

Predicting adulteration of Palm oil with Sudan IV dye using shortwave handheld spectroscopy and comparative analysis of models

Roseline Love MacArthur^{a,b}, Ernest Teye^{a,*}, Sarah Darkwa^b

^a University of Cape Coast, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

^b University of Cape Coast, Faculty of Science & Technology Education, Department of Vocational and Technical Education, Cape Coast, Ghana

ARTICLE INFO

Keywords:

Palm oil
Integrity
Sudan IV dye
Handheld spectroscopy
Multivariate technique

ABSTRACT

Shortwave handheld NIR spectroscopy coupled with multivariate algorithm was attempted to simultaneously classify and measure Sudan IV dye adulteration in palm oil samples. K-nearest neighbour (KNN) was used to develop a classification model to discriminate between authentic palm oil samples and Sudan IV dye adulterated (0.10–0.002 % w/w) ones. Principal component regression (PCR), partial least square regression (PLSR) and support vector machine regression (SVMR) algorithms were comparatively employed to quantify the addition of Sudan IV dye in authentic palm oil samples. The models were evaluated by the classification rate (R), correlation co-efficient in the calibration set/prediction set (R_p^2/R_c^2), and root mean square error of calibration/prediction (RMSEC/P). The developed multiplicative scatter correction plus KNN technique was found to accurately classify where $R = 95.48\%$ and 97.00% in the calibration set and prediction set respectively. Among the quantification model developed for measuring Sudan IV dye, standard normal variant preprocessing plus partial least square regression (SNV-PLSR) gave the best performance at $R_c^2 = 0.91$, $R_p^2 = 0.90$ and RMSEC = 0.0841, RMSEP = 0.0868 while standard normal variant preprocessing plus principal component regression gave $R_c^2 = 0.90$, $R_p^2 = 0.90$ and RMSEC = 0.0846, RMSEP = 0.870. The findings have proved that, the integrity of palm oil samples can be certified rapidly and nondestructively in terms of the presence of Sudan IV by using short wave handheld NIR spectroscopy. This offers the opportunity for incorporating NIR spectroscopy into mobile phone devices to enhance mobile detection.

1. Introduction

Palm oil obtained from *Elaeis guineensis* fruits is West Africa's most important edible oil from plant origin. It is considered a nutritious condiment in various local dishes that provides additional vital medicinal properties [1]. Nevertheless, the consumption of palm oil has been challenged with the wrongful and fraudulent addition of Sudan I V dye. Often this adulteration is done to improve the consumer perception of the oil as the addition improves colour. However, this activity popularly known as food fraud in the palm oil industry has become a global challenge. In the EU, USA and almost all other countries, the addition of Sudan I V dye to food commodities including palm oil is banned. This dye also is known as oil-soluble azo dye was originally used for imparting a rich-red, red-orange or yellow colour to textiles and plastic has found its wrongful use to adulterate spices, tomato paste, palm oil and others for the purpose of enhancing colour perception of consumers [2]. In recent time there have been reported cases of palm oil adulteration with Sudan IV dye originating from West

Africa. This led to the ban of most of the palm oil from West Africa. More specifically, Ghana, one of the finest producers of palm oil has become the culprit of Sudan IV dye adulteration leading to a ban by some foreign countries [3]. Furthermore, consumers in Ghana have been alerted by the Food and Drugs Authority to buy palm oil from reputable registered sources as market survey test conducted revealed that samples collected show the presence of Sudan dyes [4]. Sudan dyes are categorized as group 3 carcinogens with high evidence of causing cancer and other related illnesses when ingested by humans. Hence the addition of the dye to any form of food commodity renders it unfit for human consumption. Currently, most food commodities such as palm oil and spices imported into the European Union must have an accompanying report of the absence of Sudan dye [2]. This act has further resulted in a big challenge as most processors in African countries are faced with the difficulty of accessing laboratory infrastructure for their test and authentic report. Furthermore, the very few processors who attempt to send their samples to centralized laboratories are forced to wait for a long time to get their samples tested. This is as a result of the

* Corresponding author.

E-mail address: ernest.teye@ucc.edu.gh (E. Teye).

<https://doi.org/10.1016/j.vibspec.2020.103129>

Received 2 May 2020; Received in revised form 11 July 2020; Accepted 16 July 2020

Available online 18 July 2020

0924-2031/ © 2020 Elsevier B.V. All rights reserved.

slow and tedious nature of the wet chemistry analytical techniques applied. On the other hand, the known techniques used to measure Sudan dye in foods as stated by Teye and co-workers [5] have their disadvantages which do not favour rapid and timely detection of Sudan dye in palm oil and it is a disincentive for onsite detection at an affordable rate.

These aforementioned challenges, therefore, called for an alternative novel technique and near infrared spectroscopy (NIRS) could provide the solution. NIRS technique has found its usage in several fields such as petrochemical, agricultural, fisheries, and food processing industry among others to provide a rapid analytical methods for measurements [6]. Its usage is simple, fast, requires simple sample preparation, no usage of chemicals, and with little technical know-how. NIRS has been successfully used for measuring Sudan dye in chilli powder [7], palm oil adulteration with lard [8], adulteration of walnut oil [9], prediction of palm oil acidity [10] and determination of free fatty acids in palm oil [11]. Others include; determination of oil content in palm fruits [12] and the prediction of fat content in cocoa beans [13].

However, little or no studies have been conducted with NIRS for measuring Sudan IV dye in palm oil samples. The only available studies attempted classifying adulterated palm oil and authentic ones using NIRS in the range of 900–1700 nm with optimum results [5]. The objective of this research was to attempt to simultaneously differentiate and quantify Sudan IV dye in palm oil samples using portable short-wave NIRS (740–1070 nm) and multivariate algorithms.

2. Materials and methods

2.1. Sample collection and preparation

Pure palm oil samples (unadulterated) were collected from different locations at processing centres in Ghana (collected immediately after processing at the factory for surety of its purity) while the adulterated palm oil samples were made in the laboratory by spiking with different concentrations (1% - 0.002 % w/w) of Sudan IV dye (Fisher Scientific, Loughborough, England) and Fig. 1. Show all the samples used. In all, a total of 150 samples were used for the studies comprising: 30 samples of pure palm oil and 12 samples each of the spiked/ adulterated samples.

2.2. Sample spectral acquisition

The spectrum of each sample was collected in the reflectance mode by using a handheld NIR spectrometer (SCIO™ model CP-SC006) in a spectra range of 740 nm–1070 nm in a 1-nm resolution for spectra data recording. For each sample of the palm oil, 30 g was poured into glass container and scanned five times after rotating the sample cup. Before scanning the samples, the spectrometer was calibrated using the

reference cup provided by the manufacturer. The whole process was carried out at an ambient temperature of 30.3 °C with a humidity of 68 % at the Technology village laboratory of the School of Agriculture, University of Cape Coast.

2.3. Statistical analysis

Spectral data recordings stored in the cloud were downloaded unto the computer and imported into chemometric software packages (Unscrambler X 10.4; Camo Analytics, Oslo, Norway and MATLAB 2016a; MathWorks Inc., USA) using windows 10 Basic for all data processing. The analyses of data were performed by comparing different chemometric methods to determine the most suitable procedure to achieve simultaneous differentiation and quantification challenges of Sudan IV detection.

2.4. Preprocessing techniques

There are several preprocessing techniques, however, there are no clear cut standard procedures to decide the best pre-processing technique according to the equipment used for spectra acquisition and samples investigated, requiring a trial and error approach for specific applications [10]. In this study, raw spectra and other preprocessing methods such as: standard normal variant (SNV), second derivative plus Savitzky-Golay (SG) and multiplicative scattered correction (MSC) were used independently. SNV is a spectral data correction technique which possesses the power of removing slope variations and to correct scatter light effect [14] and for more information refer [15]. Second derivative is used for increasing spectra resolution and reveals unique spectral peaks. The second derivative spectra were derived with a Savitzky-Golay filter using a second-order polynomial and a 17-point window. While MSC is a tool used for the correction of both multiplicative and additive scatter effects on different particle sizes and inclination of baseline variation [15].

2.5. Data partition

In this research, the raw data set comprises a total of 150 samples. This data set was subdivided into two sets; training set (90 samples, used for building the model) and test set (60 samples; used for testing the predictability of the model). To avoid bias in selecting members for each sub-group, a well-defined approach used by other authors was used [5]. Thus; to avoid the effect of randomness caused by random sampling, the Kennard-Stone algorithm was used to systematically partition the data set into training set (calibration) and test set (prediction) as done by others [16] using MATLAB version 2017a..



Fig. 1. Palm oil samples showing different levels of Sudan dye adulterations.

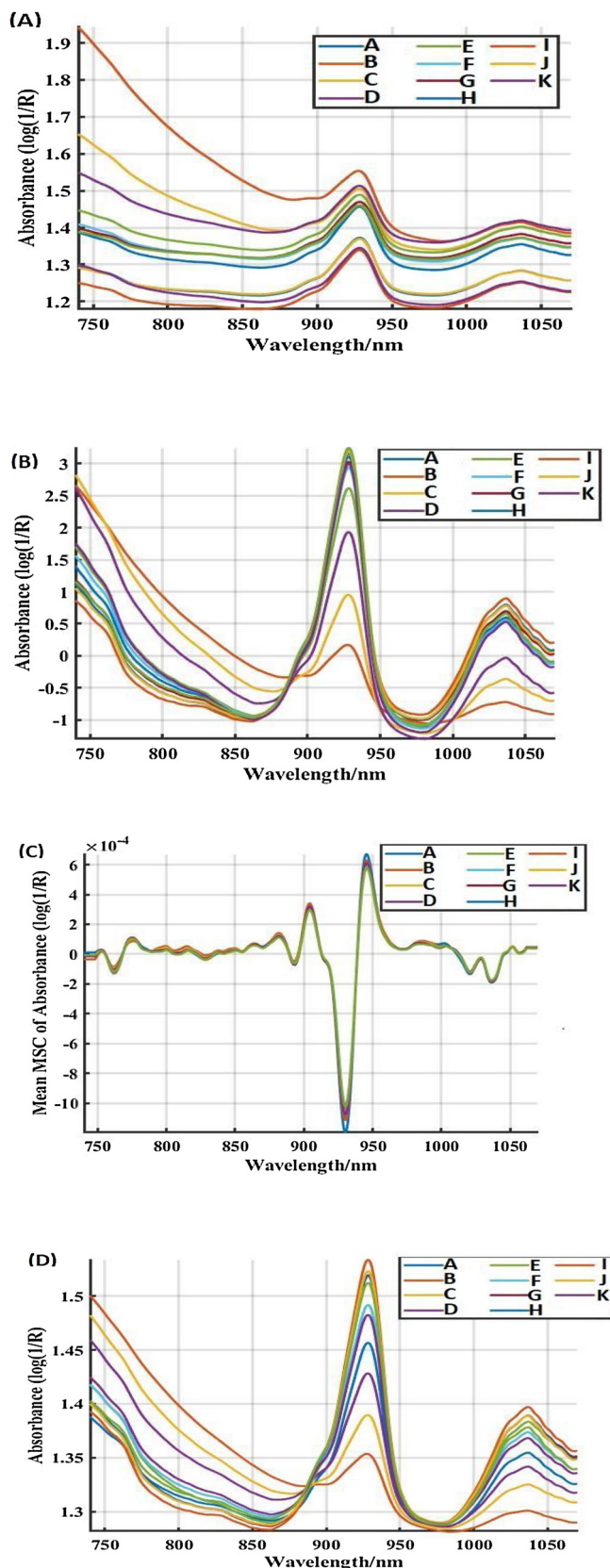


Fig. 2. Spectral profile (A) raw, (B) SNV, (C) 2nd derivative, and (D) MSC of samples.

2.6. Multivariate algorithms

2.6.1. K-nearest neighbour (KNN)

KNN is a non-parametric linear tool that performs its functions by measuring the differences or similarities between two stances. That is, it tries to classify an unknown sample based on the known classification of its neighbours, popularly known as equals with equals easily associate. It is one of the simplest supervised classification algorithms that assume that similar identities exist in close proximity. However, the major drawback of KNN is its ability to only use local prior probabilities to predict instance labels and therefore it fails to account for class distribution around the neighborhood of query instance [17]. It cannot work well if large differences are present in the number of samples in each class [18]. On the other hand it is one of the best classification tools when there is little or no prior knowledge about the distribution in the sample data set. For more information on KNN refer Berrueta and co-workers [18].

2.6.2. Principal component of regression (PCR)

Principal component regression is a multivariate algorithm which is normally used for collinear data set. The idea of PCR is to regress a response variable on the principal component of the predictor variable in sequence based on the variances of the components rather than the original predictor variables [19]. PCR solves collinearity problems and can eliminate the lesser principal components that allow some level of reducing random error. However, it has the risk of discarding useful information that ends up in used principal components. For more information about PCR kindly refer [20].

2.6.3. Partial least square regression (PLS-R)

Partial least square is a multivariate algorithm that was first introduced to functional neuroimaging [21] and has now found its use in a wide range of fields. It is bilinear modeling tool which functions by projecting the original independent variable (X-information) onto a small number of the latent variable (LV) to simplify the relationship between X and Y for predicting with the smallest number of latent variable [22]. PLS is associated with either classification (known as PLS-DA) or regression problems (known as PLSR) and its main advantage is that it can analyze data with strongly collinear, noisy and with numerous X-variables, in addition to simultaneously modelling several response variables (profiles of performance) [23]. It has also found its use in solving real multivariate challenges in the food industry. For more information refer to Krishnan and others [23,24].

2.6.4. Model validation

The performance of the classification model (KNN) was evaluated according to the prediction rate (%), sensitivity, specificity, precision in both the training set and the test set. The last three parameters depend on the values of true positive (TP), false positive (FP), true negative (TN) and false negative (FN). In a targeted class samples, TP and TN represent correctly identified samples in positive and negative classes respectively [16]. While the quantification model by using PCR and PLS were evaluated by co-efficient of correlation (R), root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and ratio of performance-to-deviation (RPD). All these parameters were calculated by equations 1–7

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Where: (4) } RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5) RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

$$(6)R = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)RPD = \frac{SD}{RMSEP}$$

n = the number of samples

y_i = the reference measurement results for sample i ,

\hat{y}_i = the estimated result for sample i when the model is constructed with sample i removed

\bar{y} = the estimated results of the model for the sample i ,

\bar{y} = the mean of the reference measurement results for all samples

SD = standard deviation

3. Results and discussion

3.1. Spectral presentation

Palm oil samples obtained from processors in Ghana were spiked with Sudan IV dye at 1%, 0.5 %, 0.25 %, 0.125 %, 0.05 %, 0.0312 %, 0.0156 %, 0.0078 %, 0.0039 %, 0.0019 %, and 0% i.e. pure samples. Fig. 2(A–D) shows the mean spectra profile of Raw (B), SNV (B), 2nd derivative-SG (C) and MSC (D) preprocessed treatments of palm oil samples used in this study. It could be observed that each preprocessing technique revealed a unique spectral profiles for the palm oil samples. More so, it was observed from the figures that all the preprocessing treatments had a similar peak intensity around 935 nm. However, for 2nd derivative preprocessed treatment, there were several useful peaks along the entire wave band with spectra intensity for pure palm oil samples and adulterated ones compare to the others. This could be attributed to the strength of derivative techniques as it known to give spectra resolution enhancement [15]. On the other hand, the wavelength range of 900–950 nm is associated with methyl and ethenyl functional groups found in Sudan IV dye, fat and oils molecular structure (a vital component of palm oil). Other peaks revealed by 2nd derivative technique were 765 nm, 885 nm, 905 nm, 930 nm, 945 nm, 1000–1005 nm. These peaks could provide vital information for accurately for determining palm oil integrity with regards to Sudan IV dye adulteration. These aforementioned wavelengths correspond to C–H 4th overtone, CH₂ and CH₃ second overtone stretching bonds found in Sudan IV dye molecular structure. The use of NIR spectrometer in the range of 740–1070 nm provided some information in this study. The information within this region is suitable for food analysis as it is associated with hydroxyl bonds (O–H) and aliphatic chain (C–H) [25]. Furthermore, it could be observed that the spectral profile showed similar peaks for pure and adulterated samples. However few differences were observed at certain wavelengths. As Sudan IV dye adulteration concentration increases the peak position and peak intensity exhibited some differences at the wavelength ranges of 765 nm, 885 nm, 900–950 nm, and 1000–1005 nm. For instance, the concentrations of 1%, 0.5 %, and 0.1 % in the palm oil caused a brighter red colour impacted onto the samples. This colour impacted is perceived by the consumer as high-quality palm oil [2] and difficult to detect it onsite by using the sensory perceptions. However, the major peaks that could provide useful information for rapid qualitative and quantitative detection could be seen in Fig. 2. These wavelength ranges correspond to hydrogen bonds (C–H, OH, SH, NH—) made up of overtones and the combination of basic vibrational peaks observed in the spectra range used. The absorption around 765, 914, 964, 980 nm could also have corresponded to oil, cellulose, protein, and water [26] found in palm oil.

3.2. Principal component analysis (PCA)

PCA was done to bring out any possible patterns. Fig. 3 shows the MSC PCA score plot of palm oil samples (pure and adulterated). PCA is known as an unsupervised pattern recognition tool which is found its usefulness in bringing out vital information to separate similar samples by the principle of grouping similar samples closer to each other. To obtain a cluster trend, the topmost two principal components (PC1 &

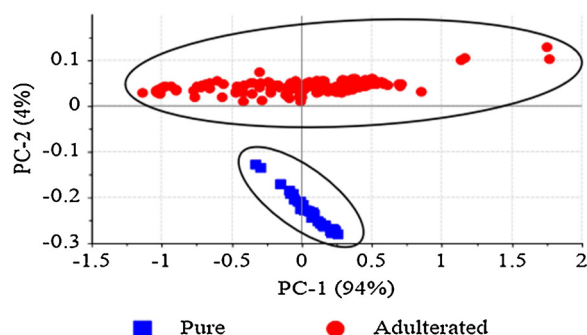


Fig. 3. MSC PCA score plot of palm oil samples.

PC2) were used. This brought out important information and removed non-useful ones, therefore similar samples were then grouped closer to each other. More so, the groupings provided a visual graphical output that could be used in determining differences within and between cluster trends. Comparing the effect of the preprocessing treatments on PCA cluster in our study, MSC PCA gave a clear separation of two groups which represented pure and adulterated palm oil samples as seen in Fig. 3. In this Figure, all the samples clustered clearly along the first two PCs plane where PC1 and PC2 can explain 94 % and 4% of the total variance in the samples respectively, giving a total accumulative contribution of 98 % variance for the 150 samples used in this study. In this study, the first principal component (PC1) covered the maximum information direction [27]. Furthermore, the two PCs could explain 98 % spectra information from all the samples used and this provided chemical compositional information in the spectra region. Pure and adulterated palm oils have considerable compositional differences in their chemical structures due the impact of Sudan IV dye on colour when used as an adulterant. After PCA, the samples were divided into training set and test set before they were inputted into the multivariate classification model.

4. Classification analysis

Supervised pattern recognition technique known as KNN was used to build a classification model for qualitative analysis of the palm oil samples. Table 1 shows the statistical performance of KNN model after pretreating data with SNV and MSC preprocessed technique. The results obtained showed that KNN performed better when MSC preprocessing data was used with a prediction rate above 95 % in both training set and test set at 3 PCs. Also, the sensitivity, specificity and precision were found to be closer to 1 (99, 95, and 98 respectively). This further revealed that the model developed is reliable and stable because, lower number of PCs normally leads to a better generalization performance of

Table 1

Statistical parameters of KNN models with various preprocessing techniques for authentication of palm oil samples.

Model Pre-processing	No. PCs	Evaluation	Performance	
			training set (90)	test set (60)
SNV	5	Prediction rate	77.83	73.44
		Sensitivity	0.88	0.82
		Specificity	0.32	0.22
		Precision	0.88	0.87
2 nd Der	7	Prediction rate	85.57	78.13
		Sensitivity	0.93	0.83
		Specificity	0.48	0.29
		Precision	0.88	0.90
MSC	3	Prediction rate	95.48	97.00
		Sensitivity	0.99	0.99
		Specificity	0.95	0.98
		Precision	0.98	0.97

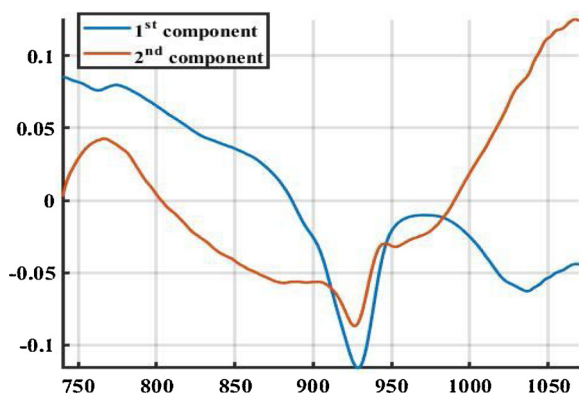


Fig. 4. PC loading score plot with two principal components.

a model [28]. The improvement of the performance of KNN by MSC compared to the others could be attributed to the effect of MSC in correcting multiplicative, additive scatter effects and inclination of baseline variation. The classification rate above 95 % provides useful concrete evidence that handheld NIR spectrometer could be used for qualitative analysis of palm oil integrity in terms of Sudan IV dye. To explain the phenomenon of the model's performance, PCA eigenvector analysis was performed on the first two PCs to reveal the major peak of influence as shown in Fig. 4. This revealed major peaks around 765 nm, 930 nm, 950 nm and 1040 nm which are associated with 1st overtone CH combination, and OH— bonding and CH₃ stretching found in the chemical structure of palm oil and Sudan IV dye [29]. Particularly the wavelengths around 930–960 nm are associated with 2nd overtone CH— stretching that could be related to functional groups of palm oil and a chemical structure in Sudan IV dye [5]. Generally, this study has provided some information that the shortwave spectrometer used could provide functional groups in the chemical formula of pure palm oil and elements in Sudan IV dye used.

Table 1 Statistical parameters of KNN models with various preprocessing techniques for authentication of palm oil samples

5. Quantitative analysis

Also, in this study the feasibility of quantifying levels of Sudan IV dye was investigated. Fig. 5 shows the best performance of two chemometric techniques for measuring the content of Sudan IV dye in palm oil samples. From this figure, PLSR performed slightly better than PCR. This could be ascribed to the fact that PLSR can analyze a strong collinearity in the data set better than PCR. Furthermore, the clouded values seen around zero (0) in Fig. 5 is ascribed to the fact that the pure samples were spiked with very small amount (0.002 %) of Sudan IV dye compared with the highest adulteration of 1% (hence a wide range). Also, from Table 2, the results obtained for RPD revealed some useful information for the models developed for palm oil integrity. In this study RPD was found to be between 2.2–3.4 with SNV performing better than MSC for both PCR and PLSR algorithms for the quantitative analysis. It could be explained that, SNV removed the slope variations and corrected the effect of scatter light that could have influenced the performance better than MSC. The best performance was obtained at RPD value above 3.4. This means the model is stable and good because the model's prediction stability and strength is considered poor if $RPD < 2$, general for $3 > RPD \geq 2$ while $RPD \geq 3$ are stable [30]. Furthermore, Li and co-workers [31] classify RPD values as: excellent model ($RPD \geq 4$), good model ($3 \leq RPD < 4$) and poor model ($RPD < 3$).

6. Conclusion

From the study, it is concluded that portable NIR spectroscopy could

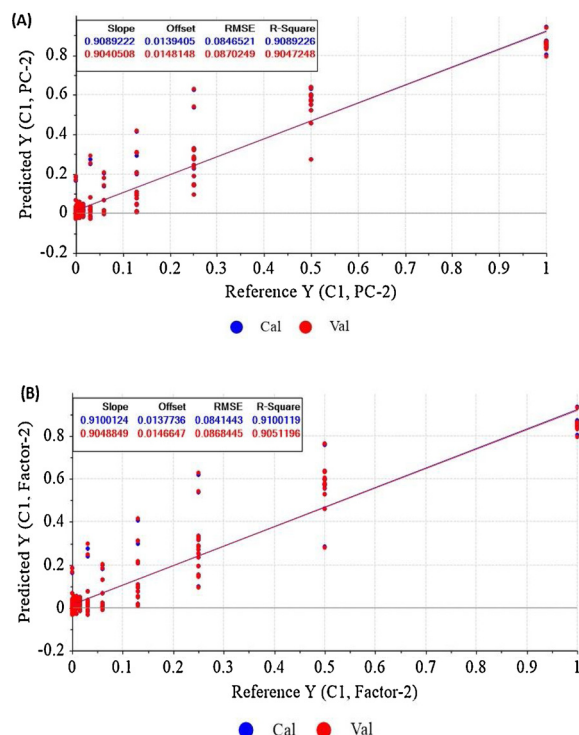


Fig. 5. The best performance of SNV plus PCR (A) and PLSR (B) model.

Table 2

Quantification of Sudan dyes (IV) adulteration in palm oil samples.

Models	Pp	PCs	Performance of regression model				
			Training set		Test set		
			R ²	RMSEC	R ²	RMSEP	RPD
PCR	MSC	5	0.81	0.122	0.80	0.125	2.219
	SNV	3	0.90	0.084	0.90	0.087	3.232
PLSR	MSC	3	0.88	0.098	0.85	0.110	2.556
	SNV	3	0.91	0.081	0.90	0.082	3.429

Pp = Preprocessing.

be used for certifying the integrity of palm oil in terms of the presence of Sudan IV dye. This feasibility study has a great potential of incorporating technology into mobile phone devices for market surveillance and onsite use. More specifically, among the preprocessing methods used MSC was able to improve the primary spectral data set for the classification model. KNN classification technique gave the optimum performance of 95.48 % in the calibration set and 97 % in prediction set. While for prediction of the levels of Sudan 1 V dye, SNV preprocessing improved the prediction model better. The results of all the regression models where PLSR gave $R^2 = 0.90$ and $RPD < 3.4$ in the prediction set. This study provides a feasibility of using shortwave handheld NIR spectrometer as an initial screening tool for quality control officer along the palm oil value chain. It will therefore provide a complementary tool to wet chemistry standard techniques.

Authors statement

R. L. MacArthur: Literature review, data collection, data analysis, and manuscript preparation.

E. Teye: Conception of manuscript, design of experiment and methodology, manuscript revision.

S. Darkwa: Data collection and manuscript revision.

Declaration of Competing Interest

The Authors: Roseline Love MacArthur, Ernest Teye, Sarah Darkwa declare that there is no conflict of interest

Acknowledgment

The authors highly appreciate the support provided by the Institute for Global Food Safety of Queen's University Belfast, UK and School of Agriculture, University of Cape Coast, Cape Coast, Ghana. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] O. Okogeri, Adulteration of crude palm oil with red dye from the leaf sheath of sorghum bicolor, *Food Sci. Qual. Manage.* 17 (2013) 1–6.
- [2] S. Genualdi, et al., Method development and survey of Sudan I–IV in palm oil and chilli spices in the Washington, DC, area, *Food Addit. Contam. Part A* 33 (4) (2016) 583–591.
- [3] R. RASFF, The Rapid Alert System for Food and Feed. Annual Report, (2008).
- [4] J. Amoako-Mensah, Assessment of the Prevalence of Palm Oil Adulteration with Sudan IV Dye in the Greater Accra Region, (2017).
- [5] E. Teye, et al., Rapid and nondestructive fraud detection of palm oil adulteration with Sudan dyes using portable NIR spectroscopic techniques, *Food Addit. Contam. Part A* 36 (11) (2019) 1589–1596.
- [6] S. Lohumi, et al., A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration, *Trends Food Sci. Technol.* 46 (1) (2015) 85–98.
- [7] S.A. Haughey, et al., The use of handheld near-infrared reflectance spectroscopy (NIRS) for the proximate analysis of poultry feed and to detect melamine adulteration of soya bean meal, *Anal. Methods* 7 (1) (2015) 181–186.
- [8] K.N. Basri, et al., Classification and quantification of palm oil adulteration via portable NIR spectroscopy, *Spectrochim. Acta A. Mol. Biomol. Spectrosc.* 173 (2017) 335–342.
- [9] B. Li, et al., Rapid detection of authenticity and adulteration of walnut oil by FTIR and fluorescence spectroscopy: a comparative study, *Food Chem.* 181 (2015) 25–30.
- [10] K.C. Kaufmann, et al., Portable NIR spectrometer for prediction of palm oil acidity, *J. Food Sci.* 84 (3) (2019) 406–411.
- [11] S.N.N.A. Rani, Prediction of Free Fatty Acid in Crude Palm Oil Using Near Infrared Spectroscopy, Universiti Teknologi Malaysia, 2015.
- [12] S. Kasemsumran, et al., A feasibility study on non-destructive determination of oil content in palm fruits by visible–near infrared spectroscopy, *J. Near Infrared Spectrosc.* 20 (6) (2012) 687–694.
- [13] E. Teye, X. Huang, Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression, *Food Anal. Methods* 8 (4) (2015) 945–953.
- [14] Q. Chen, et al., Comparisons of different regressions tools in measurement of antioxidant activity in green tea using near infrared spectroscopy, *J. Pharm. Biomed. Anal.* 60 (2012) 92–97.
- [15] H. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, *Trends Food Sci. Technol.* 18 (2) (2007) 72–83.
- [16] Q.-Q. Wang, H.-Y. Huang, Y.-Z. Wang, Geographical authentication of *Macropygia cocos* by a data fusion method combining ultra-fast liquid chromatography and Fourier transform infrared spectroscopy, *Molecules* 24 (7) (2019) 1320.
- [17] H. Dubey, Efficient and accurate kNN based classification and regression, A Master Thesis Presented to the Center for Data Engineering, International Institute of Information Technology 500 (2013) 32.
- [18] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, Supervised pattern recognition in food analysis, *J. Chromatogr. A* 1158 (1–2) (2007) 196–214.
- [19] J. Sun, A multivariate principal component regression analysis of NIR data, *J. Chemom.* 10 (1) (1996) 1–9.
- [20] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, *Anal. Chim. Acta* 185 (1986) 1–17.
- [21] A. McIntosh, et al., Spatial pattern analysis of functional brain images using partial least squares, *Neuroimage* 3 (3) (1996) 143–157.
- [22] Y. Shao, Y. He, Nondestructive measurement of the acidity of strawberry based on wavelet transform and partial least squares, 2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007, IEEE, 2007.
- [23] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2) (2001) 109–130.
- [24] A. Krishnan, et al., Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review, *Neuroimage* 56 (2) (2011) 455–475.
- [25] H.A. Rahim, S.N.N.A. Rani, Classification of fatty fat acid in palm oil using near infrared spectroscopy, *Sens. Transducers* 156 (9) (2013) 247.
- [26] Y. Ozaki, W.F. McClure, A.A. Christy, Near-Infrared Spectroscopy in Food Science and Technology, John Wiley & Sons, 2006.
- [27] A.S. Luna, et al., Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy, *Spectrochim. Acta A. Mol. Biomol. Spectrosc.* 100 (2013) 115–119.
- [28] E. Teye, et al., Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans, *Food Res. Int.* 55 (2014) 288–293.
- [29] S. Sciuto, et al., Rapid screening technique to Identify Sudan Dyes (I to IV) in adulterated tomato sauce, chilli powder, and palm oil by innovative high-resolution mass spectrometry, *J. Food Prot.* 80 (4) (2017) 640–644.
- [30] M. Arslan, et al., Rapid Screening of Phenolic Compounds from Wild *Lycium ruthenicum* Murr. Using Portable near-Infrared (NIR) Spectroscopy Coupled Multivariate Analysis, *Anal. Lett.* (2020) 1–15.
- [31] Z. Li, et al., Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for *Aspergillus flavus* contamination detection in peanut, *J. Sci. Food Agric.* 99 (13) (2019) 5671–5679.