



Use of Infrared Spectroscopy and chemometrics for rapid authentication and detection of Moroccan traditional butter adulteration

W. Terouzi^{1,2*}, F. Kzaiber^{1,2}, A. Oussama¹

¹Team of Applied Spectro- Chemometrics and environment, Faculty of Science and Technology of Beni Mellal, University of Sultan Moulay slimane, 21000- Beni Mellal, Morocco.

²Laboratory of Engineering and Applied Technologies, Hight School of Technology of Beni Mellal, University of Sultan Moulay slimane, 21000- Beni Mellal, Morocco

*Corresponding author: w.terouzi@usms.ma

Abstract

Authenticity is the most important criterion for food quality. In fact, food industries, consumers and regulatory agencies are increasingly demanding fast and effective methods to confirm authenticity or detect adulterations. The goal of present study is to use Attenuated total reflectance-Fourier transform mid-infrared (ATR-FTMIR) spectroscopy coupled with chemometrics, as rapid and green analytical method for the detection of the traditional butter adulteration. In this case, pure traditional butter and blends of traditional cow's butter with different percentages of vegetable butter (3.8–40%) and of mashed potatoes (13–36%) were measured using ATR-FTMIR spectroscopy. The spectral data were subjected to a preliminary derivative elaboration based on the Gap algorithm to reduce the noise and extract a largest number of analytical information from spectra. Firstly, cluster analysis (CA) and principal component analysis (PCA) were applied, and three distinctive clusters were recognized. Then,

Linear discriminant analysis (LDA) and support vector machines (SVM), were elaborated as classification methods. The obtained classification results showed that this approach could identify and detect, easily, traditional butter adulteration with an accuracy value of 97,22 and 100%.

Keywords : Adulteration ; ATR-FTIR spectroscopy ; Authentication ; Chemometrics tools ; mashed potatoes ; traditional cow's butter ; vegetable butter

1. Introduction

Nowadays, the traditional butter, is one of the oldest milk products in many countries. It's a water in oil emulsion, with a report of 80 g milk fat/100 g minimum and 16 g moisture/100g maximum [1]. It's essentially made from cow's milk it is known for its multiple nutritional benefits and is the basis of different traditional recipes [2, 3]. Also, traditional cow's butter can be considered expensive compared to other vegetable butters. This can cause it to be adulterated by other low and cheap ingredients. Hence, the need to ensure its quality and authenticity [2, 4]. In fact, adulteration is the major issue for health of consumers. Since the beginning of food trade, incidents concerning adulterations of relevant products are well known [5].

On the other hand, there is a great global demand to develop fast, efficient, economical and green analytical methods. In this context, several analytical methods are developed based on Fourier transform infrared (FTIR) spectroscopy with attenuated total reflectance (ATR) or transmission cell accessories for authentication, characterization, identification or classification of fats and oils [6, 7, 8, 9,10, 11, 12]. Fourier Transform Mid Infrared (FTMIR) spectroscopy is a biochemical fingerprinting technique [13]. It can be, successfully, used to product results with the same accuracy and sensitivity as the reference methods in short time [14]. ATR-FTIR is a fast, cheap, green, simple to use technique with almost no sample pre-treatment and it shows a high selectivity when associated with chemometrics tools [15], [16]. Chemometrics is efficient tool, widely used in spectroscopy to extract useful information from complex spectra containing overlapping absorption peaks, interference effects and instrumental artifacts from the data collected [17].

Therefore, this work has given an other demonstration for ATR-FTIR combined with chemometrics as efficient approach to detect traditional cow's butter adulteration.

2. Materials and methods

2.1. Sample preparation

Firstly, a serie of 16 samples of pure and authentic traditional cow's butter (B) were personally taken from farm in Fkih Ben Salah area in Beni Mellal-Khnifra region, Morocco; and preserved at (- 4°C) until preparation of blends and analysis.

Then, the adulterated samples were prepared by mixing pure traditional cow's butter (B) with vegetable butter (BB) in the range of 3.8-40% and with mashed potatoes (BP) in the range of 13-36%. All vegetable butter and potatoes used as adulterants, were purchased in a local supermarket. So, we created three classes of samples with 104 samples in total : 68 samples (calibration set) were randomly taken for establishing classification models. Other 36 samples (external validation set) were used for testing the reliability and robustness of those models.

All studied samples were kept in cold storage (- 4°C) during the nights between the days of measurements. Spectroscopic measurements were taken from samples after they had been brought into equilibrium with the room temperature of 25°C.

2.2. Instrumentation and spectral acquisition

ATR-FTMIR analysis was carried out by using a Vector 22 Bruker FTIR Spectrophotometer equipped with an attenuated total reflectance accessory (ATR single-reflexion, Diamond, incident angle 45°, Pike Miracle, Pike Technologies, Madison, USA) with DTGS detector, Globar (MIR) Source and KBr Germanium separator. All spectra were scanned in the absorbance mode from 4000 to 600 cm^{-1} , with a resolution of 4 cm^{-1} at 80 scans ; and the data were handled with OPUS logiciel.

About 1g of each studied samples were directly placed, without preparation on an Attenuated Total Reflectance cell provided with a diamond crystal. Analyses were carried out at controlled room temperature (25°C). The background was collected before every sample was measured. Between spectra, the ATR plate was cleaned by ethanol.

2.3. Chemometric tools

In this work, to extract the useful information from the ATR-FTMIR analysis of studied butter samples, we used four popular chemometric methods: Principal component analysis (PCA); Cluster analysis (CA); Linear discriminate analysis (LDA) and support vector machines (SVM). These techniques show effective results in various applications [9, 18, 19, 20, 21, 22, 23].

Cluster analysis is a non-supervised technique. It is able to classify a given population into groups (clusters), based on similarity or closeness measures. The objective principle of the distance is adopted for this aim. The agglomerative hierarchical clustering is nowadays one of the most cited methods in literature [24], providing intuitive similarity relationships between each sample and the entire data set. In hierarchical clustering, each cluster is subdivided into smaller clusters, forming a tree-shaped data structure or dendrogram. Agglomerative hierarchical clustering starts with the single-gene clusters and successively joins the closet clusters until all genes have been joined into the supercluster: The sample grouping is illustrated by a dendrogram that permits a global vision of the similarity among the samples.

PCA is a non-supervised method. As a linear projection method is based on variance, transforms the original measurement variables into new uncorrelated variables called principal components (PCs) [25, 26]. The PCs are a simple linear combination of original variables. The scores vectors describe the relationship between the samples and allow checking if they are similar or dissimilar, typical or outlier.

LDA was used as supervised linear projection technique to find directions of maximum separation from a set of samples for which class membership is known in order to be able to predict the class membership of unknown samples [27, 28].

Support Vector Machines (SVMs) is a non-linear supervised learning technique [29]. It is kernel based learning algorithms introduced by Vapnik [30, 31]. SVM is a learning technique for classification and regression that uses linear or non-linear kernel-functions to project the data into a high-dimensional feature space. Correlation is then performed in this hyperspace based on the structural risk minimization principle; i.e., aiming to increase the generalization ability of a model [32, 33].

2.4 Data pre-processing procedures and Software

To extract better analytical information, a preliminary treatment of the data seemed necessary. So, a series of pre-processing elaborations were tested on the spectral data prior to the multivariate calibration. The Savitzky–Golay [34] and Norris gap [35] algorithms were tested

for data derivatisation. Standard normal variate (SNV) and multiple scatter correction (MSC) [36] were also tested. For data pre-treatment giving best result is derivative function through the Gap algorithm, by fixing the following parameters: 2nd order, gap size 21, with uncentered data in the range 4000-2400 and 2300-600 cm⁻¹.

All chemometrics techniques were performed by the UNSCRAMBLER software version 10.2 from CAMO (Computer Aided Modelling, Trondheim, Norway). And validated with full Cross validation, the same samples are used both for model estimation and testing: leaves out only one sample at a time [37].

3. Results and discussion

Firstly, Fourier transform mid infrared (FT-MIR) spectra of calibration and validation sets were obtained. One spectrum is the average of 80 scans of the studied butter samples on FT-MIR region. Then, the FT-MIR spectra of 68 samples from each of the three studied classes, were recorded in triplicate and a mean spectrum was calculated for each sample studied. The resultant mean spectra of three classes are presented in Figure.1.

According to Figure.1.it was not a big difference between the three classes by visual inspection. This is due to the similar chemical composition of traditionnal butter and adulterants (mashed potatoes and vegetabl butter) [2, 4]. In this case, chemometrics methods appeared very ideal to give an effective solution, as they allow extracting of unspecific analytical information from the full-spectra or large regions of them.

3.1. Cluster analysis

The calibration data set obtained from derivative transformation of the FTIR data, after removing the range 2400 - 2300 cm⁻¹(is due to CO₂) was employed to perform CA, applying Ward's method. Results were reported in the form of dendrogram, shown in Figure.2. On the basis of the connecting distances three distinctive clusters were defined.

3.2. PCA modeling

CA results were confirmed by PCA elaboration of the data. The PCA model was built with the NIPALS algorithm, by considering all frequency intervals 4000–2400 and 2300–600 cm⁻¹, as X variables, and the model was validated by full Cross validation.

PCA results are shown in Figure.3. Scores plot was constructed to observe principal groupings among observations. There are three clusters. Loading plot (Figure.3. (a) ;(b) and

(c)) indicates the importance of each variable for the model and it is used to interpret the relations among variables and clusters observed in the score plot. Spatially, the plotted PC1 loadings values of the spectral data are shown, representing the regions of the spectra where the differences among studied samples are more evident. It's very noted that proteins ($1680\text{--}1540\text{ cm}^{-1}$) and lipids ($3000\text{--}1700\text{ cm}^{-1}$) contents allowed to better identify and discriminate the butters samples according to their biochemical discrepancies.

Figure.4. is the Hotelling T^2 statistics plot. It is an alternative to plotting sample leverages. The plot displays the Hotelling T^2 statistic for each sample as a line plot. The associated critical limit (with a default p-value of 5%) is displayed as a red line. The Hotelling T^2 statistic has a linear relationship to the leverage for a given sample. Its critical limit is based on an F-test. In fact, it's useful to identify outliers or detect situations where a process is operating outside normal conditions. According to Figure.4. there's no « outliers » in the data set.

3.3. PCA-LDA modeling

LDA is subject to the constraint that the number of features should not exceed the number of samples. So, our high dimensional spectral data were, firstly, reduced to PCs prior to LDA by using PCA. The original data set comprising of a large number of inter-correlated variables, was transformed into a reduced new set of variables [38, 39].

According to score plot in Figure.5. The obtained PCA-LDA model was able to discriminate calibration set with a correct classification above 98%. This result is confirmed by the Table.1. where we see clearly, that there is only one misclassified sample (BB6).

3.3.1. Classification of new samples by PCA-LDA model

In order to verify the applicability and how reliable this model in detecting every adulterated butter or assuring its authenticity, the external validation process was carried out. Established PCA-LDA model is used to predict the quality of new samples. These latter have the same matrix effects as samples of calibration set. So, the PCA-LDA model was applied to a group of unknown samples of three classes, the results are listed in Table.2. The rate of correct classification was 97,22% within the test set. In particular, thirty-seven samples were safely assigned in the three classes, while one sample SBB1 resulted classified in another class.

3.4. SVM modeling

The SVMs model was built by considering, as X variables, the spectra in the range 4000-2400 and 2300-600 cm^{-1} ; and the classification model was validated by Cross validation with segment = 38.

SVM model was constructed with a nu-SVC classification, the optimal parameter for “nu” which lies in the range 0-1, is selected as the value that give the maximum correct classification rate, $\text{nu} = 0.5$.

The SVMs result is shown in the confusion matrix (Table.3), which indicates how many samples were classified in each class, and the prediction matrix, which indicates the classification determined for each sample in the training set. According to Table.3. all the samples are perfectly classified, except one sample (BB6).

Then, this result is confirmed by classification plot (Figure.6). Based on the characteristic wavelengths of lipid and protein, detected as significant variables responsible for the distinction between the three classes in loadings plot of PCA, all the classes resulted perfectly separated from the other ones, except for a single sample (BB6).

3.4.1.Classification of new samples by SVM model

In this step, the final classification model was applied on the prediction (external validation) set, to classify each sample into the classes previously established (Table.4).

Table.4 illustrate the classification results with the comparison between the predicted results of each class and the theoretical reference classes. The rate of correct classification was 100%.

In the light of the obtained results, it has been proved that the proposed method allow the correct classification and detection of traditional cow's butter adulteration. The both PCA-LDA and SVMs models carried out from transformed infrared spectra gave a good ranking rate of 97,22 and 100% respectively.

Finally, we arrived to develop a new application of the ATR-FTMIR combined with classification techniques as a rapid, inexpensive, non poluant and non destructive analytical tool, useful to detect traditional butter adulteration and to ensure its authenticity and safety.

FIGURES AND TABLES

1 FIGURES

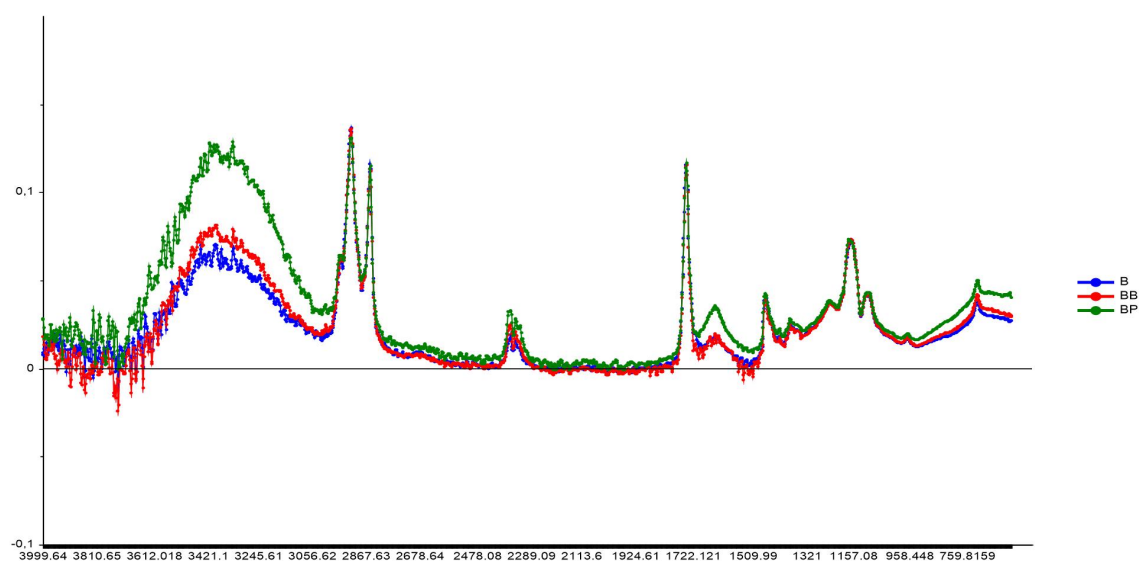


Figure.1. The mean spectra calculated for each class: Pure traditional cow's butter (B) ; Binary mixture with vegetable butter (BB) and with potatoes (BP), at MIR region of 4000–600 cm-1

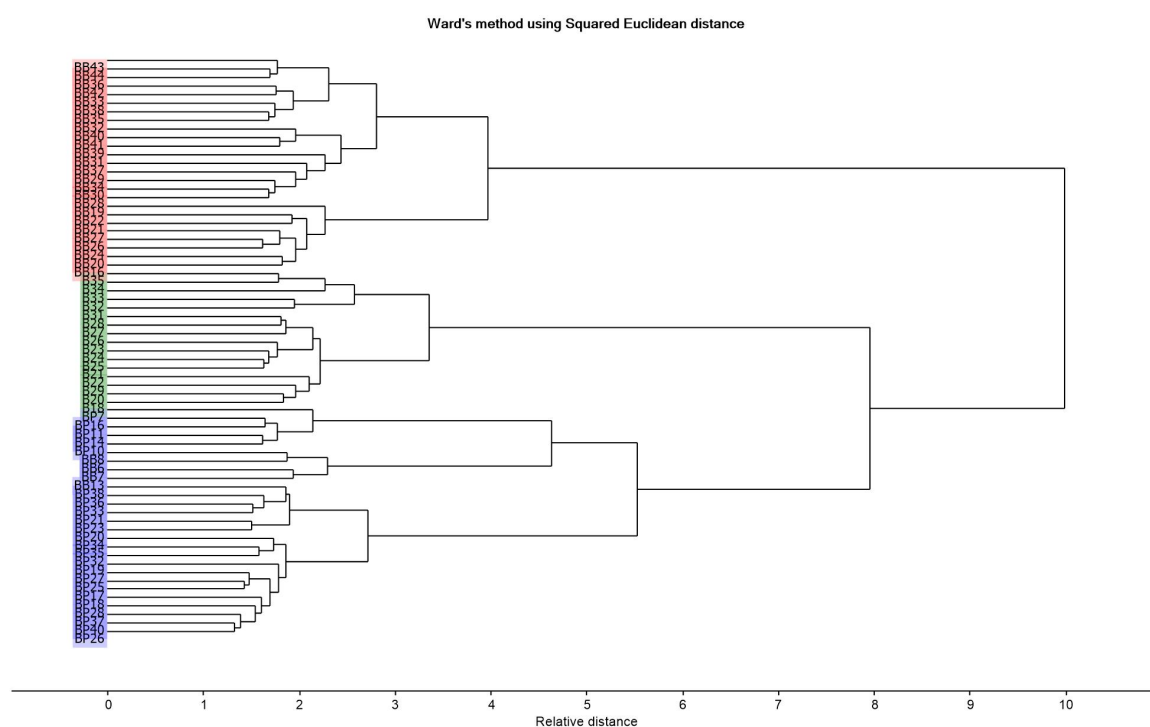


Figure.2. Dendrogram by CA analysis on the calibration set

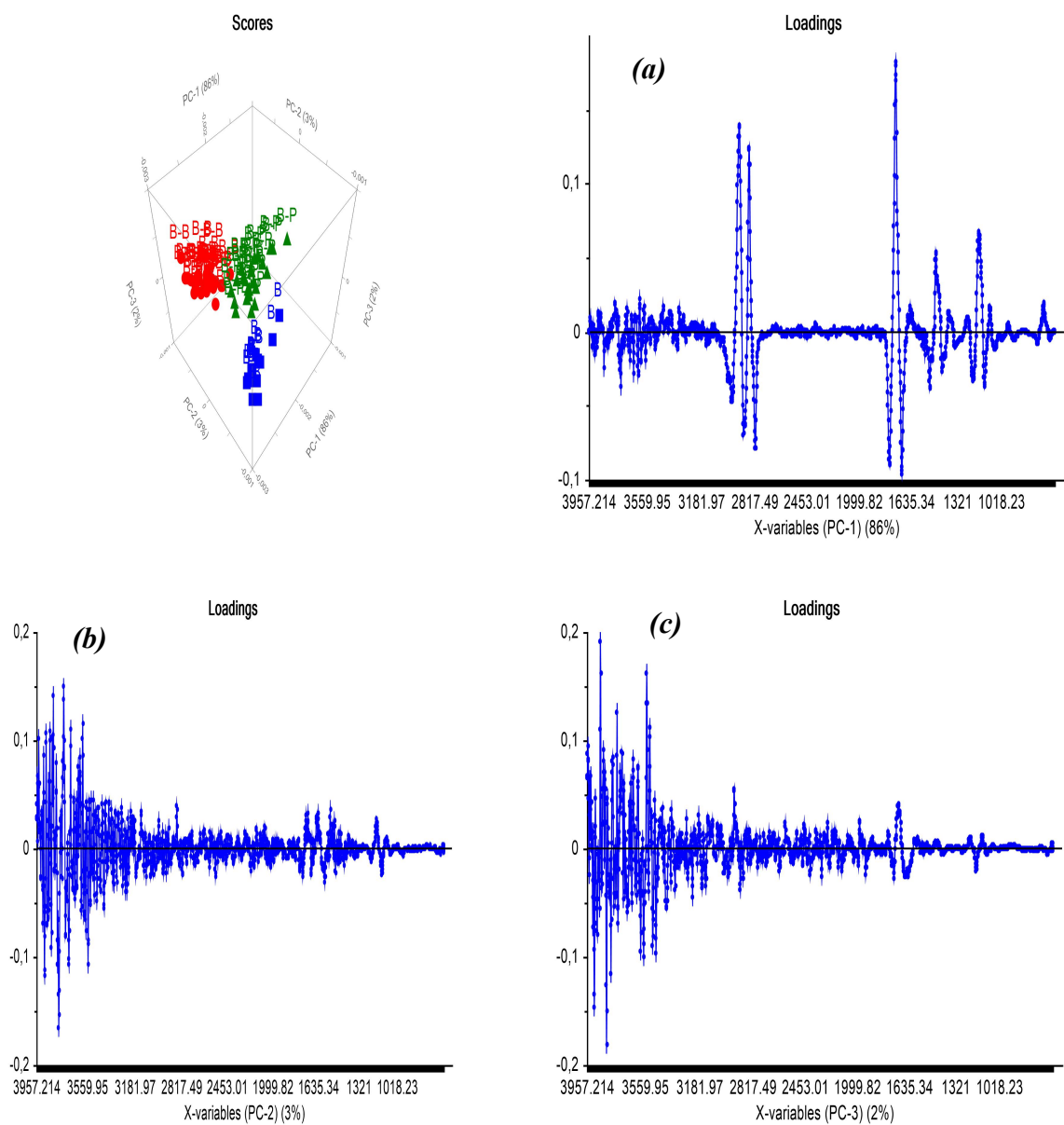


Figure.3. 3D-Score plot by PCA analysis on the calibration set: B, BB and BP;(a) first principal component; (b) second principal component ; (c) third principal component

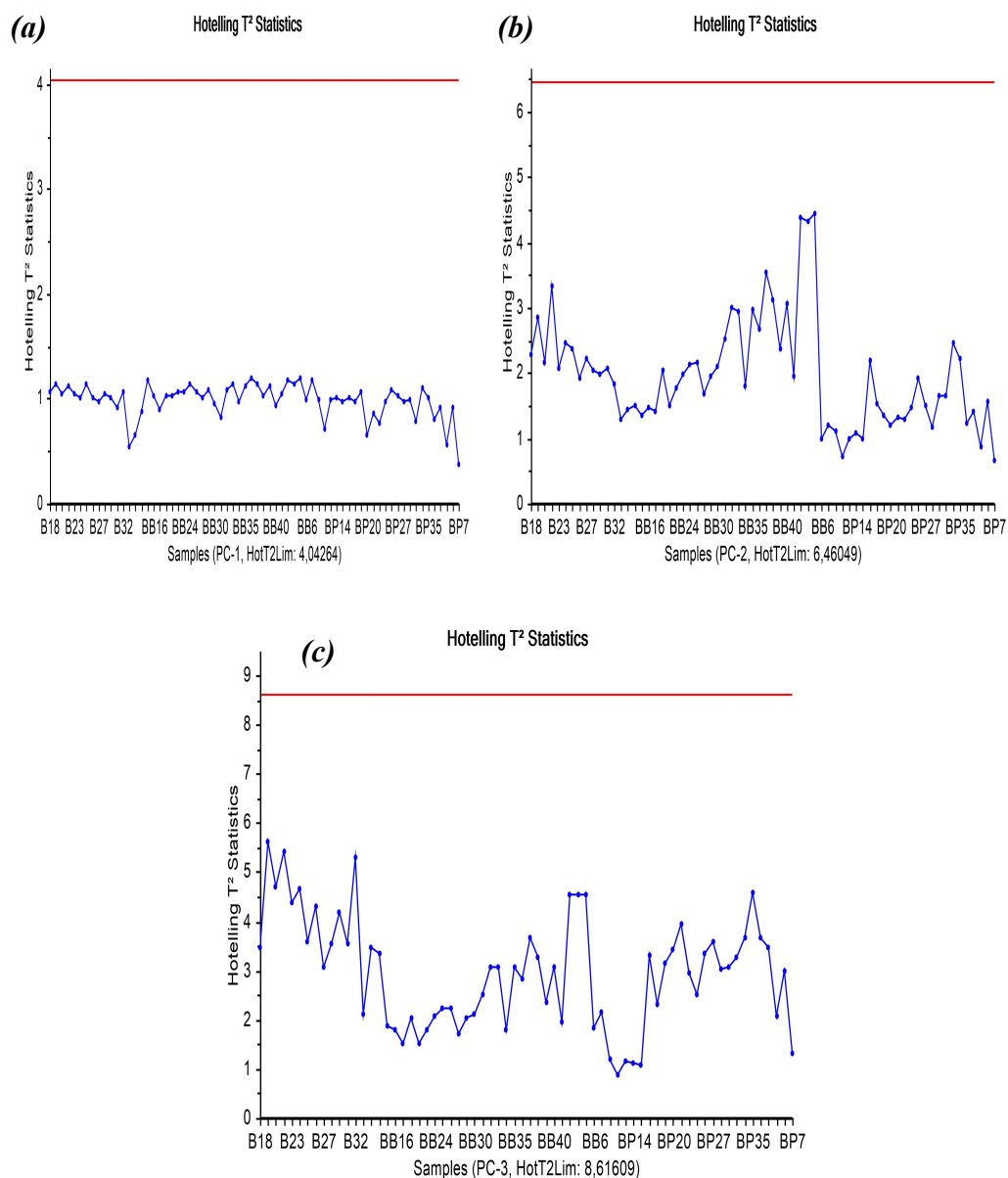


Figure.4. Hotelling T^2 statistics;(a) first principal component; (b) second principal component ; (c) third principal component

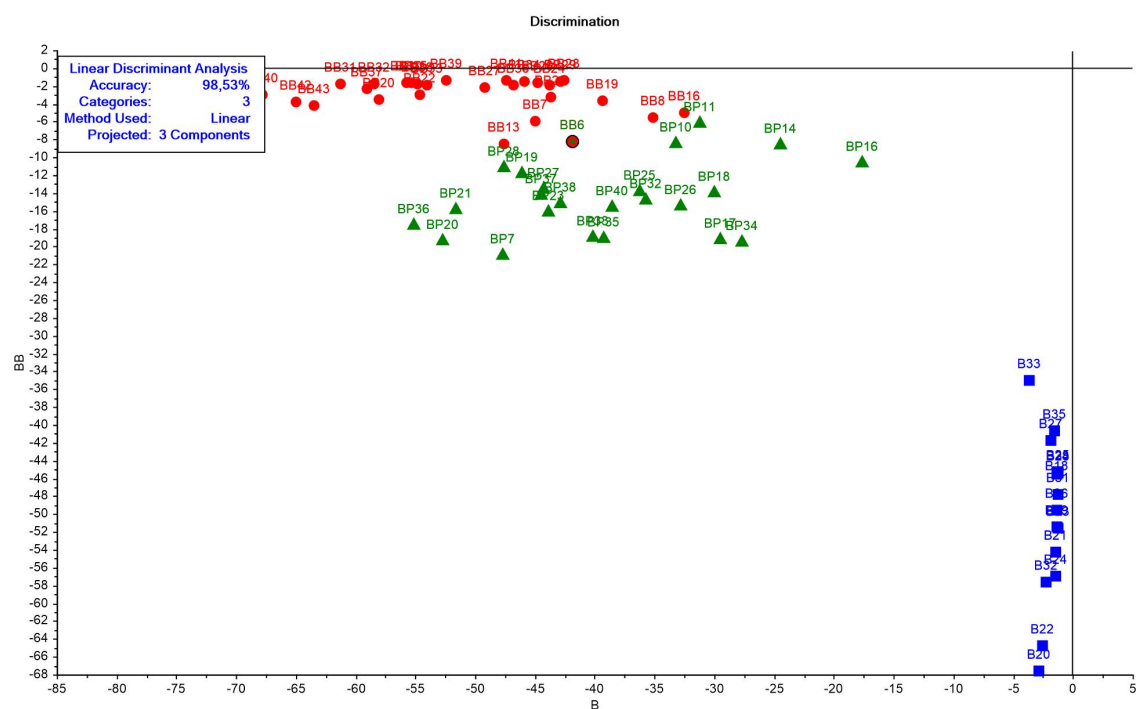


Figure. 5. Discriminant value plot for classes B and BB by calibration set: PCA-LDA application.

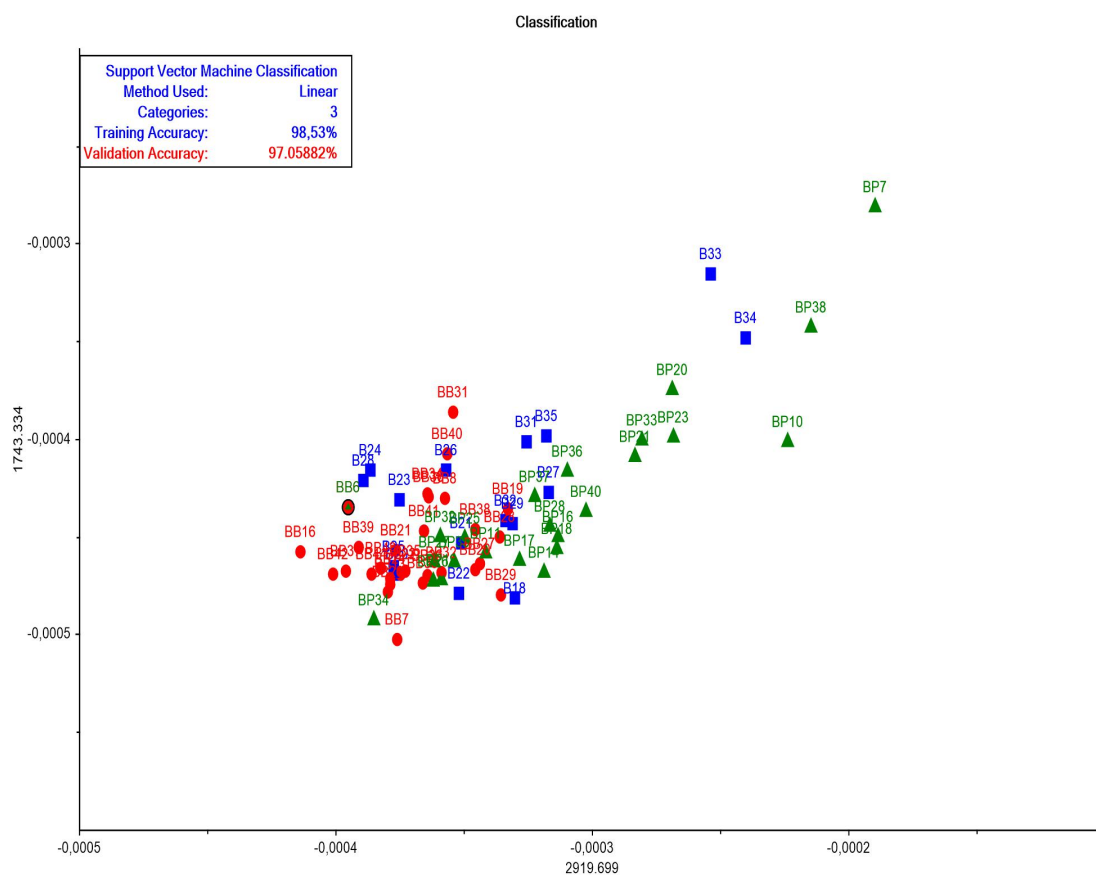


Figure.6. 2D score plot of classification results by SVM on calibration, with 2 wavelenths 2919 / 1743cm-1.

2 TABLES

Table .1. Confusion matrix of calibration set, carried out by PCA-LDA

Confusion matrix	B	BB	BP
Predicted	1	2	3
B	16	0	0
BB	0	28	0
BP	0	1	23

Table .2. Prediction matrix from application of the PCA-LDA classification model on the test set

Samples	B	BB	BP	Prediction
SB1	-2,128912	-38,717	-26,5097	B
SB2	-2,79373	-67,65346	-54,75658	B
SB3	-1,221444	-51,59914	-40,93527	B
SB4	-1,322393	-49,49378	-38,84457	B
SB5	-1,21376	-45,35158	-34,54589	B
SB6	-3,88595	-29,31243	-24,66796	B
SB7	-2,270581	-57,61974	-49,91995	B
SB8	-1,255718	-45,3704	-33,07034	B
SB9	-1,420079	-57,01016	-43,55743	B
SBB1	-29,69818	-8,703928	-6,416878	BP
SBB2	-52,23812	-2,319152	-21,92376	BB
SBB3	-43,68601	-1,900523	-16,98788	BB
SBB4	-48,65948	-1,334221	-14,82174	BB
SBB5	-77,76894	-4,414285	-21,70582	BB
SBB6	-55,24537	-1,44416	-14,97938	BB
SBB7	-55,6201	-2,74799	-10,94668	BB
SBB8	-46,69864	-2,628403	-8,227732	BB
SBB9	-38,85781	-7,968797	-16,82379	BB

Samples	B	BB	BP	Prediction
SBB10	-41,91574	-5,731844	-14,71617	BB
SBB11	-39,42472	-7,937254	-15,76689	BB
SBB12	-52,5403	-1,583055	-12,69373	BB
SBB13	-32,58458	-3,122275	-11,39774	BB
SBB14	-28,39509	-4,752233	-8,118093	BB
SBP1	-23,55257	-11,84522	-3,244364	BP
SBP2	-28,2391	-20,47558	-2,56382	BP
SBP3	-39,93977	-15,69444	-1,279514	BP
SBP4	-43,63299	-12,37474	-1,392289	BP
SBP5	-40,22206	-21,73471	-2,717322	BP
SBP6	-35,84037	-15,45045	-1,29155	BP
SBP7	-46,11365	-11,85668	-1,674345	BP
SBP8	-40,34827	-15,27582	-1,193655	BP
SBP9	-49,90308	-19,46367	-2,337693	BP
SBP10	-40,072	-21,00345	-2,00206	BP
SBP11	-45,31904	-15,57475	-1,532331	BP
SBP12	-26,18751	-15,07573	-2,173025	BP
SBP13	-27,2151	-15,18552	-1,995079	BP
SBP14	-40,15684	-14,77965	-1,16477	BP
SBP15	-45,77583	-18,35797	-1,825016	BP
SBP16	-45,18692	-15,16903	-1,694538	BP
SBP17	-20,9222	-12,89472	-4,076475	BP
SBP18	-29,13976	-6,720535	-3,838634	BP

Table .3. Confusion matrix of calibration set, carried out by SVM

Confusion matrix	B	BB	BP
Predicted	1	2	3
B	16	0	0
BB	0	28	0
BP	0	1	23

Table .4. Prediction matrix from application of the SVM classification model on the test set

Samples	Prediction	Reference class
SB1	B	B
SB2	B	B
SB3	B	B
SB4	B	B
SB5	B	B
SB6	B	B
SB7	B	B
SB8	B	B
SB9	B	B
SBB1	BB	BB
SBB2	BB	BB
SBB3	BB	BB
SBB4	BB	BB
SBB5	BB	BB
SBB6	BB	BB
SBB7	BB	BB
SBB8	BB	BB
SBB9	BB	BB
SBB10	BB	BB

Samples	Prediction	Reference class
SBB11	BB	BB
SBB12	BB	BB
SBB13	BB	BB
SBB14	BB	BB
SBP1	BP	BP
SBP2	BP	BP
SBP3	BP	BP
SBP4	BP	BP
SBP5	BP	BP
SBP6	BP	BP
SBP7	BP	BP
SBP8	BP	BP
SBP9	BP	BP
SBP10	BP	BP
SBP11	BP	BP
SBP12	BP	BP
SBP13	BP	BP
SBP14	BP	BP
SBP15	BP	BP
SBP16	BP	BP
SBP17	BP	BP
SBP18	BP	BP

References

- [1] Chia-Lin Lee, Huei-Lin Liao, Wan-Chen Lee, Chao-Kai Hsu, Fwu-Chyn Hsueh, Jyh-Quan Pan, Chih-Hsuan Chu, Chieh-Ting Wei, Ming-Ju Chen (2018). Standards and labeling of milk fat and spread products in different countries. Journal of food and drug analysis, 26, 4 69 - 480.

- [2] Terouzi, W., Kzaiber, F., Gorfti, A., Oussama, A (2017). Application of multivariate analysis to predict adulteration of Moroccan traditional butter by mashed potatoes. *International Journal of Engineering Research and Allied Sciences (IJERAS)* ISSN : 2455-9660, 01(10), 8-12.
- [3] Nikhil M. Jirankalgikar, Subrata De (2014). Detection of tallow adulteration in cow ghee by derivative spectrophotometry. *Journal of Natural Science, Biology and Medicine*, 5 (2), 317-319.
- [4] Terouzi, W., Ait Elkadi, Y., Gorfti, A., Kzaiber, F., Oussama, A (2018). Multivariate regression methods with infrared spectroscopy to detect the falsification of traditional butter. *Invention Journal of Research Technology in Engineering & Management (IJRTEM)* ISSN: 2455-3689, 2 (7), 87-96.
- [5] Esslinger, S., Riedl, J., Faulh-Hassek, C (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International*, 60, 189–204.
- [6] G. Dobson (2001). Lipid analysis by vibrational spectroscopy. *European Journal of Lipid Science and Technology*, 103, 815–840.
- [7] Y. W. Lai, E. K. Kemsley & R. H. Wilson (1994). Potential of Fourier transform infrared spectroscopy for the authentication of vegetable oils. *Journal of Agricultural and Food Chemistry*, 42, 1154–1159.
- [8] D. B. Dahlberg, S. M. Lee, S. J. Wenger & J. A. Vargo (1997). Classification of vegetable oils by FT-IR. *Applied Spectroscopy*, 51, 1118–1124.
- [9] N. Dupuy, L. Duponchel, J. P. Huvenne, B. Sombret & P. Legrand (1996). Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra. *Food Chemistry*, 57, 245–251.
- [10] B. F. Ozen, I. Weiss & L. J. Mauer (2003). Dietary Supplement Oil Classification and Detection of Adulteration Using Fourier Transform Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry*, 57, 5871–5876.
- [11] M. Safar, D. Bertrand, P. Robert, M. F. Devaux & C. Genot (1994). Characterization of edible oils, butters and margarines by Fourier transform infrared spectroscopy with attenuated total reflectance. *Chemists- Society*. 71, 371–377.
- [12] M. De Luca, W. Terouzi, G. Ioele, F. Kzaiber, A. Oussama, F. Oliverio, R. Tauler, G. Ragno (2011). Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. *Food Chemistry*, 124, 1113–1118.

- [13] N. Nicolaou, Y. Xu, R. Goodacre (2010). Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. *J. Dairy Sci.*, 93, 5651-5660.
- [14] N. Nicolaou and R. Goodacre (2008). Rapid and quantitative detection of the microbial spoilage in milk using Fourier transform infrared spectroscopy and chemometrics. *Analyst*, 133, 1424-1431.
- [15] G. Downey (1996). Authentication of food and food ingredients by near infrared spectroscopy. *J. Near Infrared Spectrosc.* 4 (1), 47–61.
- [16] G. Downey (1998). Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *Trends Anal. Chem.* 17, 418–424.
- [17] Y. Roggo, P. Chaluz, L. Maurer, C. Lema-Martinez, A. Edmond & N. Jent (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, 44, 683–700.
- [18] J. Yang & J. Yu Yang (2003). Why can LDA be performed in PCA transformed space. *Pattern Recogn.* 36, 563-566.
- [19] Terouzi, W., Platikanov, S., Anna de Juan Capdevila, A., and Oussama, A. (2013). Classification of olives from Moroccan regions by using direct FT-IR analysis: Application of support vector machines (SVM). *International Journal of Innovation and Applied Studies*. ISSN 2028-9324, 3 (2), 493-503.
- [20] Terouzi, W., Ait yacine, Z., Kzaiber, F., Oussama, A. (2016). Determining the geographical origin of olives by chemometric tools applied to ATR-FTIR spectra of mesocarp. *Asian Academic Research Journal of Multidisciplinary*, 3 (6), ISSN: 2319-2801, 176-192.
- [21] W. Terouzi, A. Oussama (2016). Characterization And Discrimination of Monocultivar Loquat (*Eriobotrya Japonica*) Fruit From two Areas in Beni Mellal-Khenifra Region. *International Journal of Research in Engineering and Science (IJRES)*, ISSN (Print): 2320-9356, 4 (7), 40-48.
- [22] Belousov AI, Verzhakov SA, von Frese J (2002). Application aspects of support vector machines. *J. Chemometrics*, 16, 482–489.
- [23] Broun, M.P.S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T.S., Manuel Ares, Jr. and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *PNAS*, 97 (1), 262-267.

- [24] A. P. Reynolds, G. Richards, B. dela Iglesia and V. J. Rayward-Smith (2006). Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms*, 5, 475–504.
- [25] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S.D.E. Jong, P.J., Lewi. Smeyers Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, The Netherlands, 1998, p. 88.
- [26] S. Wold, K.Esbensen & P. Geladi. Principal component analysis. *Chemom. Intell. Lab. Syst.*, 2, 1987, 37-52.
- [27] R. Garrido-Delgado, L. Arce, A.V. Guamán, A.. Pardo, S. Marco, M. Valcárcel (2011). Direct coupling of a gas–liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools. *Talanta* 84, 471–479.
- [28] A.Gori, R. M. Maggio, L. Cerretani, M. Nocetti, M. Fiorenza Caboni (2012). Discrimination of grated cheeses by Fourier transform infrared spectroscopy coupled with chemometric techniques. *International Dairy Journal* 23, 115-120.
- [29] Xu, L., Shao, X.G. *Chemometric Methods*. Beijing: Science Press, 2006, 130 -133.
- [30] Vapnik, V., Chervonenkis, A. *Theory of Pattern Recogn*, Nauka, Moscow, 1974.
- [31] Vapnik, V. *The nature of statistical learning theory*, Springer Verlag, 1995.
- [32] Vapnik, V. *The Nature of Statistical Learning Theory*. Second edition. New York: Springer, 1999.
- [33] Drucker, H., Burges, C.J., Kaufman, L., Smola, A (1997). and Vapnik, V. Support Vector Regression Machines. *Adv Neural Inf Process Syst*, 9, 155-161.
- [34] Savitzky, A. & Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry*, 36, 1964, 1627–1639.
- [35] Norris, K.H. & Williams, P.C. Optimization of mathematical treatments of raw near infrared signal in the measurement of protein in hard Red Spring wheat, I: influence of particle size. *Cereal Chemistry*, 62, 1984,158–165.
- [36] Iñón, F. A., Garrigues, J. M., Garrigues, S., Molina, A., and de la Guardia, M. (2003). Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. *Analytica Chimica Acta*, 489, 59–75.
- [37] KH. Esbensen. *Multivariate data analysis-in practice. An introduction to multivariate data analysis and experimental desing*.Oslo: CAMO. 2000.
- [38] Scott, I.M., Lina, W., Liakata, M., Wood, J.E., Vermeer, C.P., Allaway, D., Ward, J.L., Draper, J., Beale, M.H., Corol, D.I., Baker, J.M., King, R.D. (2013). Merits of random

forests emerge in evaluation of chemometric classifiers by external validation. *Analytica Chimica Acta* 801, 22– 33.

- [39] Ankit Bansal, Vikas Chhabra, Ravindra K. Rawal, Simant Sharma (2014). Chemometrics: A new scenario in herbal drug standardization. *Journal of Pharmaceutical Analysis*, <http://dx.doi.org/10.1016/j.jpha.2013.12.001>.