



BCSIR

Available online at www.banglajol.info

Bangladesh J. Sci. Ind. Res. 50(4), 233-240, 2015

BANGLADESH JOURNAL
OF SCIENTIFIC AND
INDUSTRIAL RESEARCH

E-mail: bjisir07@gmail.com

ATR-FTIR spectroscopy and chemometric techniques for the identification of edible vegetable oils

M. M. Karim^{1*}, S. F. E. Karim^{2*}, A. A. Rana¹, S. M. Masum¹, A. Mondol¹ and S. S. Israt¹

¹Department of Applied Chemistry and Chemical Engineering, University of Dhaka, Dhaka-1000, Bangladesh.

²Department of Banking and Insurance, University of Dhaka, Dhaka-1000, Bangladesh.

Abstract

There is a large variety and trademarks of vegetable oils in Bangladesh. The oils have characteristics very similar to each other and often cannot be classified by the observation of color, odor or taste. This paper proposes a vibrational spectroscopic method like FTIR in association with chemometric techniques to classify vegetable oils like: sunflower, mustard, sesame, soybean, castor, olive and palm oils from different manufacturers. In the FTIR spectra of oil, as information about fatty acid composition is concentrated in the range of 4400-200 cm⁻¹ principal component analysis (PCA) was applied on the standardized full FTIR spectral data of this region for vegetable oils to totally capture the FTIR spectral pattern; seven varieties of vegetable oils could be successfully classified from their PCA scores. PCA of FTIR spectra of different known vegetable oils is used to determine the identity of several unknown vegetable oils. The unknowns are then analyzed, plotted, and identified based on their proximity to the known in principal component space. For the multivariate analysis PCA and soft independent modeling of class analogy (SIMCA) and support vector machines (SVMs) were used. 85% and 14% variability of data was explained by PC1 and PC2 respectively. PC1 has strong positive correlation with soybean, sunflower, palm and olive oil while strong negative correlation with mustard, castor and sesame oil. Soybean oils are positively and sesame oils are negatively correlated with PC2. Unknown oil samples can be identified properly by used supervised methods i.e. SIMCA, SVM by developing model with the help of PCA. The major interest of this method using chemometric analysis of spectral data is in their rapidity, since no chemical treatment of samples is required.

Keywords: Vegetable oil; Principal component analysis (PCA); Soft independent modeling of class analogy (SIMCA); Support vector machine (SVM)

Introduction

Edible vegetable oils are widely used in cooking oils, salad oils, liquid and solid shortenings, spreads and ingredients in several foods, like bakery products and fried foods (O'Brien, 2009). Due to the large variety of brands and trademarks of vegetable oils in the Bangladesh market, it is common not to know for sure if the substance that is being bought is really a product without adulteration. Moreover, vegetable oils have characteristics quite similar to each other and often cannot be distinguished by only observing the color, odor or taste.

As food safety and quality that are considered important issues are directly related to the human health and social progress people increasingly look for trusted brands of food products, and expect manufacturers and retailers to provide high quality products (Gori *et al.*, 2012). Increased consumer awareness of food safety and quality issues has led to the development of new and increasingly sophisticated techniques for food product classification.

Fats and oils (Feinberg *et al.*, 1987) are among the oldest classes of chemical compounds used by humans. They are composed of glyceric esters of fatty acids and are distinguished by their physical state: fats are solid at ambient temperature, whereas oils are liquid. The principal variation in composition is the chain length and degree of unsaturation of the component fatty acids (Sadeghi-Jorabchi *et al.*, 1991). Oils such as sunflower seed oil have a high concentration of monounsaturated and polyunsaturated fatty acids. Olive oil is 100% of olive origin, which corresponds with a high level of monounsaturated fatty acids.

Since the infrared (IR) spectra contain significant information about all the components of a complex mixture, fourier transform infrared (FTIR) spectroscopy is a very powerful and general technique for investigating the structure and monitoring the changes in organic compounds (e.g., composite materials) and inorganic nanoparticles (Rana

*Corresponding author: E-mail: moinulcd@yahoo.com, sfek05@gmail.com

et al., 2015; Ferdous *et al.*, 2015). Fourier transform infrared (FTIR) spectroscopy with attenuated total reflectance (ATR) has been used to authenticate, identify or classify fats and oils (Lai *et al.*, 1994; Dahlberg *et al.*, 1997; Dupuy *et al.*, 1996; Ozen *et al.*, 2003; Safar *et al.*, 1994). Generally, an ATR accessory is used with the FTIR spectrometer for oil and fat study, due to its ability to handle liquid samples easily.

In combination with chemometric techniques such as principal component analysis (PCA) (Martens, 1979; Dupuy *et al.*, 1995) vibrational spectroscopy like ATR-FTIR allows classification of foods to be undertaken without any chemical analysis. The main advantage is that no prior information on the mixtures is needed since the significant information is extracted during the statistical treatment.

Recently, multiple linear regression analysis using FTIR spectroscopy has been used often for establishing calibration equations for determining various constituents in agricultural and food products. However, useful information may be lost in this process because only a few wavelengths are utilized in this statistical method. Therefore, it is necessary to capture the complete FTIR spectral pattern to obtain full information. Further, an attempt has begun to apply FTIR data for qualifying or classifying purposes in addition to quantifying purposes.

Principal component analysis (PCA) (Massart *et al.*, 1991), a data compression method in which little information is lost, has also been applied for analyzing FTIR spectral patterns (Cowe and Mc Nicol 1985; Cowe *et al.*, 1988; Bertrand *et al.*, 1985; Bertrand *et al.*, 1987; Downey *et al.*, 1990; Robert *et al.*, 1987). However, when PCA is applied, there might be possibility that samples are classified according to their level of moisture or to their particles size, not to their own peculiar compositional characteristics, because not only absorption characteristics but also scattering determines the pattern of their FTIR spectra. The FTIR spectral pattern of oil represents its fatty acid composition because FTIR absorption bands 200-4400 cm^{-1} are assigned to straight carbon chains and cis double bonds that reflect fatty acid moieties in fat molecules. Because oil consists almost totally of triglycerides and there may be no need to consider either interference from water absorption or the size effect of scattering particles; the characteristic wavelengths with high loading weight seem to be assigned more easily for oils than for other materials. So, the relationship between the PCA results and the assignments of wavelengths with high loading weight can be characterized more properly. In this study, PCA was applied to the measured full-range FTIR spectral data of some vegetable oils to classify them, and the feature of PCA applied to the FTIR spectral was also investigated. Principal component

analysis (PCA) was used for unsupervised pattern recognition while soft independent modeling of class analogy (SIMCA) (Veras *et al.*, 2010) and support vector machines (SVMs) were used for supervised pattern recognition.

Theory

PCA treats spectral peak position as vectors (x_1, x_2, \dots, x_n) and forms linear combinations of the vector assigning a weight (a_1, a_2, \dots, a_n) to each vector. The weight is chosen from each set of peak positions. This maximization of the variation is subject to the constraint that the sum of the squared weight is equal to one. The linear combinations created are called principal components and can be expressed in the form $y = a_1x_1 + a_2x_2 + \dots + a_nx_n$ (Duntman, 1989; Jackson, 1983; Gordon, 1981).

When two principal components are used as the axes on which to plot the peak position data, the result is that most of the variation present in the original n-dimensional data set can be expressed in a two-dimensional scatter plot. By assessing the proximity of the unknown to the known in this principal components space, unknowns can be classified.

PCA is not useful for differentiating defined classes, since the class information is not used in the construction of the model and PCA just describes the overall variation in the data. However PCA can be coupled with the class information in order to give classification models by means of soft independent modeling of class analogy (SIMCA) (Beebe *et al.*, 1998). SIMCA was the first class modeling technique introduced in chemistry and nowadays is one of the best-known modeling classification methods; it is defined 'soft' since no hypothesis on the distribution of variables is made, and 'independent' since the classes are modeled one at a time (i.e. each class model is developed independently). Basically, a SIMCA model consists of a collection of G PCA models, one for each of G defined classes. Therefore, PCA is separately calculated on the objects of each class; since the number of significant components can be different for each category, cross-validation has been proposed as a way of choosing the number of retained components of each class model. In this way, SIMCA defines G subspaces (class models); then, a new object is projected in each subspace and compared to it in order to assess its distance from the class. Finally, the object assignment is obtained by comparing the distances of the object from the class models. Even if SIMCA is often a useful classification method, it has also some disadvantages. Primarily, the class models in SIMCA are calculated with the aim of describing variation within each class: when PCA is applied on each category, it finds the directions of maximum variance in the class space. Consequently, no attempt is made to find directions that

separate classes, on the opposite of, for example, partial least squares discriminant analysis (PLS-DA), which directly models the classes on the basis of the descriptors (Marengo *et al.*, 2006)

Support vector machines (SVMs) have been developed by Vapnik (Vapnik, 1998) and details about SVM classifier can be found in literature (Amendolia *et al.*, 2003; Cristianini and Shawe-Taylor, 2000). SVM is also showing high performances in practical applications. In very simple terms, an SVM corresponds to a linear method in a very high dimensional feature space that is nonlinearly related to the input space. Even though we think of it as a linear algorithm in a high dimensional feature space; in practice, it does not involve any computations in that high dimensional space.

Support vector machines map input vectors to a higher dimensional space where a maximal separating hyper plane is constructed (Amendolia *et al.*, 2003; Cristianini and Shawe-Taylor, 2000). Two parallel hyper planes are constructed on each side of the hyper plane that separates the data. The hyper plane is the hyper plane that maximizes the distance between the two parallel hyper planes. An assumption is made that the larger the margin or distance between these parallel hyper planes the better the generalization error of the classifier will be.

Materials and methods

Materials

Seven different vegetable oils i.e. soybean, sesame, mustard, sunflower, olive, palm and castor oil were purchased from the local super market as standard and unknowns from open market. At least eight replicate spectra of each standard & two replicate spectra of unknown oils were taken to establish the training data set.

Instrumentation and software

Attenuated total reflectance-fourier transform infrared spectroscopy (ATR-FTIR) spectra were recorded on Shimadzu FTIR Model IR Prestige-21 (Shimadzu Corp., Kyoto, Japan) with Pike Miracle ATR and analyzed using IR solution version 1.30 software. The sampling station was equipped with an overhead ATR accessory comprising of transfer optics within the chamber through which infrared radiation is directed to a detachable ATR zinc selenide crystal mounted in a shallow trough for sample containment. Chemometric analysis including PCA (Principal Component Analysis), SVM (Support Vector Machine), SIMCA (Soft Independent Modeling of Class

Analogy) were performed using Unscrambler X software (Camo, USA) (version 9.7 and demo version 10.3).

Spectral acquisition

Each sample (approximately 0.5 mL) was applied to the ATR plate by direct transfer from the bottle. Multiple spectra (eight times) were collected for each sample. All spectra were recorded from 4000 to 500 cm^{-1} at a resolution of 1cm^{-1} . The ATR crystal was carefully cleaned with pure chloroform to eliminate the presence of oil/fat residues between measurements and dried with soft tissue after each experiment to ensure a clean crystal surface so as to obtain the best possible sample spectra.

Results and discussion

One spectrum of each of the unknown oils was obtained for establishing test set. A plot of superimposing spectra from training set is shown in the Fig. 1.

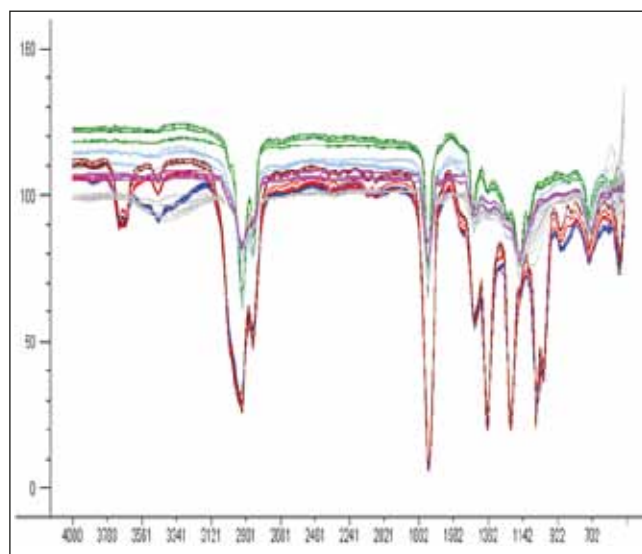


Fig. 1. The FTIR spectra of 7 different types of vegetable oils: sunflower, mustard, sesame, soybean, castor, olive, palm oil where X-axis and Y-axis represent wave number/ cm^{-1} and absorbance respectively

The FTIR spectra of sunflower, mustard, sesame, soybean, castor, olive, palm oil are similar in appearance. It is difficult to identify the oils by the variation within the FTIR spectra. In this study, the PCA is “trained” by acquiring spectra of sunflower, mustard, sesame, soybean, castor, olive, palm oil and determining the peak positions for each spectrum.

Principal component analysis (PCA)

As there are a large number of variables, PCA is used to reduce these numbers of variables by correlating them with one another. PCA is performed using unscrambler X software for training set of sample i.e. total 56 samples of sunflower, mustard, sesame, soybean, castor, olive, palm oil where 8 samples were taken from each of the oil. Here FTIR spectra wavelengths range from 4000 cm^{-1} to 500 cm^{-1} . Fig. 2. shows 98.24% of variable of data explained by first two PCs i.e. PC1 and PC2. So it can be said that first two PCs are enough for explaining variation of data.

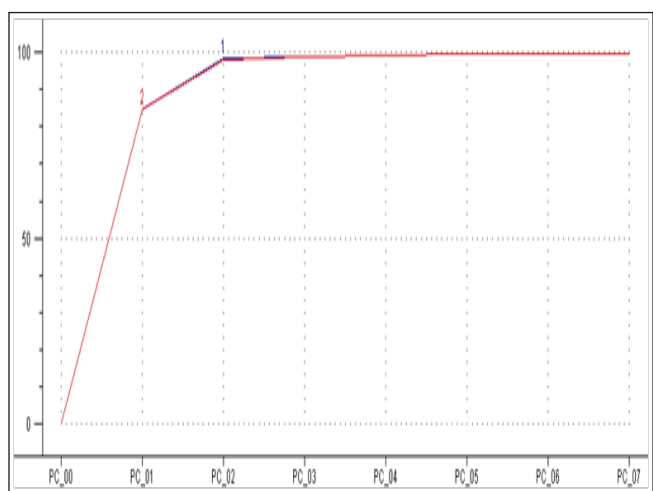


Fig. 2. Total Variance explained by PCs

Fig. 3. shows loading plot of PC1 & PC2. From here we observed the wave lengths which are more correlated with the given PCs.



Fig. 3. Loading plot

In Fig. 4 score plot is observed. It may be mentioned that soybean, sunflower, palm and olive oils are highly and positively correlated with PC1 and mustard, castor and

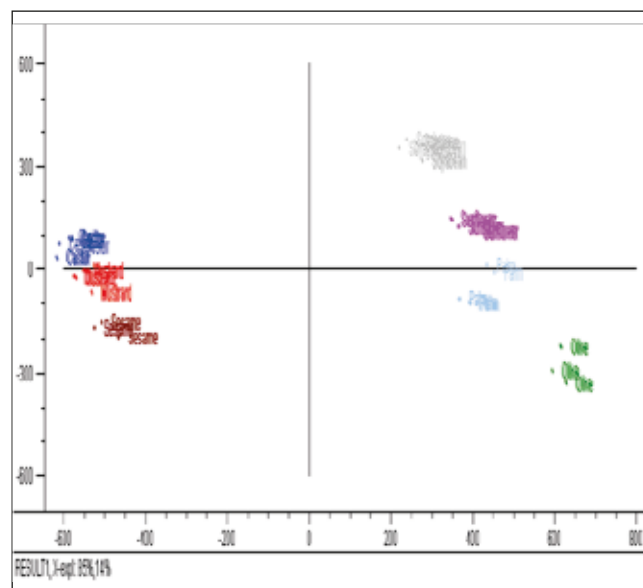


Fig. 4. Score plot of PC1 vs PC2

sesame oils are highly but negatively correlated with PC1. It is also found that 85% variability of the data explained by PC1 whereas PC2 explained only 14% variability. Soybean oils are positively correlated with PC2 and sesame and olive oil are negatively correlated. But sunflower, mustard, palm and castor oil are not significantly correlated with PC2.

Table I. Classification of unknown oil by SIMCA

Sample	castor	mustard	Olive	Palm	Sesame	Soyabean	sunflower
CastorUN1	*						
CastorUN2	*						
MustardUN1		*					
MustardUN2		*					
OliveUN1			*				
OliveUN2			*				
PalmUN1				*			
PalmUN2				*			
SesameUN1					*		
SesameUN2					*		
SoyabeanUN1						*	
SoyabeanUN2						*	
SunflowerUN1							*
SunflowerUN2							*

Independent modeling of class analogy (SIMCA)

Soft independent modeling of class analogy (SIMCA) is used for the classification of unknown oil. As SIMCA is based on PCA, individual PCA model for each oil is created using data from training set. With the help of each individual PCA model unknown oil of test set can be easily classified. Table I shows the classification of unknown oils by using SIMCA.

From the Table I it is observed that unknown oils are positioned nicely with their respective group.

Fig. 5 shows Coomans plot of castor oil versus mustard oil at 5% significant level. It is observed from Fig. 5 that unknown sample of castor and mustard oil is classified nicely at the given significance level. This is also true for other oils.

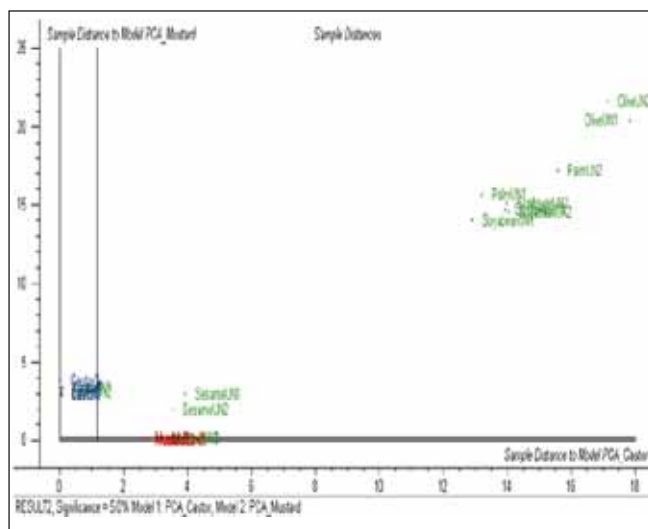


Fig. 5. Coomans plot for castor and mustard oil

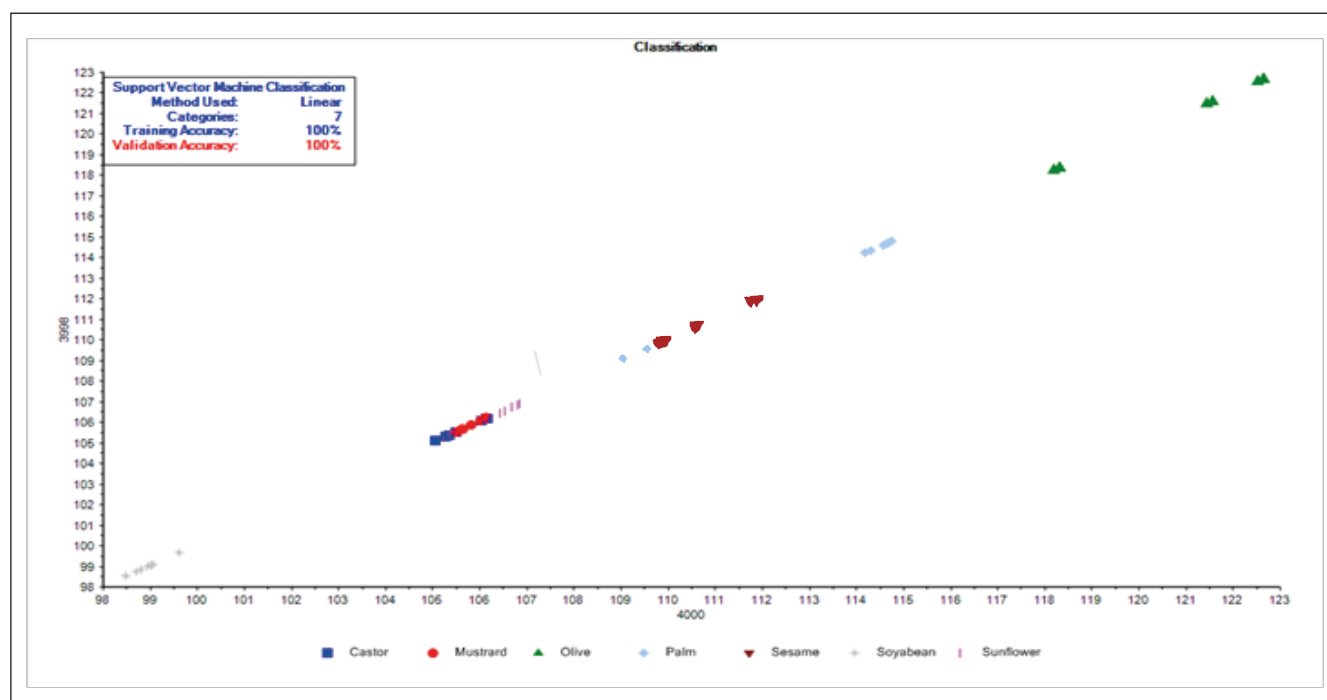


Fig. 6. SVM classification model for training data set

Support vector machine (SVM)

We also used support vector machine (SVM) classification method for our training data set to classify the individual oil into individual group. Here we used categorical variable i.e. type of oil as a classifier. Fig. 6 shows how the oil is classified in their respective group.

It is observed from the Fig. 6 that soybean, sunflower, palm and sesame are grouped separately. But mustard and castor oil are overlapped each other. Palm oils are separated into two groups. It is also observed that olive oil is grouped uniquely but there is also subgroup into the group.

We are also trying to predict our test set of unknown oil by using SVM classification model. The output of the prediction is given in Table 2 where we observed that unknown oils are classified as with their respective group of test data set. As SVM is a supervised method of classification or discrimination, which produces a model using training data set (Fig. 6). We can identify unknown oil by its nearest position of known group. In Table II standard column represent the group which is made on the basis of training data. In sample column it is observed that unknown samples are positioned as their respective class.

Table II. Classification of test samples by SVM classification model

Sample	Standard
Castor UN1	Castor
Castor UN2	Castor
Mustard UN1	Mustard
Mustard UN2	Mustard
Olive UN1	Olive
Olive UN2	Olive
Palm UN1	Palm
Palm UN2	Palm
Sesame UN1	Sesame
Sesame UN2	Sesame
Soy bean UN1	Soyabean
Soy bean UN2	Soyabean
Sunflower UN1	Sunflower
Sunflower UN2	Sunflower

Conclusion

ATR-FTIR spectroscopy combined with multivariate calibration is powerful in discriminating or identification among edible oils (sunflower, mustard, sesame, soybean, castor, olive, palm oil) through comparing the spectral data of these oils in the range of 4400–200 cm^{-1} without the need for running expensive standard chemical methods. The results obtained herewith are expected to be useful for the determination and routine test purpose. Apart from that, present study suggests a reliable quantitatively method for detecting adulteration of the olive oil with corn or sunflower oils.

References

- Amendolia SR, Cossu G, Ganadu ML, Golosio B, Masala GL and Mura GM (2003), A comparative study of k-nearest neighbour, support vector machine and multi-layer perception for thalassemia screening, *Chemometrics and Intelligent Laboratory Systems* **69**(1): 13-20.
- Beebe KR, Pell RJ and Seasholtz M B (1998), *Chemometrics: A practical guide*, John Wiley & Sons, Inc., New York, ISBN 0-471-12451-6.
- Bertrand D, Lila M, Furtoss V, Robert P and Downey G (1987), Application of principal component analysis to the prediction of lucerne forage protein content and in vitro dry matter digestibility by NIR spectroscopy, *Journal of the Science of Food and Agriculture* **41**(4): 299-307.
- Bertrand D, Robert P and Loisel W (1985), Identification of some wheat varieties by near infrared reflectance spectroscopy, *Journal of the Science of Food and Agriculture* **36**(11): 1120-1124.
- Cristianini N and Shawe-Taylor J (2000), *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press.
- Cowe IA, Mc Nicol JW and Cuthbertson DC (1988), Reconstruction of constituents spectra from near infrared sample spectra using principal components, *Analyst* **113**(2): 269-272.
- Cowe I A and Mc Nicol J W (1985), The use of principal component in the analysis of near-infrared spectra, *Applied Spectroscopy* **39**(2): 257-266.
- Dahlberg DB, Lee SM, Wenger SJ and Vargo JA (1997), Classification of vegetable oils by FT-IR, *Applied Spectroscopy* **51**: 1118–1124.
- Dupuy N, Duponchel L, Huvenne JP, Sombret B and Legrand P (1996), Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra, *Food Chemistry* **57**: 245–251.
- Dupuy N, Huvenne JP, Duponchel L and Legrand P (1995), Classification of green coffees by FTIR analysis of dry extract, *Applied Spectroscopy* **49**(5): 580-585.
- Downey G, Robert P, Bertrand D and Kelly PM (1990), Classification of commercial skim milk powders according to heat treatment using factorial discriminant analysis of near-infrared reflectance spectra, *Applied Spectroscopy* **44**(1): 150-155.
- Dunteman GH (1989), *Principal components analysis*. In Sage University series on quantitative applications in the social sciences, 07-069, Sage publications, Newbury Park, CA.
- Feinberg M, Favier JC and Ireland-Ripert J (1987), *Repertoire General des Aliments. Table de Composition des Corps Gras*, Tech. Dot. Lavoisier/Ciqual-Regal, INRA, Paris.
- Gori A, Cevoli C, Fabbri A, Caboni MF and Losi G (2012), A rapid method to discriminate season of production and feeding regimen of butters based on infrared spectroscopy and artificial neural networks, *Journal of Food Engineering* **109**(3): 525-530.
- Gordon A D (1981), *Classification*, Chapman & Hall, New York.
- Hossan MJ, Gafur MA, Karim MM and Rana AA (2015), Mechanical properties of Gelatin-Hydroxyapatite composite for bone tissue engineering, *Bangladesh J. Sci. Ind. Res.* **50**(1): 15-20.
- Islam ASMA, Ferdous T, Das AK, Karim MM and Masum SM (2015), Photocatalytic degradation of direct brown RN dye in the presence of ZnO nanoparticles, *Bangladesh J. Sci. Ind. Res.* **50**(1): 1-6.
- Jackson BB (1983), *Multivariate data analysis*, Richard D. Irwin, Inc, Homewood IL.
- Lai YW, Kemsley EK and Wilson RH (1994), Potential of Fourier transform infrared spectroscopy for the authentication of vegetable oils, *Journal of Agricultural and Food Chemistry* **42**:1154–1159.
- Marengo E, Bobba M, Robotti E and Liparota MC (2006), Modeling of the polluting emissions from a cement

- production plant by partial least-squares, principal component regression, and artificial neural networks, *Environmental Science and Technology* **40**: 272–280.
- Massart DL, Brereton RG, Dussy RE, Hopke PK, Spiegelman CH and Wegscheider W (1991), *Chemometrics tutorials*: Elsevier, Amsterdam, ISBN 0-444-88837-3.
- Martens H (1979), Factor analysis of chemical mixture, *Anal. Chim. Acta* **112**: 423–442.
- O'Brien RD (2009), *Fats and Oils: Formulating and Processing for Applications*, 3rd Ed. CRC Press, Florida.
- Ozen BF, Weiss I and Mauer LJ (2003), Dietary supplement oil classification and detection of adulteration using Fourier transform infrared spectroscopy, *Journal of Agricultural and Food Chemistry* **57**: 5871–5876.
- Robert P, Bertrand D, Devaux MF and Grappin R (1987), Multivariate analysis applied to near-infrared spectra of milk, *Analytical chemistry* **59**(17): 2187–2191.
- Safar M, Bertrand D, Robert P, Devaux MF and Genot C (1994), Characterization of edible oils, butters and margarines by Fourier transform infrared spectroscopy with attenuated total reflectance, *Journal of the American Oil Chemists Society* **71**: 371–377.
- Sadeghi-Jorabchi H, Wilson RH, Belton P, Edward-Weeb JD and Coxon DT (1991), Quantitative analysis of oils and fats by Fourier transform Raman spectroscopy, *Spectrochim. Acta* **47A**: 1449–1458.
- Veras G, Gomes A A, Silva A C, Brito ALB, Almeida PBA and Medeiros EP (2010), Classification of biodiesel using NIR spectrometry and multivariate techniques, *Talanta* **83**: 565–568.
- Vapnik VN and Vapnik V (1998), *Statistical learning theory* (Vol. 1) New York: Wiley.

Received: 27 August 2015; Revised: 6 September 2015;

Accepted: 19 October 2015.