Wageningen Academic Publishers

Evaluation of extra-virgin olive oil adulteration using FTIR spectroscopy combined with multivariate algorithms

Y. Xu, M.M. Hassan, F.Y.H. Kutsanedzie, H.H. Li and Q.S. Chen*

School of Food & Biological Engineering, Jiangsu University, Zhenjiang 212013, China P.R.; qschen@ujs.edu.cn

Received: 21 May 2018 / Accepted: 13 July 2018

RESEARCH ARTICLE

Abstract

Fourier transfer infrared (FTIR) spectroscopy is a fast and reliable technique for the authentication of adulteration in the extra-virgin olive oil (EVOO) for quality control and market management. To verify the authenticity of EVOO, the feasibility of FTIR spectroscopy coupled with multivariate calibration has been investigated. Three different multivariate calibrations including linear discrimination analysis, back propagation artificial neural network and least-squares support vector machine (LS-SVM) were studied to make models with the acquired spectra and compared by correlation coefficients in the prediction set. The results demonstrated that the LS-SVM model was superior to others based on its optimum discrimination rates of 100 and 92.5% in the training and prediction set respectively. Besides, all misclassified samples were low-level adulterated EVOO ones with concentrations of 2.5%. This work demonstrates that the FTIR spectroscopy technique combined with an appropriately selected multivariate calibration could be promising to detect of adulterations in EVOO.

Keywords: adulteration, oil quality, Fourier transfer infrared spectroscopy, least-squares support vector machine

1. Introduction

Extra-virgin olive oil (EVOO), which is obtained from the olive, a crop of the olive tree (Oleaeuropaea; family Oleaceae) through the mechanical procedure, is a popular commodity that is extensively consumed throughout the world. It is widely used in cooking, soaps, cosmetics and pharmaceuticals products due to its flavour and bioactive properties (Callao and Ruisánchez, 2018). Increasing researches have demonstrated that large amounts of antioxidants and monounsaturated fatty acids have the potential to reduce DNA-aimed oxidative damages and the risk of coronary heart disease (Vann *et al.*, 2015; Yang *et al.*, 2017a). A relationship between the cancer prevention and olive oil-involved diet are also reported by preliminary studies, and that of colon cancer partially confirmed (Apetrei and Apetrei, 2014; De *et al.*, 2011). Moreover, multi-factors including cultivar, environment and cultural practices affect the uniqueness and quality of specific EVOOs, which makes them a valuable commodity to command a high price (Apetrei and Apetrei, 2014). Owing to the stated reasons, EVOO is vulnerable to be adulterated with other lower-valued edible oils. Thus, efficient and reliable detection tools to authenticate the adulterations in extra-virgin olive oil are of great significance.

Current approaches for adulterated EVOO detection include high-performance liquid chromatography (HPLC) (De *et al.*, 2011), gas chromatography/mass spectrometry (GC/MS) (Yang *et al.*, 2013), nuclear magnetic resonance spectroscopy (Xu *et al.*, 2014) and inductively coupled plasma optical emission spectroscopy (ICP-OES) (Gonzálvez *et al.*, 2010). Nevertheless, all the above mentioned classic methods have the demerits of being time-consuming, cost-intensive and demand high technical know-how; which outweighs the merits of high precision and selectivity. Alternatively, laser-induced fluorescence spectroscopy (Mu *et al.*, 2015), surface-enhanced Raman spectroscopy (El-Abassy *et al.*, 2010), ion mobility spectrometry (Garrido-Delgado *et al.*, 2018) and biosensor methods including optical thin-film biosensor chips and biosensor-based multi-sensitive platform (Oussama *et al.*, 2012; Vasconcelos *et al.*, 2015) emerged with the superiority of simple instrumentations and easy operations, but they demand sample pretreatments

and preparation which make them as well as lack on-line or in-line applications.

Recently, Fourier transfer infrared (FTIR) spectroscopy, as an advanced and excellent analytical technique, has gained considerable interests for its various applications in the food industry. Compared to other available methods, the technique reveals several advantages such as simplicity, rapidity, environmentally and requires minimal sample preprocessing. So far, the infrared spectroscopy technique, combined with multivariate calibration, has been used for the detection of olive oil and other food products (Su and Sun, 2018; Vasconcelos *et al.*, 2015). The cited papers further reported that FTIR spectroscopy has a high potential for the detection of adulterated olive oil combined with multivariate calibrations. However, the majority of the multivariate calibrations that have been used in those papers included partial least square discriminant analysis, principal component analysis (PCA), and cluster analysis and so on (Fadzlillah *et al.*, 2014; Georgouli *et al.*, 2016; Jiménez-Carvelo *et al.*, 2017). To date, there are few studies reported attempts on the use of nonlinear multivariate calibrations combined with FTIR for the detection of adulteration in EVOO (Lohumi *et al.*, 2015; Yang *et al.*, 2017b). Considering the subtle differences between EVOO spectra and low-concentration adulterated ones, the linear tools alone may not provide a complete solution to the classification problem; therefore, it is of great significance to attempt different linear and nonlinear tools for the detection of adulteration in EVOO.

This paper attempts to verify the authenticity of EVOO using FTIR spectroscopy technique coupled with multivariate analysis. Herein, linear discriminant analysis (LDA), back propagation artificial neural network (BP-ANN) and least-squares support vector machine (LS-SVM) were used comparatively for the selection of an optimal model to provide a promising method for the detection of adulteration in EVOO.

2. Materials and methods

Samples preparation

In this work, all the edible oils-EVOO, pure peanut oil (PPO) and pure rapeseed oil (PRO) were obtained from the local Auchan supermarket; and were deemed as pure samples. Herein, pure EVOO samples were obtained without further treatments; they were from the same brand (NUMA) and batch and mixed before scanning to eliminate the effect of non-uniformity. In this study, the purchased EVOO was defined as pure EVOO samples, and the other purchased edible oils (PPO and PRO) were used as pure/adulterated ones. Three kinds of adulterated samples were prepared via the blending of various concentration of purchased oils: PPO adulterations were made with the mixing of EVOO and varying concentration of PPO, and other two were blended with PRO and a mixture of PPO and PRO (w/w 1:1), respectively. Table 1 represents the batches, types, number of samples, and adulterated concentration of all adulterated samples. These samples were obtained by the random incremental addition of each adulterant as man-made in the range of adulteration provided. Finally, a total of 160 samples including 120 adulterated samples and 40 pure samples were used in this work.

Fatty acid analysis

0.9 ml hexane and 0.2 ml of 1 M sodium methoxide were added to 0.05 g oil sample and the resultant mixture shaken for 2 min. Thereafter, the clear supernatant was taken for further GC-MS analysis. The GC parameters were set to: column, Rtx-5 MS (30 m × 0.25 mm, 0.25 μm; Restex Corp., Bellefonte, PA, USA); the pressure was 84.6 kPa; the total flow rate was 24 ml/min; column flow was 1 ml/min; line speed was 37.5 cm/s; and carrier gas was of high purity helium used. The programmed temperature was held for 2 min at 130 °C, then increased to 200 °C (10 °C/min), and finally held at 270 °C for 4 min (2 °C/min); split ratio was set to 50:1. The mass spectrometer was operated in electron impact ionisation (70 eV) and full scan (40-800 m/z) mode using an injection volume of 1 μl.

Table 1. Batches, types, number of samples, and adulterated concentration of adulterated samples.

Batches	Adulterated other oils	Number of samples	Adulterated concentrations
1st	pure peanut oil	30	3.72~50.50%
	pure rapeseed oil	30	2.02~51.77%
	pure peanut oil/pure rapeseed oil	30	1.95~41.38%
2nd	pure peanut oil	10	3.88~44.78%
	pure rapeseed oil	10	2.03~42.36%
	pure peanut oil/pure rapeseed oil	10	1.98~38.93%

Fourier transfer infrared spectra collection

Spectra were acquired by an FTIR spectrometer (Excalibur 3100, Varian Inc. Palo Alto, CA, USA) equipped with a liquid nitrogen cooled mercury cadmium telluride detector. The sample holder was coupled with an overhead attenuated total reflection (ATR) accessory with germanium crystal (PIKE Technologies, Madison, WI, USA). Each FTIR spectrum was collected in the mid-IR region from 4,000-700/cm at a resolution of 8/cm after averaging 32 scans. Multi-reflection spectra of the samples were obtained and presented in absorbance units after taking into account the background spectrum of air. The ATR crystal was deeply cleaned with ethanol and distilled water and was dried between successive measurements to prevent any cross-contamination.

Multivariate calibrations

In this work, three different multivariate algorithms, including LDA, BP-ANN and LS-SVM were systemically studied for modelling and compared their performance. Before modelling, all the samples were divided into two subsets, namely the training set and the prediction set. Samples in the training set were used to build the model and ones in the prediction set were utilised to evaluate the performance of the final model. Specifically, for the enhancement of the high robustness and the generalisation of developed models, two batches of adulterations were artificially made as follows: the adulteration concentrations of the 1st batch samples covered that of the 2nd batch. Accordingly, 120 samples with 30 pure samples and 90 adulterated samples of the 1st batch were contained in the training set; and 40 samples with 10 pure ones and 30 adulterated samples of the 2nd batch were contained in the prediction set.

Software

All algorithms were run in Matlab R2009a (Mathworks Inc., Natick, MY, USA) in Windows 7. Varian Resolutions Pro 4.0 Software (Varian Inc., Palo Alto, CA, USA) was used for spectrum data acquisition.

3. Results

Spectra investigation and characteristic variables extraction

Infrared spectroscopy plays a vital role in molecular structures identification due to the multi-information recorded and the possibility of assigning certain absorption peaks to functional groups. In the analysis of edible oils, most peaks and troughs of the spectrum were associated with certain functional groups. Figure 1 shows the raw spectra (Figure 1A) and 2nd-derivative (Figure 1B) spectra

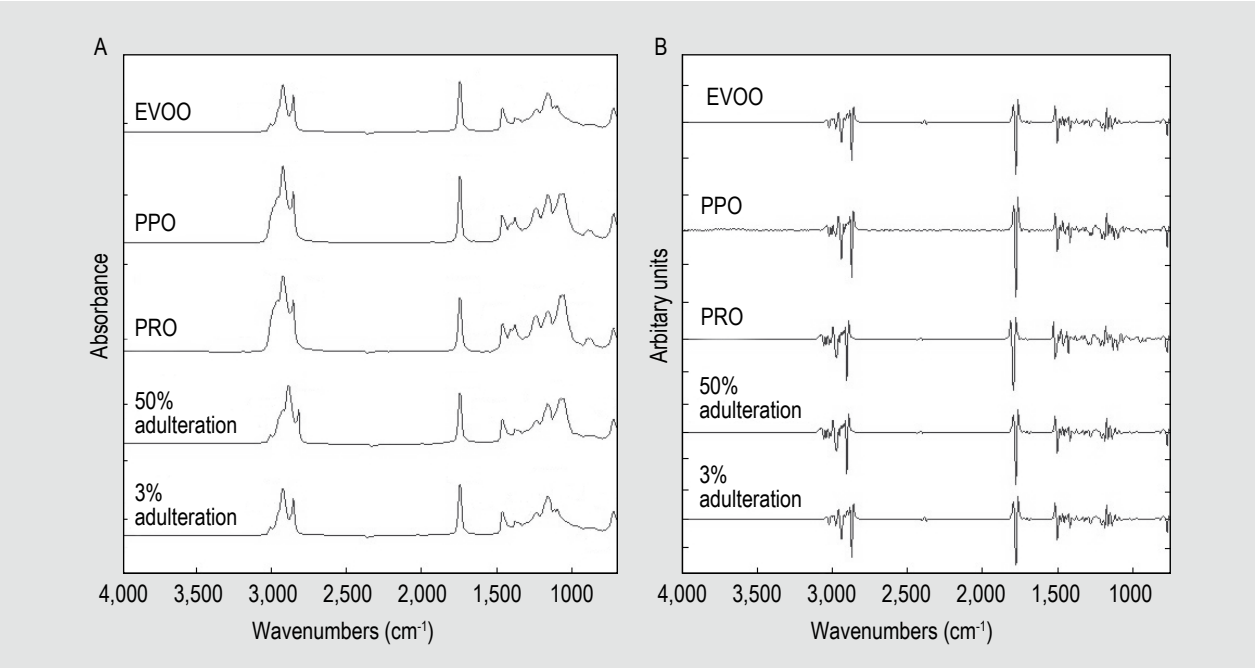

Figure 1. Fourier transfer infrared spectra data collected from extra-virgin olive oil (EVOO), pure peanut oil (PPO), pure rapeseed oil (PRO) and two adulterated oil samples (A) and their corresponding 2nd derivative spectra (B). The two adulterated samples were made from PPO, PRO, EVOO in percentage weight proportions of 25, 25, 50; and 1.5, 1.5, 97 respectively. These spectra showed the differences between the pure oils and those adulterated. The 2nd derivative spectra were obtained from spectra processing to remove slope variation and to correct the scatter effects which may influence the performance of a model.

from EVOO, PPO, PRO and adulterated samples. For the former three pure oils, it can be seen that even though they have the similar waveforms in both raw and 2nd derivative spectra, there were still obvious differences that can be observed in specific bands between them. Such appearances can be attributed to the characteristic absorption bands for common triglycerides and the similar chemical composition of EVOO, PPO and PRO, in terms of fatty acid compositions (Table 2); while the differences among the three spectra can be ascribed to each oil type unique characteristic revealed at some wave numbers, shown in Table 3 (LermaGarcía *et al.*, 2010; Rohman and Man, 2010). Observing from Table 3, differences of wave number bands were mainly focused on the range of 721-1,743/cm and 2,852-3,005/cm, which were assigned to the bending vibration caused by C-H, the stretching of C=O and the symmetric stretching vibration of the bond in this function group =C-O-C that absorbs the energy of infrared light to form the FTIR absorption peak (Rohman and Man, 2010; Zhao *et al.*, 2015). The spectrum of adulterated samples also confirmed the existence of differences in varying adulteration concentrations (Figure 1). Moreover, observing from the 50% and 3% adulterated EVOO samples in Figure 1, the differences between 50% adulterated EVOO and 3% adulterated EVOO became small, and with the decreased concentrations of the adulterants, the spectra of low concentration adulteration were much more like that of EVOO. These revealed that FTIR spectroscopy technique can be used as a powerful tool for detection of the adulteration when combined with chemometric algorithms. Meanwhile, the spectra of those specific wave numbers were selected as the characteristic variables of pattern recognition for the present work.

Principal component analysis

Principal component analysis (PCA) is a kind of unsupervised recognition approaches that is widely used to extract information from a multivariate dataset via data

Table 2. The main component of fatty acids in extra-virgin olive oil, pure peanut oil, and pure rapeseed oil used in this study.

Fatty acid (%, weight/weight)[1]	Extra-virgin olive oil	Pure peanut oil	Pure rapeseed oil
Myristic (C14:0)	0.01±0.00	0.01±0.00	0.05±0.01
Palmitic (C16:0)	10.22±0.25	12.2±0.20	5.33±0.32
Stearic (C18:0)	3.03±0.14	3.74±0.13	2.80±0.08
Oleic (C18:1)	73.98±1.53	42.7±1.12	55.99±0.92
Linoleic (C18:2)	7.08±0.12	34.9±0.49	20.91±0.35
Linolenic (C18:3)	0.32±0.013	0.06±0.002	7.23±0.34
Arachidic(C20:0)	0.51±0.02	1.55±0.09	0.69±0.04

[1] Each value in the table represents the means ± the standard deviation (n=3).

Table 3. Functional groups and modes of vibration in the spectrum of edible plant oil.

Variable (X)	Wavenumber (cm^{-1})	Functional group assignment
X1	721	overlapping of the methylene (-CH_2) rocking vibration and to the out of plane vibration of *cis*-disubstituted olefins
X2	850	=CH_2 wagging
X3	962	bending vibration of CH functional groups of isolated *trans*-olefin
X4	1,030	C-O stretching
X5	1,098	stretching vibration of the CO ester group
X6	1,117	
X7	1,160	C-O stretching
X8	1,236	
X9	1,377	bending vibrations of CH_2 groups
X10	1,402	=C-H bending vibration
X11	1,417	rocking vibrations of CH bonds of *cis*-disubstituted olefins
X12	1,465	bending vibrations of the CH_2 and CH_3 aliphatic groups
X13	1,743	Ester carbonyl functional group of the triglycerides
X14	2,852	asymmetrical and symmetrical stretching vibration of methylene (-CH_2) group
X15	2,924	
X16	3,005	*cis* double-bond stretching

reconstruction and dimensional reduction. In this work, the array of 16 characteristic variables was extracted from the spectra data; however, they were cross-sensitive among some functional groups such as X4, X7 and X8 (Table 3). Thus, there were some collinear variables (i.e. overlapped information) among the 16 characteristic variables, and the overlapped information might as well introduce serious difficulty in the study. To avoid serious issues caused by overlapped spectra information, PCA was employed to extract principal components (PCs) vectors via eliminating redundant information from useful ones. Moreover, a visualising dimensional space was provided by PCA, which indicated the data cluster trends. Besides, the appearance of the different groups in cluster trend along with the principal component axes further explains the existence of diverse clusters associated with their respective categories (Xu *et al.*, 2017). Herein, a scatter plot was generated according to PC1, PC2 and PC3 issued from PCA. Figure 2 demonstrates two 2-dimensional (2D) scatter plots represented by PC1-PC2 (Figure 2A) and PC1-PC3 (Figure 2B), respectively. Samples in the training set were remarked as 'o' for EVOO, '*' for PPO adulteration, '□' for PRO adulteration, and '☆' for PPO and PRO adulteration (Xu *et al.*, 2015). Results showed that the former three PCs could account for 55.67, 29.99 and 8.98% of variances, respectively. Thus, the 3D space issuing from PC1, PC2 and PC3 can explain 94.64% of the information from the 16 characteristic variables.

Furthermore, EVOO samples (in red circles) were able to be differentiated from the other adulterated samples as seen in Figure 2. Such classification in 2D scatters plots could be explained by the theory of infrared spectroscopy and PCA algorithm. EVOO is often different from PPO and PRO in some ingredients based on their unique functional groups (Table 3); meanwhile, certain groups possess their own absorption bands in spectroscopy. Based on the PCs loadings plot shown in Figure 3, peaks at 850/cm, 1,030/cm, 1,098/cm and 1,117/cm were observed for the largest loadings for PC1. Among them, peaks at 850/cm and 1,030/cm are linked to $=CH_2$ wagging and C-O stretching respectively; whereas 1,098/cm and 1,117/cm corresponded to the stretching vibration of the CO ester group. Peaks for the largest loadings for PC2 were observed at 962/cm and 1,160/cm, where 962/cm is associated with the bending vibration of CH functional groups of isolated trans-olefin; and 1,160/cm is assigned to C-O stretching. PC3 loading peak at 721/cm is associated with the overlapping of methylene ($-CH_2$) rocking vibration and to the out of plane vibration of cis-disubstituted olefins. Thus, it is reasonable to expect some differences between the EVOO samples and other samples adulterated with PPO or PRO exist in the spectroscopy. Although with decreased concentration of the adulterants, some spectral differences may not be easily detected by the naked eye, PCA as a data reconstruction and dimensional reduction approach is capable of indicating the cluster trend between EVOO samples and adulterated samples in visualising dimensional space. As for these

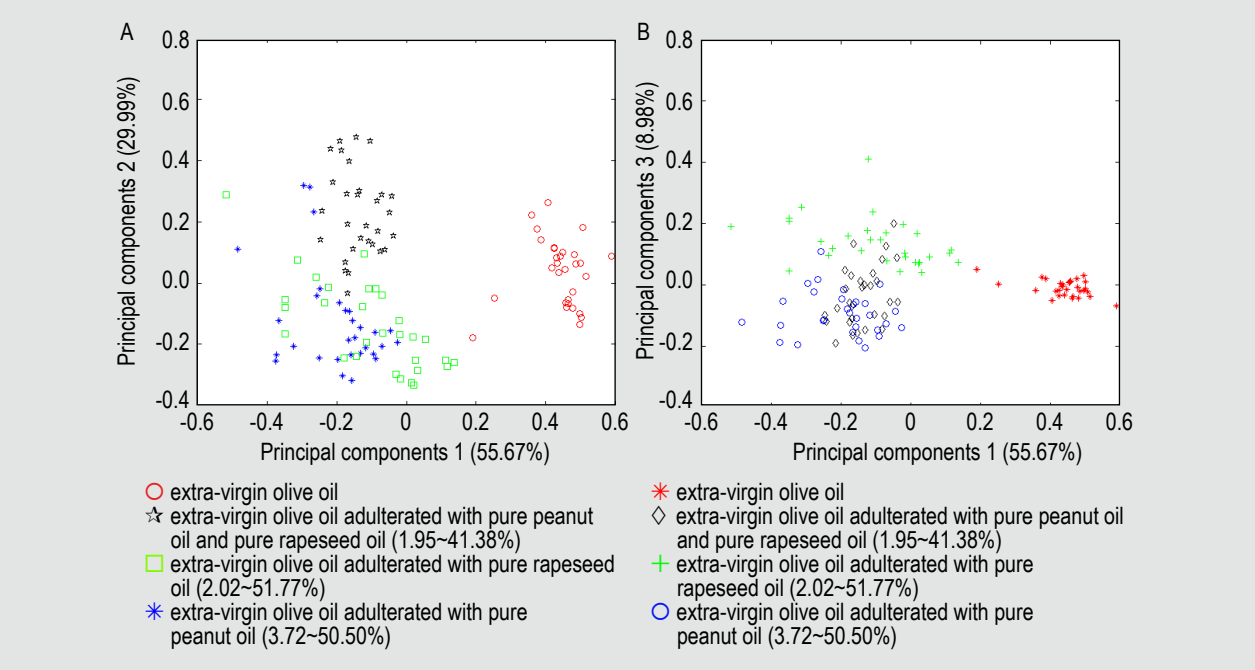

Figure 2. Score cluster plot for all training samples with principal components 1-principal components 2 (A) and principal components 1-principal components 3 (B). The scatter plot based on principal component analysis was used to extract information from a multivariate to achieve dimension reduction via the creation of new variables referred to as principal components; with principal components 1, 2 and 3 accounting for 55.67, 29.99 and 8.98% of original obtained information, respectively.

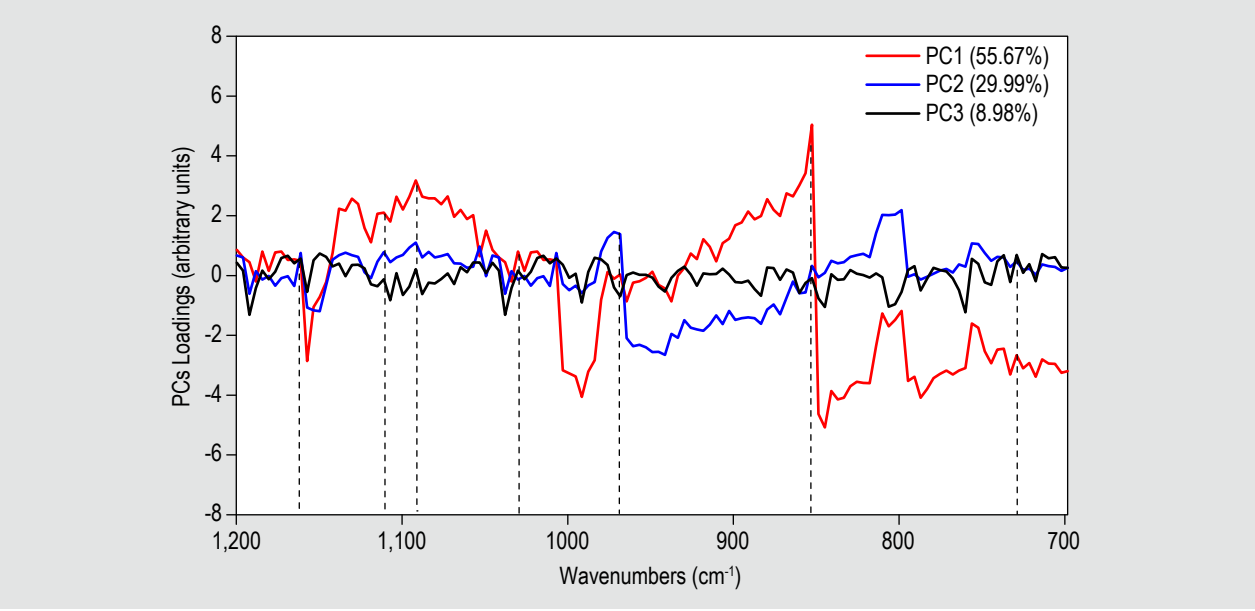

Figure 3. Plot of top three principal components (PCs) loadings against wavenumbers of oils spectra with Fourier transfer infrared system.

overlapped samples, it can be inferred that the presence of the adulterants is too low to be detected by spectroscopy and therefore other advanced chemometrics should be attempted.

Discrimination results

Supervised pattern recognition refers to the setting of a priori knowledge concerning the membership of different samples. It works as follow: firstly, a classification model is developed with the training samples in categories; afterwards, independent samples are used to evaluate the performance of the model. Eventually, the results are computed via the comparison of their predicted categories with their respective genuine one (Fyh *et al.*, 2017). Before modelling, PCA used as a variable extracting tool of the 16 characteristic values from each sample to allow the principal component factors to be selected for the input of the models. Optimised PCs were obtained according to the cross-validation R_P. Subsequently, different algorithms (i.e. LDA, BP-ANN, and LS-SVM) were attempted comparatively, for achieving a most appropriate classification model.

Optimal results of linear discriminant analysis model

PCs were extracted by PCA to develop the LDA discrimination model, and the optimal number of PCs was achieved according to the highest discrimination rates by R_P. Finally, the performances of the built LDA model were re-evaluated by some independent samples in the prediction set. Figure 4A revealed that the results of LDA models with different PCs; the optimal LDA model was obtained with the fourth PCs, and the discrimination results of the LDA model were 99.17% and 85% in the training and prediction set respectively.

Optimal results of back propagation artificial neural network model

In the application of BP-ANN, PCs extracted by PCA were selected as the input of BP-ANN and then optimised by R_P. Similarly, other parameters were also optimised as follows: 3 nodes were set in the hidden layer; 'tansig' function was set as the scale function and the learning rate factor and the coefficient of momentum were set as default values. Figure 4B revealed the results of BP-ANN model with varying PCs. Ultimately, the BP-ANN model was achieved with PCs=4 for the input layer and 8 nodes in the hidden layer. Thus, the optimal BP-ANN model was obtained with a 4-8-1 topological architecture, and the discrimination rate was 100 and 82.5% in the training and prediction set respectively.

Optimal results of least-squares support vector machine model

The kernel function plays a vital role in SVM operation. Generally, SVM has three classical kernel functions including polynomial kernel function (Equation 1), radial basis function (RBF) kernel function (Equation 2) and sigmoid kernel function (Equation 3) (Jiang *et al.*, 2012).

Polynomial kernel function:

$$K(x_i,x_j)=(1+x_ix_j)^{\sigma} \quad (1)$$

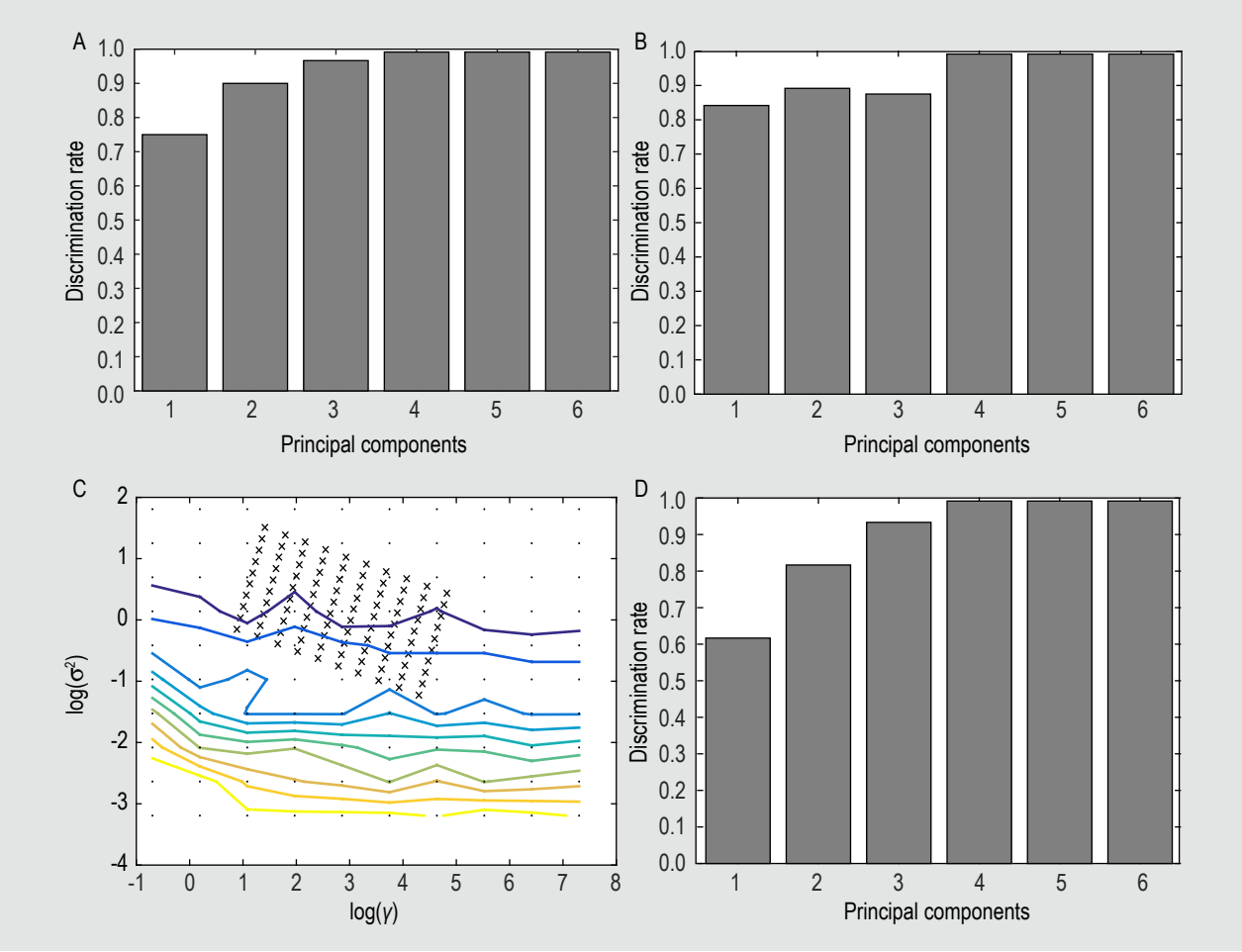

Figure 4. Discrimination rates of linear discriminant analysis model (A) and back propagation artificial neural network model (B) and least-squares support vector machine (D) model according to different principal components. Discrimination rates indicate the accuracy of linear discriminant analysis, back propagation artificial neural network and least-squares support vector machine models performance in the detecting adulterations in the test samples. Contour plot of the optimisation of the parameters by cross-validation (C). g is the regularisation parameter and σ^2 is the kernel parameter (σ^2). The method of 'grid search' was used to obtain the best two parameters (g and σ^2); and the contour map with lines of different colours means the different models obtained with the corresponding pairs of parameters (g and σ^2) the outcome of support vector machine optimisation from which two corresponding pairs on a line. γ=regularisation parameter.

RBF kernel function:

$$K(x_i,x_j)=\exp(-\frac{||x_i-x_j||^2}{2\sigma^2}) \quad (2)$$

Sigmoid kernel function:

$$K(x_i,x_j)=\tanh(\sigma(x_i x_j)+\upsilon) \quad (3)$$

In particular, RBF kernel function, which computes simpler and faster when compared to others, is attempted in this work. It is also deemed as the best choice when a prior experience is lacking. Moreover, there are two important parameters, namely the regularisation parameter (γ) and the kernel parameter (σ^2), determines the performance of SVM model. Thus, the method of 'grid search' was used to obtain the best two parameters (γ, σ^2) in two steps (Ouyang *et al.*, 2014; Xu *et al.*, 2015): first, a grid length in 10×10 (marked in '·') was applied by a larger step. Then, another grid length in a 10×10 (marked in '×') was used by much smaller step to achieve the best combination of γ and σ^2. Herein, cost functions were defined as error estimations to optimise the LS-SVM model, and pairs of (γ, σ^2) were selected with the lowest cost function.

Figure 4C illustrated a contour map where varying cost values were represented in specific colours. Additionally, log σ^2 ranged from -4 to 2 and log γ ranged from -1 to 8 were attempted with an interval of 1. Thus, σ^2=7 and γ=10 values were optimised. Meanwhile, the optimal number of PCs was also determined based on the highest discrimination rate via R_P. Figure 4D showed the discrimination rates of the LS-SVM model according to different PCs. Therefore,

the optimal model was obtained with the fourth (4) PCs, and the back-discrimination rate of the final LS-SVM model yielded 100 and 92.5% in the training and prediction set respectively.

4. Discussion

As regard searching an accurate model for the detection of adulterated EVOO using FTIR spectroscopy, three classification algorithms including LDA, BP-ANN and LS-SVM, were studied systematically and comparatively. The overall results are summarised in Table 4.

Firstly, it was observed that all the models had 4 PCs. The reasons could be explained by the chemical structure and algorithm theory. From the chemical structure, it is reasonable that there are similarities among three pure oils; and however, after the purposive adulteration, the differences among oils were further weakened while the similarities enhanced. Accordingly, less PCs cannot represent the individual differences. From the algorithm theory, the PCs have significant effects on a discrimination model when it can explain more than 1% of the variances. However, in this PCA analysis, each PCs can only explain less than 1% of the variance except for the top 4 of PCs. Considering the inconspicuous differences between EVOO samples and adulterated sample spectra, the performance of model would be weakened if PCs=4 is removed in the model calibration. On the contrary, the performance of the model would not be strengthened, or even be weakened, if more PCs are added. Besides, the number of PCs was also confirmed by R_P in this study. Therefore, the achieved results demonstrated that the optimal number of PCs for all models (LDA, BP-ANN, and LS-SVM) as 4.

Secondly, remarkable differences could be observed among the performance of three models. In the LDA model, the discrimination rate was 99.17% in the training set, with one misclassified adulterated sample with the adulterated concentration of 1.95%; and 85% in the prediction set, with one, misclassified pure EVOO and five misclassified adulteration with adulterated concentrations of 2.03, 2.40, 2.18, 3.06, and 1.98% respectively. In the BP-ANN model, the back discrimination rate was 100% in the training set; and 82.5% in the prediction set, with one, misclassified pure EVOO and six misclassified adulteration with the adulterated concentrations of 3.88, 2.03, 2.40, 2.18, 3.06, and 1.98% respectively. In the LS-SVM model, the result in training set (100%) was the same as that of the BP-ANN model; and 92.5% in the prediction set, with

Table 4. Comparison of the identification results of linear discriminant analysis, back propagation artificial neural network, and least-squares support vector machine models.[1]

Subsets	Models	Samples types	Samples number	Discrimination results		
				EVOO	Adulteration	Discrimination rate (%)
Training set	LDA	EVOO	30	30	0	99.17 [2]
		adulteration	90	1	89	
	BP-ANN	EVOO	30	30	0	100.00
		adulteration	90	0	90	
	LS-SVM	EVOO	30	30	0	100.00
		adulteration	90	0	90	
Prediction set	LDA	EVOO	10	9	1	85.00 [3]
		adulteration	30	5	25	
	BP-ANN	EVOO	10	9	1	82.50 [4]
		adulteration	30	6	25	
	LS-SVM	EVOO	10	10	0	92.50 [5]
		adulteration	30	3	27	

[1] LDA: linear discriminant analysis; BP-ANN: back propagation artificial neural network; LS-SVM: least-squares support vector machine; EVOO: extra-virgin olive oil.

[2] linear discriminant analysis model in the training set: one adulteration sample was misclassified into, and its adulteration concentration is 1.95%.

[3] linear discriminant analysis model in the prediction set: one extra-virgin olive oil sample was misclassified; five adulteration samples were misclassified, and their adulteration concentration are 2.03, 2.40, 2.18, 3.06, and 1.98, respectively.

[4] back propagation artificial neural network model in the prediction set: one extra-virgin olive oil sample was misclassified; six adulteration samples were misclassified, and their adulteration concentration are 3.88, 2.03, 2.40, 2.18, 3.06, and 1.98%, respectively.

[5] least-squares support vector machine model in the prediction set: three adulteration samples were misclassified, and their adulteration concentration are 2.03, 2.18, and 1.98%, respectively.

three, misclassified adulterated samples with adulterated concentrations of 2.03, 2.18, and 1.98% respectively. Results were concluded as follows: first, adulterated samples were more inclined to be misclassified in this work; second, the concentration of the misclassified adulterant samples was lower than other adulterated samples. Such outcome basically was attributed to the similar components between PPO, PRO and EVOO, which had similar spectra among different samples, and further resulted in misclassification between EVOO samples and adulterated ones. Thirdly, two pure EVOOs were misclassified in the prediction set of LDA and BP-ANN model. The outliers in this study are interpreted as the results of the unbalanced number of pure EVOOs and the adulteration. Generally, the development of a classification model was depended on an adequate number of adulteration and the corresponding real ones. Whereas, considering the practical feature of application conditions and the comprehensive verify of the feasible of the proposed approach, 120 of adulterated EVOO samples and 40 of pure EVOO samples were attempted here. Accordingly, the minority samples (pure EVOO samples) failed to provide sufficient information to support the calibration in the discrimination model.

A comparison of discrimination rate among three models is seen in Table 4. Herein, LS-SVM achieved remarkable performance to more than both LDA model and BP-ANN model: In the training set, nonlinear tools (BP-ANN and LS-SVM) demonstrated slightly superior to the linear tool (PLS) with the classification rates of 100 and 99.17% respectively; whereas, in the prediction set, LS-SVM model made substantial progress in discrimination and gave the best results as compared to LDA model (85%) and BP-ANN model (82.5%). More importantly, there were no misclassified pure EVOO samples in LS-SVM model, which confirmed the capacity of solving issues of the unbalanced number of sample. The reasons of varying performance of three models are outlined as follows: first, non-linear algorithms are typically better than the linear ones at self-learning and self-adjustment, which caused the stronger classification abilities in BP-ANN and LS-SVM as compared to the LDA. Second, BP-ANN, as a commonly used nonlinear tool, works according to the experiential risk minimisation (ERM), which cannot give considerations to empirical risk and confidence interval at the same time. In other words, BP-ANN tends to suffer generalisation, leading to over-fitted models in some cases. Overfitting enabled a good discrimination result in the training set, whereas a bad one in the prediction set. As a result, it is reasonable that a big difference exists between the training set and prediction set in the BP-ANN model. Third, LS-SVM as another nonlinear algorithm used here operates based on the structural risk minimisation (SRM). SRM describes a general model of capacity control and provides a trade-off between hypothesis space complexity and the quality of fitting the training data (Wang and You, 2013). It allows one to trade-off errors in the training sample against improved generalisation performance; and then considers a more general case when the hierarchy of classes is chosen in response to the data. Finally, a result is presented on the generalisation performance of classifiers with a 'large margin'. More details about the impressive generalisation performance of the maximal margin hyperplane algorithm descriptions are found in (Vapnik and Chervonenkis, 2015). Table 5 shows the classification results of different techniques applied to oil adulteration detection. It is obvious that traditional approaches like HPLC, GC/MS and ICP-OES may achieve higher classification rate than that presented in this study, but these aforementioned conventional techniques have demerits such being time-consuming and cost-intensive. In addition, methods which are easy to operate like laser-induced fluorescence, Raman

Table 5. Comparison of different methods in the study of adulteration of oil products.[1]

Methods	Algorithms	Classification rate (%)	Reference
HPLC	PLS-DA	100	De *et al.*, 2011
GC/MS	PLS-LDA	90	Yang *et al.*, 2013
ICP-OES	DA	93.65	Gonzálvez *et al.*, 2010
Laser-induced fluorescence	Back-propagation neural network	92.2	Mu *et al.*, 2015
Raman spectroscopy	PLS	97.1	El-Abassy *et al.*, 2010
FTIR spectroscopy	LDA	88.3	Vasconcelos *et al.*, 2015
IMS	PLS	72.1	Garrido-Delgado *et al.*, 2018
FTIR spectroscopy	CLPP+kNN	74.65	Georgouli *et al.*, 2016
FTIR spectroscopy	LS-SVM	92.5	this study

[1] HPLC: high-performance liquid chromatography; PLS-DA: partial least square discriminant analysis; PLS-LDA: partial least square linear discriminant analysis; GC/MS: gas chromatography/mass spectrometry; ICP-OES: inductively coupled plasma optical emission spectroscopy; DA: discriminant analysis; PLS: partial least square; FTIR: Fourier transfer infrared; LDA: linear discriminant analysis; IMS: ion mobility spectrometry; CLPP: continuous locality preserving projections; kNN: k-nearest neighbors; LS-SVM: least-squares support vector machine.

spectroscopy and ion mobility spectrometry also need sample pretreatments. Compared with above methods, FTIR spectroscopy proves to be a simple technique for rapid detection of adulteration in EVOO. Thus, compared with ERM, SRM reveals its superior property of minimising upper bounds on the expected risk. In short, LS-SVM as an efficient and reliable recognition tool showed the best performance compared with the other two algorithms in solving the complicated data classification.

5. Conclusions

In this study, FTIR spectroscopy independently coupled to LDA, BP-ANN and LS-SVM was applied as a rapid and reliable technique for the rapid identification of adulteration in EVOO. Among the three algorithms, LS-SVM model achieved discrimination rate of 92.5% in the prediction set which proved superior to LDA (85%) and BP-ANN (82.5%) model. The overall results demonstrate that FTIR spectroscopy combined with LS-SVM can yield remarkable improvements in the rapid detection of adulterated EVOO. FTIR spectroscopy technique combined with appropriately selected algorithms holds promise as a powerful analytical to with high potential for monitoring oil quality and safety. This has great implications for safeguarding consumer health threats and producers profitability so far as adulteration of oil remains a grave concern.

Acknowledgements

This work has been financially supported by the National Natural Science Foundation of China (31772063), a key R&D Program of Jiangsu Province (BE2017357).

Conflict of interest

Authors have no conflict of interest.

References

Apetrei, I.M. and Apetrei, C., 2014. Detection of virgin olive oil adulteration using a voltammetric e-tongue. Computers & Electronics in Agriculture 108: 148-154.

Callao, M.P. and Ruisánchez, I., 2018. An overview of multivariate qualitative methods for food fraud detection. Food Control 86: 283-293.

De, L.M.-E.P., Bosque-Sendra, J.M., Bro, R. and Cuadros-Rodríguez, L., 2011. Discriminating olive and non-olive oils using HPLC-CAD and chemometrics. Analytical & Bioanalytical Chemistry 399: 2083-2092.

El-Abassy, R.M., Donfack, P. and Materny, A., 2010. Visible Raman spectroscopy for the discrimination of olive oils from different vegetable oils and the detection of adulteration. Journal of Raman Spectroscopy 40: 1284-1289.

Fadzlillah, N.A., Man, Y.B.C. and Rohman, A., 2014. FTIR spectroscopy combined with chemometric for analysis of sesame oil adulterated with corn oil. International Journal of Food Properties 17: 1275-1282.

Fyh, K., Chen, Q., Hassan, M.M., Yang, M., Sun, H. and Rahman, M.H., 2017. Near infrared system coupled chemometric algorithms for enumeration of total fungi count in cocoa beans neat solution. Food Chemistry 240: 231-238.

Garrido-Delgado, R., Muñoz-Pérez, M.E. and Arce, L., 2018. Detection of adulteration in extra virgin olive oils by using UV-IMS and chemometric analysis. Food Control 85: 292-299.

Georgouli, K., Rincon, J.M.D. and Koidis, A., 2016. Continuous statistical modelling for rapid detection of adulteration of extra virgin olive oil using mid infrared and Raman spectroscopic data. Food Chemistry 217: 735-742.

Gonzálvez, A., Armenta, S. and Guardia, M.D.L., 2010. Adulteration detection of argan oil by inductively coupled plasma optical emission spectrometry. Food Chemistry 121: 878-886.

Jiang, H., Liu, G., Xiao, X., Yu, S., Mei, C. and Ding, Y., 2012. Classification of Chinese soybean paste by Fourier Transform Near-Infrared (FT-NIR) spectroscopy and different supervised pattern recognition. Food Analytical Methods 5: 928-934.

Jiménez-Carvelo, A.M., Osorio, M.T., Koidis, A., González-Casado, A. and Cuadros-Rodríguez, L., 2017. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. LWT – Food Science and Technology 86: 174-184.

LermaGarcía, M.J., RamisRamos, G., HerreroMartínez, J.M. and SimóAlfonso, E.F., 2010. Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. Food Chemistry 118: 78-83.

Lohumi, S., Lee, S., Lee, H. and Cho, B.K., 2015. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. Trends in Food Science and Technology 46: 85-98.

Mu, T., Chen, S., Zhang, Y., Chen, H., Guo, P. and Meng, F., 2015. Classification of motor oil using laser-induced fluorescence and phosphorescence. Analytical Letters 49(8): 1233-1239.

Oussama, A., Elabadi, F., Platikanov, S., Kzaiber, F. and Tauler, R., 2012. Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores. Journal of the American Oil Chemists Society 89: 1807-1812.

Ouyang, Q., Zhao, J. and Chen, Q., 2014. Instrumental intelligent test of food sensory quality as mimic of human panel test combining multiple cross-perception sensors and data fusion. Analytica Chimica Acta 841: 68-76.

Rohman, A. and Man, Y.B.C., 2010. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. Food Research International 43: 886-892.

Su, W.H. and Sun, D.W., 2018. Fourier Transform Infrared and Raman and Hyperspectral Imaging techniques for quality determinations of powdery foods: a review. Comprehensive Reviews in Food Science & Food Safety 17(1): 104-112.

Vann, K.R., Sedgeman, C.A., Jacob, G., Avi, G.G. and Neil, O., 2015. Effects of olive metabolites on DNA cleavage mediated by human type II topoisomerases. Biochemistry 54: 4531-4541.

Vapnik, V.N. and Chervonenkis, A.Y., 2015. On the Uniform convergence of relative frequencies of events to their probabilities. Springer International Publishing, New York, NY, USA, pp. 264-280.

Vasconcelos, M., Coelho, L. and Barros, A., 2015. Study of adulteration of extra virgin olive oil with peanut oil using FTIR spectroscopy and chemometrics. Cogent Food and Agriculture 1: 1-13.

Wang, C.W. and You, W.H., 2013. Boosting-SVM: effective learning with reduced data dimension. Applied Intelligence 39: 465-474.

Xu, Y., Kutsanedzie, F.Y.H., Sun, H., Wang, M., Chen, Q., Guo, Z. and Wu, J., 2017. Rapid Pseudomonas species identification from chicken by integrating colorimetric sensors with near-infrared spectroscopy. Food Analytical Methods 11(4): 1199-1208.

Xu, Y., Li, H., Chen, Q., Zhao, J. and Ouyang, Q., 2015. Rapid detection of adulteration in extra-virgin olive oil using three-dimensional fluorescence spectra technology with selected multivariate calibrations. International Journal of Food Properties 18: 2085-2098.

Xu, Z., Morris, R.H., Bencsik, M. and Newton, M.I., 2014. Detection of virgin olive oil adulteration using low field unilateral NMR. Sensors (Basel) 14(2): 2028-2035.

Yang, M., Chen, Q., Kutsanedzie, F.Y.H., Yang, X., Guo, Z. and Ouyang, Q., 2017a. Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms. Measurement 103: 179-185.

Yang, R., Dong, G., Sun, X., Yang, Y., Liu, H., Du, Y., Jin, H. and Zhang, W., 2017b. Discrimination of sesame oil adulterated with corn oil using information fusion of synchronous and asynchronous two-dimensional near-mid infrared spectroscopy. European Journal of Lipid Science and Technology 119(9): 1600459.

Yang, Y., Ferro, M.D., Cavaco, I. and Liang, Y., 2013. Detection and identification of extra virgin olive oil adulteration by GC-MS combined with chemometrics. Journal of Agricultural & Food Chemistry 61: 3693-3702.

Zhao, X., Dong, D., Zheng, W., Jiao, L. and Lang, Y., 2015. Discrimination of adulterated sesame oil using mid-infrared spectroscopy and chemometrics. Food Analytical Methods 8: 2308-2314.