Food Research International 55 (2014) 288–293

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans

Ernest Teye [a,b], Xing-yi Huang [a,*], Wu Lei [a], Huang Dai [a]

[a] School of Food and Biological Engineering, Jiangsu University, Xuefu Road 301, Zhenjiang 212013, Jiangsu, PR China
[b] School of Agriculture, Department of Agricultural Engineering, University of Cape Coast, Cape Coast, Ghana

ARTICLE INFO

Article history:
Received 7 June 2013
Accepted 13 November 2013

Keywords:
FT-NIR spectroscopy
Cocoa beans
Identification
Quantification
Adulteration
SVM
Si-PLS

ABSTRACT

Fourier transform near-infrared (FT-NIR) spectroscopy combined with Support Vector Machine (SVM) and synergy interval partial least square (Si-PLS) was attempted in this study for cocoa bean authentication. SVM was used to develop an identification model to discriminate between fermented cocoa beans (FC), unfermented cocoa beans (UFC) and adulterated cocoa bean (5–40 wt/wt.% content of UFC). Si-PLS model was used to quantify the addition of UFC in FC. SVM model accurately discriminated the cocoa bean samples used. After cross-validation, the optimal identification rate was 100% in both the training set and prediction set at three principal components. For quantitative analysis, Si-PLS model was evaluated according to root mean square error of prediction (RMSEP) and coefficient of correlation in prediction (R_{pred}). The results revealed that Si-PLS model in this work was promising. The optimal performance of Si-PLS model showed an excellent predictive potential, RMSEP = 1.68 and R_{pred} = 0.98 in the prediction set. The overall results indicated that FT-NIR spectroscopy together with an appropriate multivariate algorithm could be employed for rapid identification of fermented and unfermented cocoa beans as well as the quantification of UFC down to 5% in FC for quality control management.

1. Introduction

Cocoa beans are from the pod of a tree crop known as *Theobroma cacao*. It is mainly cultivated in tropical and subtropical countries. In terms of world production, Ivory Coast and Ghana contribute about 65%. Cocoa bean is the main raw material for chocolate and confectionary products. Quality assurance of cocoa beans has attracted worldwide interest in order to control and prevent cocoa bean adulterations with regard to wrongful mixing of fermented cocoa beans with unfermented ones. After harvesting of the cocoa pod, the beans are taken out and fermented for 4–6 days and dried to a moisture content of 7%. This process is very important because it improves the final quality of the product and the processors always demand fermented cocoa beans. However, some farmers willfully mix unfermented cocoa beans with fermented ones especially when they are under pressure to meet the demand from buyers (Aikpokpodion & Dongo, 2010). Unfermented cocoa bean has astringent taste and unpleasant flavor, hence needs to be fermented (Nielsen, Snitkjaer, & van den Berg, 2008). Fermentation is a very vital first postharvest activity which leads to cocoa beans with the desired chocolate characteristics and this step is known to significantly affect the physicochemical properties of the cocoa bean (Aremu, Agiang, & Ayatse, 1995; Saltini, Akkerman, & Frosch, 2013). Comparatively, unfermented cocoa bean has higher polyphenols, starch and unpleasant taste and flavor (Aculey et al., 2010; Kim & Keeney, 2006; Wollgast & Anklam, 2000). Studies revealed that, chocolates produced from unfermented cocoa beans have no cocoa flavor and are very bitter and astringent (Biehl & Voigt, 1996). This problem is of great concern to chocolate producers and cocoa processors in general. The analytical methods used for cocoa bean examination such as: high performance liquid chromatography (Adamson et al., 1999; Kim & Keeney, 1983; Ramli, Yatim, Said, & Hok, 2001), fermentation index (Ilangantileke, Wahyudi, & Bailon, 1991; Romero-Cortes et al., 2013), sensory test, gas chromatography mass spectroscopy, and cut test (Aculey et al., 2010), although reliable and accurate, they are often, tedious, time consuming, expensive, destructive and require an elaborate sample preparation. Therefore, there is a search for a solution to overcome these short comings. Optimistically, Near Infrared Spectroscopy (NIR) has shown a great potential to provide the solution.

NIR spectroscopy has been used for food analysis since 1938 (Huck, Guggenbichler, & Bonn, 2005; Williams & Stevensen, 1990). It is currently known to be an advanced and excellent analytical technique which has found its use for qualitative and quantitative analyses in the food and other industries (Blanco & Villarroya, 2002; Cen & He, 2007; Roggo et al., 2007). The advantages of NIR spectroscopy over the traditional analytical methods are: simple, rapid, non-invasive, semi/non-destructive, and environmentally friendly (no chemical involved) and requires minimal or no sample preparation (Esteban-Dıez, Gonzalez-

* Corresponding author. Tel.: +86 51188792368/00233 243170302.
E-mail addresses: teyernest@gmail.com (E. Teye), h_xingyi@163.com (X. Huang).

0963-9969/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodres.2013.11.021

Saiz, & Pizarro, 2004; Rodriguez-Saona & Allendorf, 2011; Vesela et al., 2007).

Numerous studies have used NIR spectroscopy to investigate authentication of various foods such as; tea (Chen, Zhao, Liu, & Cai, 2008), coffee (Downey, Boussion, & Beauchêne, 1994) and other uses including; the detection of the addition of barley to coffee (Ebrahimi-Najafabadi et al., 2012), coffee varietal differentiation (Esteban-Diez, Gonzalez-Saiz, Saenz-Gonzalez, & Pizarro, 2007), qualitative identification of tea categories (Chen, Zhao, & Lin, 2009), discrimination between fermented and unfermented rooibos (Schulz, Joubert, & Schütze, 2003), quantification of fat, nitrogen and moisture of cocoa powder (Vesela et al., 2007), predictive analysis of cocoa procyanidins (Whitacre et al., 2003), differentiation of Ghana cocoa beans (Teye, Huang, Dai, & Chen, 2013) and verification of cocoa powder authenticity (Trilcova et al., 2004) and compositional analysis of powdered cocoa product (Permanyer & Perez, 2008). These studies mentioned earlier show that, NIR spectroscopy technique could be utilized for the identification and determination of adulterants in cocoa beans. However, information on the application of NIR spectroscopy technology together with multivariate algorithms for the identification and detection of fermented, unfermented and adulterated cocoa beans is not available in literature up to now. The objectives of this study therefore were to use FT-NIR spectroscopy together with Support Vector Machine (SVM), partial least squares regression models for the identification of fermented and unfermented cocoa beans and the prediction of the fraudulent addition of unfermented cocoa beans in fermented cocoa bean samples respectively. SVM multivariate technique was selected for the identification problem because it is one of the frequently used supervised pattern recognition methods and it has proven to provide excellent result compared with others (Chen et al., 2009; Teye et al., 2013). The prediction of unfermented cocoa beans in fermented cocoa beans was done with PLS models, either on full spectra (PLS) or subinterval spectra selection synergy interval partial least squares (Si-PLS). Si-PLS model has been used in recent times for food products and found to be superior to others (Cai, Chen, Wan, & Zhao, 2011; Chen, Zhao, Liu, Cai, & Liu, 2008).

2. Materials and methods

2.1. Sample and sample preparation

In this study, three main groups of samples were used: fermented cocoa beans ready for export (FC) = 26 samples, unfermented cocoa bean (UFC) = 26 samples and adulterated (ADC) = 80 samples. The samples were collected from different producers from all the cocoa growing regions of Ghana and from the shipyard of quality control division of the Ghana cocoa board. The samples were then sealed in air-tight bag and transported from Ghana to China; Jiangsu University (School of Food and Biological Engineering laboratory) for further analysis. The adulterated samples were prepared in the laboratory. All the samples were ground separately for 15 s in multi-purpose grinder (QE-100, Zhejiang YiLi Tool Co., Ltd, China). Grinder container was allowed to cool each sample to avoid the loss of volatile compounds. Each ground sample was sieved with a 500 μm mesh before further analysis. The adulterated samples were prepared by accurately weighing portions of 5–40 wt/wt.% content of UFC in FC content and shaken vigorously to ensure total homogenization. ADC was prepared to cover as much the perceived range as possible that pertains in the market. The samples were immediately analyzed.

2.2. FT-NIR spectral analysis

The spectrum of each sample was taken in the reflectance mode using the Antaris II Near Infrared Spectrophotometer (Thermo Electron Company, USA) with an integrating sphere. 10 g of each sample was collected into a standard sample cup and scanned three times after rotating the cup for 120°. The whole experiments were conducted at an ambient temperature of 27 ± 1 °C and a relative humidity of 54%. Each spectrum was an average of 32 scans with a spectra range of 4000–10,000 cm^{-1} and the raw data set was measured in 3.856 cm^{-1} interval, resulting in 1557 variables.

2.3. Chemometric techniques

All algorithms were carried out in Matlab Version 7.14 (Mathworks Inc., USA) with Windows 7 ultimate for data processing. The chemometric techniques used in this research were: Standard Normal Variate (SNV), Direct Orthogonal Signal Correction (DOSC), Principal Component Analysis (PCA), Support Vector Machine (SVM) and Partial Least Squares (PLS) models. PCA of the spectra was performed after preprocessing with SNV and DOSC. SNV was used to remove slope variation and to correct scatter effects (Barnes, Dhanoa, & Lister, 1989), while DOSC method developed by Westerhuis and co-workers was also done to remove spectral variance sources (Luypaert, Heuerding, Massart, & Heyden, 2007; Westerhuis, de Jong, & Smilde, 2001). PCA is an exploratory method that reduces the dimension of the data matrix and compresses the information into a few interpretable variables called principal components (PCs), which are linear combinations of the original variables (Luna, da Silva, Pinho, Ferré, & Boqué, 2013). The first three PCs (PC1, PC2 and PC3) give useful information on the score plot such that, similar samples are clustered closer to each other. SVM was performed to solve the classification problem after PCA gave the cluster trend of the samples. SVM is a non-linear supervised pattern recognition method, which embodies structural risk minimization principle and has a strong potential of self-learning and self-adjustment. It is described elsewhere (Chen, Zhao, Fang, & Wang, 2007), while PLS models (PLS and Si-PLS) were used for developing the predictive model for the measurements of unfermented cocoa bean in fermented ones as used by other authors (Chen, Zhao, Liu, & Cai, 2008; Chen, Zhao, Liu, Cai, & Liu, 2008).

2.4. Calibration and prediction set

The data set used in this experiment was made up of 132 samples. These were divided into two subsets called: calibration set (90 samples) and prediction set (42 samples). The calibration set was used to develop the model, while the prediction set was used for evaluating the actual predictive ability of the developed models. The individual samples in each set were selected in order to come to a 2/1 division of calibration set/prediction set. To avoid bias in subset division, the subset was done as follows: for every five samples, three spectra were randomly selected as the calibration set while the remaining samples were used as the prediction set.

2.5. Classification model

In this study, three main cocoa bean groups were classified namely; fermented, unfermented and adulterated samples. SNV and DOSC preprocessing methods were comparatively applied as the initial data pretreatment techniques. After PCA, SNV and DOSC provided a new data variable for building the SVM classification model. The number of PCs as an important parameter was optimized by cross validation. Cross validation was done by the leave one out procedure where: in the training set, one spectrum is deleted and the model is built with the remaining spectra in the training set. Then the left out sample is predicted with the model and the sequence is repeated with leaving each of the samples of the training set out.

2.6. Quantification models

In this section, the model for prediction of the amount of UFC in FC (% weight basis) was developed using Si-PLS technique. This model was compared with the classical partial least square (PLS). The calibration