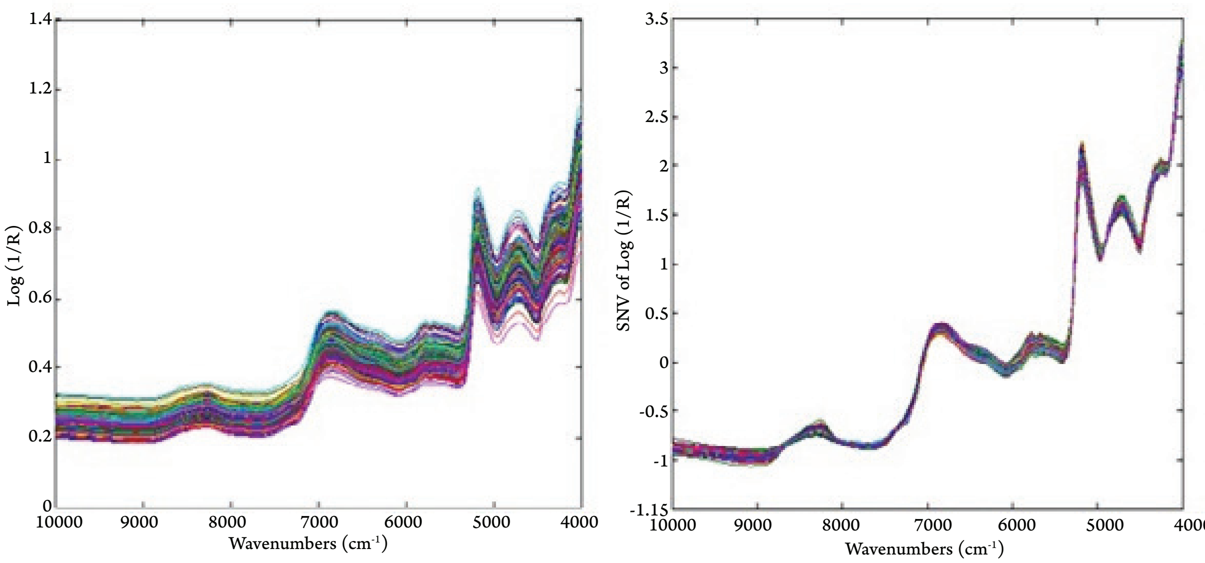

Figure 1: Spectra of Five Cocoa Bean Cultivars From (a) Raw Data and SNV Pre-processing Data

(10000-9000 cm^{-1} and 5000-4000 cm^{-1}) showing high level of noise. Therefore the rest of the band in the spectrum that could provide useful information belongs to the vibration range of 5000-9000 cm^{-1}. In these vibrational range, there are Carbonyl group, C-H stretch & C-H deformation, S-H, N-H, CH_2 and $-CH_3$ corresponding to phytochemicals such as polyphenols, proteins, alkaloids, volatile & non-volatile acid and other compounds found in cocoa beans.

Chemometric Methods

Unsupervised Classification Methods

All the spectra data set from the NIR was used in the Principal component analysis, PCA is a linear and an unsupervised pattern recognition method for visualizing data trends in a dimensional space. It is a for feature reduction technique which is the foundation for multivariate data treatment [13]. It works by reducing dimensions of the data matrix and compressing the information into interpretable variables called principal components (PCs) which are orthogonal [14]. To observe a clear cluster trends of the sample used, a scatter plot was obtained using the topmost three principal components (PC1, PC2 and PC3). This brings out important information and removes non useful ones, and similar samples were clustered closer to each other; this can be seen in Fig 2. In this regard, the visual graphical output could provide information that can be used for determining differences within and between cluster trends.

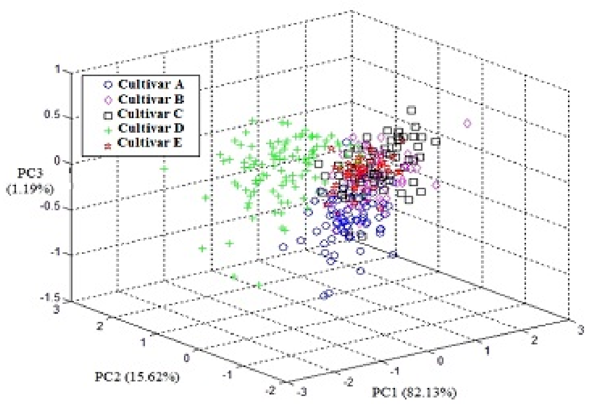

Figure 2: Score Plot of the Topmost Three Principal Component for All Sample

A = IMC85 x IMC47, B = PA7 x PA150, C = PA150 x Pound7, D = Pd10 x Pd15, E = T63/967 x T65/238

Supervised Classification Methods

Linear Discriminant Analysis (LDA)

LDA is the most frequently used technique (linear and parametric) among the supervised pattern recognition methods. It is commonly used to find linear combination of features which best differentiates two or more groups of events. The principle of LDA is based on the determination of linear discrimination functions which brings out clearly the ratio between class variance and reduces the ratio of within-class variance [15]. According to Berrueta and co-workers in LDA, classes are supposed to follow a multivariate normal distribution and be linearly separated [16], LDA can be considered as PCA. However the number of principal component factor is crucial to the performance of LDA discrimination model. In this study, PCA was used as an input data in LDA model. The samples used were grouped into training set (192 samples) and prediction set (128 samples) as shown in Table 1. The training set was used to build the model while the prediction set was used to test the model. It can be seen in Table 2 that the performance of LDA was good however at a high number of PCs and the efficiency of the model is determined by the number of PCs.

Support Vector Machine

SVM is a non-linear supervised learning method which was developed by Vapnik and co-workers for two-group classification problems [17]. It works by obtaining the optimal boundary of two groups in a vector space independent on the probabilistic arrangements of vectors in training set. When the linear boundary in low dimension input space is not enough to separate two classes, SVM can create a hyperplane that allows linear separation in the higher dimension feature space [16]. In this study, PCs obtained from PCA were used as input and samples were grouped into training set and prediction set as shown in Table 1 and Table 3 presents the performance of SVM model for the two sets.

Table 2: The Performance of LDA Model for Cocoa Bean Samples

Principal Components (PCs)	Discrimination Rate (%)	
	Training Set	**Prediction Set**
1	44.9	44.5
2	62.9	62.3
3	91.5	92.3
4	94.7	94.2
5	**93.1**	**93.9**
6	93.6	93.4
7	93.1	93.5
8	92.6	92.3
9	96.8	97.3
10	98.4	97.3

Table 3: The Performance of SVM Model for Cocoa Bean Samples

Principal Components (PCs)	Discrimination Rate (%)	
	Training Set	**Prediction Set**
1	47.9	41.8
2	86.8	83.3
3	95.8	94.4
4	99.3	98.6
5	**100**	**100**

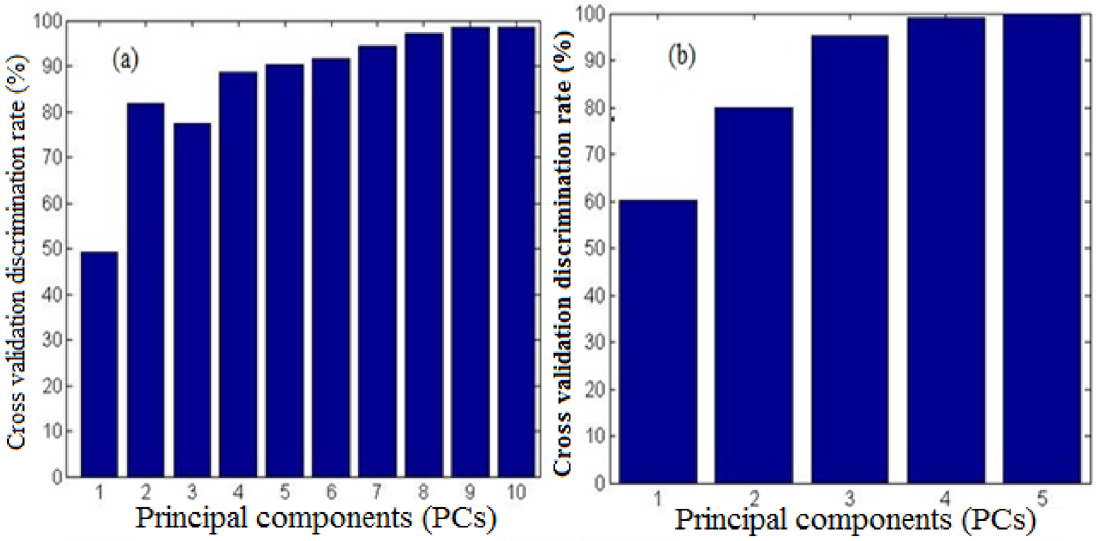

Figure 3: Cross Validation Discrimination Rate of (a) LDA Model and (b) SVM Model

DISCUSSION

Fig 2 showed the total variance contribution rate was 98.92% for the topmost three PCs which was higher than the one for Tea varieties [18], meaning PC1, PC2, and PC3 could extract the chemical compositional information in the pre-processed NIR spectra. The maximum total variance was very satisfactory though there was slight overlapping in the cluster trends. This could be explained by the slightly similar differences in chemical properties that existed in each as a result of their botanical properties. However, PCA is not a classification method hence; the top PCs provided an input data for the two classification models used. Since numbers of PCs influence the performance of the model as seen in Tables 2 and 3, the performance of the models were cross validated to ensure its stability and the outcome for both LDA and SVM are shown in Fig 3a and 3b respectively.

With respect to the two multivariate algorithm techniques used, SVM non-linear model was superior to the LDA linear one as shown in Table 3. The optimal performance of SVM model was 5 PCs compared to 10 PCs for LDA. High number of PCs could result in low generalisation in the performance. On the other hand, SVM has a stronger capability of self-learning and self adjustment. As in the case of cocoa beans varieties there are complex mixtures that are inter & intra connected so that linear algorithms is not effective in such situation. Furthermore, SVM model embodies structural risk minimization principle where upper bound is reduced on the expected risk [19, 20].

The following conclusions can be drawn from this research; Near Infrared Spectroscopy technique coupled with chemometric techniques is a powerful tool for accurate identification of cocoa bean cultivars. Support vector machine (SVM) performed better than the linear discriminant analysis (LDA) in developing discrimination models. It is concluded that FT-NIR Spectroscopic tool coupled SVM model can be exploited for the qualitative authentication of cocoa bean cultivars to aid quality control in the cocoa industry.

ACKNOWLEDGEMENTS

The authors are grateful University of Cape Coast for their financial support. The supply of cocoa beans varieties by Cocoa Research Institute of Ghana (CRIG) is also cordially acknowledged.

CONFLICTS OF INTEREST

There is no conflict of interest for the present study.

REFERENCES

1. Schinella G, Mosca S, Cienfuegos-Jovellanos E, Pasamar MÁ, Muguerza B, Ramón D, et al. Antioxidant properties of polyphenol-rich cocoa products industrially processed. Food Res Int. 2010;43(6):1614-23. DOI: 10.1016/j.foodres.2010.04.032
2. Brunetto M, Gutiérrez L, Delgado Y, Gallignani M, Zambrano A, Gómez Á, et al. Determination of theobromine, theophylline and caffeine in cocoa samples by a high-performance liquid chromatographic method with on-line sample cleanup in a switching-column system. Food Chem. 2007;100(2):459-67.
3. Mazor Jolić S, Radojčić Redovniković I, Marković K, Ivanec Šipušić Đ, Delonga K. Changes of phenolic compounds and antioxidant capacity in cocoa beans processing. Int J Food Sci Technol. 2011;46(9):1793-800. DOI: 10.1111/j.1365-2621.2011.02670.x
4. Nazaruddin R, Seng L, Hassan O, Said M. Effect of pulp preconditioning on the content of polyphenols in cocoa beans during fermentation. Indust Crop Produc. 2006;24(1):87-94.
5. Chen Q, Zhao J, Zhang H, Muhua L, Fang M. Qualitative identification of tea by near infrared spectroscopy based on soft independent modelling of class analogy pattern recognition. Near Infrared Spectroscop. 2005;13(6):327-32.
6. Veselá A, Barros A, Synytsya A, Delgadillo I, Čopíková J, Coimbra M. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. Analyt Chim Act. 2007;601(1):77-86.
7. Moros J, Iñón F, Garrigues S, de la Guardia M. Near-infrared diffuse reflectance spectroscopy and neural networks for measuring nutritional parameters in chocolate samples. Analyt Chim Act. 2007;584(1):215-22.
8. Tarkošová J, Čopíková J. Fourier transform near infrared spectroscopy applied to analysis of chocolate. J Near Infrared Spectroscop. 2000;8(4):251-7.
9. Teye E, Huang X. Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection

multivariate regression. Food Analyt Method. 2015;8(4):945-53.
10. Teye E, Huang X, Sam-Amoah LK, Takrama J, Boison D, Botchway F, et al. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. Food Chem. 2015;176:403-10.
11. Teye E, Huang X, Takrama J, Haiyang G. Integrating NIR Spectroscopy and Electronic Tongue Together with Chemometric Analysis for Accurate Classification of Cocoa Bean Varieties. J Food Process Engine. 2014;37(6):560-6.
12. Sunoj S, Igathinathane C, Visvanathan R. Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. Comput Electron Agric. 2016;124:234-42.
13. Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Analysis. 2007;44(3):683-700. DOI: 10.1016/j.jpba.2007.03.023
14. Luna AS, da Silva AP, Pinho JSA, Ferré J, Boqué R. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. Spectrochimic Act Part A Mol Biomol Spectroscopy. 2013;100(0):115-9. DOI: 10.1016/j.saa.2012.02.085
15. Chen Q, Cai J, Wan X, Zhao J. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. LWT Food Sci Technol. 2011;44(10):2053-8. DOI: 10.1016/j.lwt.2011.05.015
16. Berrueta LA, Alonso-Salces RM, Héberger K. Supervised pattern recognition in food analysis. J Chromatography A. 2007;1158(1–2):196-214. DOI: 10.1016/j.chroma.2007.05.024
17. Cortes C, Vapnik V. Support-vector networks. Machine Learn. 1995;20(3):273-97.
18. Chen Q, Zhao J, Liu M, Cai J. Nondestructive identification of tea (Camellia sinensis L.) varieties using FT-NIR spectroscopy and pattern recognition. Czech J Food Sci Vol. 2008;26(5):360-7.
19. Chen Q, Zhao J, Lin H. Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. Spectrochimic Act Part A Mol Biomol Spectroscopy. 2009;72(4):845-50. DOI: 10.1016/j.saa.2008.12.002
20. Lin H, Chen Q, Zhao J, Zhou P. Determination of free amino acid content in Radix Pseudostellariae using near infrared (NIR) spectroscopy and different multivariate calibrations. J Pharmaceut Biomed Analysis. 2009;50(5):803-8. DOI: 10.1016/j.jpba.2009.06.040