

Chemometrics Assisted Method for Classification of Mango Juice as Adulterated or Safe with over Use of Artificial Colours by UV Spectroscopic Data

Mohammad Nashir Uddin^{1,*}, Ajit Kumar Majumder², Abu Tareq Mohammad Abdullah³,
Md. Alamgir Kabir³

¹BCSIR Laboratories, Dhaka, Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

²Department of Statistics, Jahangirnagar University, Dhaka, Bangladesh

³Institute of Food Science and Technology (IFST), Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, Bangladesh

Email address:

m2nashir@yahoo.com (M. N. Uddin)

*Corresponding author

To cite this article:

Mohammad Nashir Uddin, Ajit Kumar Majumder, Abu Tareq Mohammad Abdullah, Md. Alamgir Kabir. Chemometrics Assisted Method for Classification of Mango Juice as Adulterated or Safe with over Use of Artificial Colours by UV Spectroscopic Data. *Journal of Food and Nutrition Sciences*. Vol. 5, No. 2, 2017, pp. 51-56. doi: 10.11648/j.jfns.20170502.15

Received: December 10, 2016; **Accepted:** March 15, 2017; **Published:** March 28, 2017

Abstract: Tartazine, Sunset Yellow and Beta Carotene are commonly used artificial colours in commercial mango juices in order to make them attractive to the consumers, even though these synthetic colours have hazardous effect on health. Therefore, it is very often necessary to classify juices as adulterated with heavy use of these colours or not. In the present study, two chemometric techniques, Artificial Neural Network (ANN) and Partial Least Square-Discrimination Analysis (PLS-DA) have been assessed for their efficiencies for classification. Here, UV spectroscopic data are used as input. Three techniques, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), Savitzky-Golay (S-G) filtering have been evaluated for their de-noising performance, and select the best one. Before calibration, spectral data are de-noised with MSC as is proved most efficient for de-noising UV spectral data. Spectral range 371-533 nm has been used for calibration ultimately. ANN shows better classification results than PLS-DA for all colours. Finally, the study is proposing a simpler and cheaper method for classification of mango juice as adulterated or safe with over use of artificial colours by applying ANN in de-noised spectroscopic data.

Keywords: Artificial Colours, Chemometrics, De-Noising, Classification, ANN, PLS-DA

1. Introduction

Use of excessive artificial colours in fruit juices is a very common practice in Bangladesh. Artificial colours and synthetic sweetener are added in the commercial juice in addition to natural fruit juice to make the juice more attractive. Synthetic colours such as, Tartrazine, Sunset Yellow and Beta Carotene are commonly used in commercial mango juice. These colours are very much hazardous for health, and their uses are banned or restricted in many part of the world [1-4]. Children are the most vulnerable group to be affected by harmful effect of artificial colours [5].

Traditionally, chemical methods are used to quantify artificial colours in food. Usually, chromatographic methods like High Performance Liquid Chromatography (HPLC) [6], High-performance ion chromatography (HPIC) [7] and Micellar Electrokinetic Capillary Chromatography (MEKC) [8] are used for this purpose.

For determination of synthetic colours in different foods, spectroscopic methods are also very popular [9, 10]. Artificial colours are extracted, identified in foods by spectrophotometric techniques and validated the results by chromatographic techniques in some researches [11, 12].

However, Chemometrics assisted spectroscopic methods for analysis of artificial colours in foods are hardly reported.

In chemical methods, chemical standards are used for each test which are very costly. Besides, they need to go through sample preparation process. On the contrary, chemometric techniques with spectroscopic data are non-destructive, simple, cheap and do not generate any chemical waste as no standard chemical is required in this process.

Chemometrics is the science of extracting information from chemical systems by data-driven means. In chemometrics, patterns in the data are modeled and these models are then routinely applied to future data in order to predict the same quality characteristics of a product [13]. One of the important branches of Chemometrics is Classification.

Classification means to assign an individual (sample) to one or more categories based on a set of measurements used to describe or characterize the object itself [14]. It is very often necessary to classify foods on the basis of their raw materials, species of ingredients, brand, geographical origin, category of product etc. to identify their specialty or traceability. This is one the process to identify the authenticity of food. For classification of food, chemometric methods are getting popularity among food analysts and scientists. A recent study has been noticed where multivariate classification techniques were used with UV-Vis spectroscopic data for determining adulteration in sauce [15]. In classification by multivariate methods with spectroscopic instruments, spectral data points are inputs and categories or classes are outputs.

The spectra data acquired from instrument contain spectra background information and noises which are necessary to be reduced. Several de-noising techniques are used in such studies. Multivariate Scatter Correction (MSC), Savitzky-Golay filtering, Standard Normal Variate (SNV), single and double exponentials are very frequently used. Their performance greatly vary when they are used in different types of sample.

There are several methods in chemometrics to classify food for their authenticity, traceability, origin or contained raw materials. Artificial Neural Network (ANN) has got popularity among food scientists and food quality controlling authorities for classification of food. Partial Least Squares-Discriminant Analysis (PLS-DA) is also used for classification in this field of study.

Therefore, the objective of the present work is to develop an easy and cost effective chemometrics assisted method to classify commercial mango juices as adulterated with heavy use of artificial colours or not by using UV spectroscopic data. For this, an appropriate spectroscopic range has been selected first. Then, de-noising performance of three techniques, Multivariate Scatter Correction (MSC), Savitzky-Golay filtering, Standard Normal Variate (SNV) have been assessed and choose the best one. Next, classification performance of two chemometric classification techniques, ANN and PLS-DA have been evaluated, and select the better

one. Finally, with the best alternatives, a chemometrics assisted methods has been proposed for classification of commercial mango juice as adulterated or safe with over use of artificial colours.

2. Materials and Methods

2.1. Preparation of Standard Mixture Solutions

Standard mixture solutions of three artificial colours usually used in mango juice, Tartazine, Sunset Yellow and Beta carotene, were prepared. Nine different concentrations of Tartrazine (0, 0.5, 1, 2, 4, 8, 12, 16, 20 mgL⁻¹), Sunset Yellow (0, 0.5, 1.0, 1.5, 2.0, 2.5, 3, 5, 10 mgL⁻¹) and Beta carotene (0, 0.5, 1.0, 1.5, 2.0, 2.5, 3, 5, 10 mgL⁻¹) were used to prepare synthetic mixture solutions. Here "Orthogonal Experimental Design" was used to statistically maximize the information in the outputs. Thus 81 mixtures were prepared with different concentrations of these three colours. Mixture solutions were used in Ultra-violet (UV) Spectrophotometer, to get spectral data. At last, with known concentration of artificial colours and spectral data of 81 synthetic mixtures, we developed, validated and tested methods to classify commercial mango juice as adulterated with excessive use of these artificial colours.

Colours are pH sensitive. So mixture solutions were prepared by dissolving different concentrations of Tartrazine, Sunset Yellow and Beta Carotene, determined by Orthogonal experimental design, into Acetate Buffer to keep pH level 3.6. Finally, the mixture solutions of artificial colours were run in UV to get their respective spectra.

Buffer solution preparation: Acetate Buffer (0.1 M) was prepared by dissolving 8.2 gm Sodium Acetate in 1 L De-ionized water. pH was adjusted at 3.6 with Acetic acid.

Colour solutions are prepared according to the following steps:

- 1gm of colour was taken and dissolved in 100ml acetate buffer (0.1M) to prepare 10000 mg/L colour stock solution
- 1ml of 10000mg/L colour solution was taken to make 100 ml with acetate buffer and concentration is 100mg/L
- Suggested (considered) volume of colours were taken to make 50ml with acetate buffer to prepare target sample.

2.2. Ultra Violet (UV) Measurements

Ultra violet (UV) spectrophotometer (Analytikjena, AG, 1998-2002, Germany, Model: Specor D 205) connected to software of Win ASPEC (Version 2.2.1.0) was used to obtain UV spectra of samples. Mixture solution samples of artificial colours were placed in the instrument at controlled ambient temperature. UV spectra were collected in frequency 300-700 nm and at an interval of 1nm. All spectra were rationed against a background sample spectrum. There spectra were recorded as absorbance values at each data point in triplicate.

2.3. Pre-Processing of Spectral Data

Spectra pre-processing is one of the main factors affecting model robustness and prediction capabilities with spectroscopic data. Spectral pre-processing techniques are used to remove any irrelevant information which cannot be handled by the regression techniques. Several pre-processing methods have been developed for this purpose. Appropriate spectral range selection and de-noising are two most frequently applied preprocessing techniques used in food authentication studies prior to modeling with spectral data [16].

2.3.1. Spectral Range Selection

All parts of a spectrum are not equally important to predict the considered variable. So it is necessary to find out which portion of the spectrum are contributing more to predict efficiently. This is an trial and error approach as important part not always lays in a particular part of the range, rather it varies sample to sample on the basis of the chemical components and their composition contain in the sample.

When chemometric methods are employed to build calibration models, the absorbing regions of the spectra dictate the number of variables used as input in the multivariate statistical model. The spectral range should include information on the concentration variation of the analyte monitored and other matrix constituents while excluding regions dominated by noise or other artifacts that might be incorporated into the model. Proper choice of the region can also minimize the problem of nonlinearity in the data. Suitable region may be identified by plotting the correlation spectra or by simply plotting the raw spectra and looking for various peaks in them. Spectral region selection and prediction or classification on the basis of these selected regions of spectra are used in many research in chemistry [17-19]. Spectral presentation of UV spectra of artificial colour mixture solutions are shown in figure 1.

From spectra of UV it is clear that absorbance before wavelength 370 (nm) are very highly instrumental noised and above 530 contain least information for prediction as no peak and least variance of absorbance in this part of the spectra. So wave length range 371-533 (nm) has been selected for calibration, validation and test chemometric classification models.

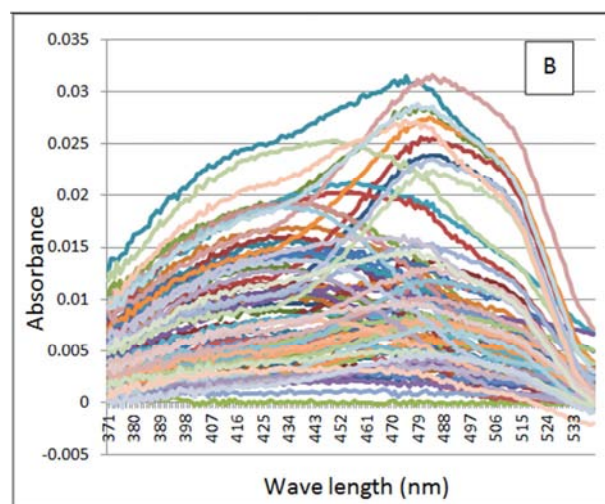
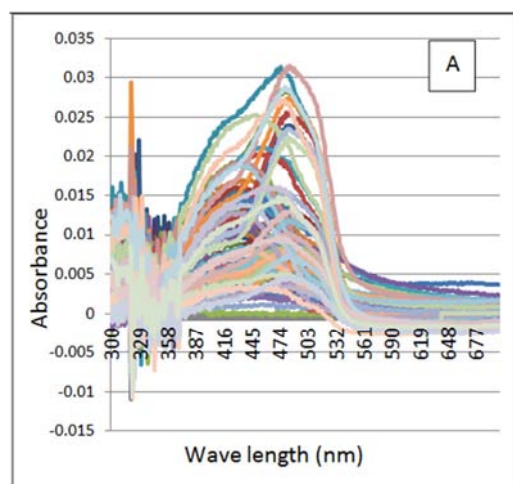


Figure 1. UV Spectra of artificial colour mixtures: (A) Raw spectra, (B) Selected spectra.

2.3.2. De-noising of UV Spectral Data

Among several spectral data de-noising techniques, performance of Standard Normal Variate (SNV), Multiplicative scatter correction (MSC) and Savitzky-Golay algorithm are assessed in the present studies with UV spectral data for artificial colours. Two de-noise efficiency measures such as Signal to Noise Ratio (SNR) and Root Mean Squared Error (RMSE) are used here. A comparative picture of these de-noising techniques are presented in table 1.

Table 1. Comparison of de-noising methods Artificial colour UV data.

De-noising method	SNR	RMSE
Multiplicative scatter correction (MSC)	4.8966	0.0026
Savitzky-Golay filtering	5.2264	0.0025
Standard Normal Variate (SNV)	-43.1346	0.4503

From the above table it is evident that, in terms of both SNR and RMSE, Multiplicative Scatter Correction (MSC) shows the best de-noising performance among these techniques, next is Savitzky-Golay filtering and then Standard Normal Variate (SNV) of UV spectral data. As MSC shows the best de-noising performance, UV spectral data have been de-noised with it before applying them in classification techniques in the study.

2.4. Assigned Classes of Artificial Colours

In the present study, for sake of classification we divide the artificial colour concentrations into two groups (Group 1 and Group 2) which are shown in table 2. The ranges have been selected on the basis of some previous studies [20-22].

Table 2. Assigned classes of artificial colours.

Artificial colours	Groups	Range	Number of samples (%)
Tartrazine	Group 1	< 8 mg/L	35 (43%)
	Group 2	≥ 8 mg/L	46 (57%)
Sunset Yellow	Group 1	< 2 mg/L	44 (54%)
	Group 2	≥ 2 mg/L	37 (46%)
Beta Carotene	Group 1	< 2 mg/L	44 (54%)
	Group 2	≥ 2 mg/L	37 (46%)

2.4. Classification Techniques

2.4.1. Partial Least Squares Discriminant Analysis (PLS-DA)

Partial least squares regression (PLSR) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the independent variable (X) and dependent variable (Y) data are projected to new spaces, the PLS family of methods are known as bilinear factor models. Partial Least Squares-Discriminant Analysis (PLS-DA) is a variant used when the Y is categorical. Instead of calibrating for a continuous variable, one calibrates for group membership (categories). The resulting models are evaluated in terms of their predictive ability to predict the Y variable of new and unknown samples (standard error of prediction, SEP). Discriminant Analysis (DA) and other classification tools are used in several food authentication studies [23-26].

2.4.2. Artificial Neural Networks (ANN)

ANN was initially developed according to the elementary principle of operation of the (human) neural system. Since then, a very large variety of networks have been constructed.

All are composed of units (neurons), and connections between them, which together determine the behaviour of the network. The choice of the network type depends on the problem to be solved; the back-propagation gradient network is the most frequently used [27-29]. This network consists of three or more neuron layers: one input layer, one output layer and at least one hidden layer. In most cases, a network with only one hidden layer is used to restrict calculation time, especially when the results obtained are satisfactory. All the neurons of each layer (except the neurons of the last one) are connected by an axon to each neuron of the next layer.

In practical, Levenberg-Marquardt back-propagation neural network is used in the present study. Number of hidden neurons is 10 and the sigmoid activation function is used in each training.

3. Results and Discussions

3.1. Classification by PLS-DA

In the present study, the PLS-DA is used to develop model by classifying synthetic mixture solutions as heavy or limited existence of artificial colours namely, Tartrazine, Sunset Yellow and Beta Carotene. Classification results by PLS-DA are presented in terms of classification error rate, non-error rate, classification error rate CV and non-error rate CV in table 3.

Table 3. Classification artificial colours concentration by PLS-DA.

Artificial Colours	Dataset	Error rate (%)	Non error rate (%)	Error rate CV	Non error rate CV
Tartrazine	Training (60)	1.0	99.0	0.02	0.98
	Test (21)	6.0	94.0	0.06	0.94
Sunset Yellow	Training (60)	0.0	1.0	0.33	0.67
	Test (21)	18.0	82.0	0.21	0.76
Beta Carotene	Training (60)	0.0	1.0	0.49	0.57
	Test (21)	40.0	60.0	0.67	0.33

From above table it is evident that, for Tartrazine, error rate and error rate CV is less both in training (1% and 0.02 respectively) and for test dataset they are 6% and 6.0. For Sunset Yellow and Beta Carotene performance of PLS-DA with training dataset is very well (0% error), but for test dataset error level is not satisfactory (18% and 40% respectively), and error rate CVs are very high accordingly.

3.2. Classification by ANN

ANN is used for development of model for classification of mango juice as adulterated or safe with over use of Tartrazine, Sunset Yellow and Beta Carotene by using UV spectral data of their synthetic mixture solutions. Classification results of ANN are presented in terms of Root Mean Square Error (RMSE) and Percentage Error of Misclassification (%E) in table 4.

For Tartrazine percentage of misclassification error is 3.51 for training dataset, and for validation and test datasets these are 0% prospectively. For Sunset Yellow, %E is 1.75% for

training data and for validation and test data it is zero percent. Finally, for Beta Carotene, misclassification error percentage is zero for all datasets.

So, a very good classification results by ANN is shown in the above table. Therefore, ANN can be used for classification of commercial mango juice as they have heavy concentration of Tartrazine, Sunset Yellow and synthetic Beta Carotene or not.

Table 4. Classification of artificial colours by ANN.

Artificial Colours	Data sets(Size)	RMSE	(%E)
Tartrazine	Training (57)	0.2079	3.509
	Validation (12)	0.0216	0.000
	Test (12)	0.1570	0.000
Sunset Yellow	Training (57)	0.2023	1.754
	Validation (12)	0.1621	0.000
	Test (12)	0.2949	0.000
Beta Carotene	Training (57)	0.0700	0.000
	Validation (12)	0.0995	0.000
	Test (12)	0.1824	0.000

4. Conclusions

The mixture solutions of artificial colours are classified by using PLS-DA with de-noised UV spectroscopic data. It is evident from the results that in PLS-DA, misclassification errors are high for all colours (6%, 18% and 14% for test data). So, PLS-DA cannot be used for classifying mango juices as adulterated or not with heavy use of artificial colours, Tartrazine, Sunset Yellow and Beta Carotene. On the other hand, in ANN, misclassification errors are very low (0% for validation and test data) which means ANN produces a very good classification results for all datasets; therefore, it can be used for classifying mango juices as adulterated or not with heavy use of artificial colours, Tartrazine, Sunset Yellow and Beta Carotene. It is also evident that results are acceptable for classification based on artificial colours.

One chemometric method has been proposed through this study where UV spectral range 371-533 (nm) has to be taken first. Next, de-noising technique MSC has to be used as it shows better performance than other methods compared here. Finally, for classifications ANN should be used in this method as it shows excellent classification performance.

In the proposed method, no standard chemical is used. So, this method is cost effective, time saving and do not generate any chemical waste. Newly proposed chemometric method could save a huge amount of quality testing cost for mango juice producing companies, quality regulating authorities and food testing laboratories.

References

- [1] Babu S and Shenoliker IS (1995). Health and Nutritional Implications of Food Colors. *Indian Journal of Medical Research*, 102: 245-255.
- [2] Sarah Chapman (2011). Guidelines on approaches to the replacement of Tartrazine, Allura RED, Ponceau 4R, Quinoline Yellow, Sunset Yellow and Carmoisine in food and beverages. https://www.food.gov.uk/sites/default/files/multimedia/pdfs/publication/guide_lines_six_colours.pdf.
- [3] "Food additives". CBC News, 29 September, 2008. <http://www.cbc.ca/news2/background/foodsafety/additives.html>.
- [4] "The additives which could be banned". The Telegraph, 2009. <http://www.telegraph.co.uk/news/health/news/3453522/The-additives-which-could-be-banned.html>.
- [5] Sahar SAS and Manal MEMS (2012). The Effects of Using Color Foods of Children on Immunity Properties and Liver, Kidney on Rats. *Food and Nutrition Sciences*, 3: 897-904.
- [6] Garcia-Falcon MS and Simal-Gandara J (2005). Determination of food dyes in soft drinks containing natural pigments by liquid chromatography with minimal clean-up. *Food Control*, 16: 293-297.
- [7] Chen Q, Mou S, Hou X, Riviello JM and Ni Z (1998). Determination of eight synthetic food colourants in drinks by high performance ion chromatography. *Journal of Chromatography A*, 827: 73-81.
- [8] Chou SS, Lin YH, Cheng CC and Hwang DF (2002). Determination of synthetic colours in soft drinks and confectioneries by micellarelectrokinetic capillary chromatography. *Journal of Food Science*, 67(4): 1314-1318.
- [9] González M, Lobo MG, Méndez J and Carnero A (2005). Detection of colour adulteration in cochineals by spectrophotometric determination of yellow and red pigment groups. *Food Control*, 16(2): 105-112.
- [10] Rong Li, Jiang Z and Liu Y (2008). Direct Solid-phase Spectrophotometric Determination of Tartrazine in Soft Drinks Using β -Cyclodextrin Polymer as Support. *Journal of Food and Drug Analysis*, 16(5): 91-96.
- [11] Ou S, Xiashi Z, Yanli F and Weixing M (2014). Determination of Sunset Yellow and Tartrazine in Food Samples by Combining Ionic Liquid-Based Aqueous Two-Phase System with High Performance Liquid Chromatography. *Journal of Analytical Methods in Chemistry*, 2014: 1-8.
- [12] Bachalla N (2016). Identification of synthetic food colors adulteration by paper chromatography and spectrophotometric methods. *International Archives of Integrated Medicine*, 3(6): 182-191.
- [13] Kirali R and Ferreira MMC (2006). The past, present, and future of chemometrics worldwide: some etymological, linguistic, and bibliometric investigations. *Journal of Chemometrics*, 20: 247-271.
- [14] Rasmus Bro (2013). Chemometrics in Food Chemistry. Elsevier, the Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK. pp: 171-227.
- [15] Carolina VDA, Serena R, Liliana A and Rodríguez MS (2016). UV-Visible Spectroscopy and Multivariate Classification as a Screening Tool for Determining the Adulteration of Sauces. *Food Analytical Methods*, 9(11): 3117-3124.
- [16] Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W and Theron KI (2007). Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, 46: 99-118.
- [17] Paradkar MM, Sivakeasva S and Irudayaraj J (2003). Discrimination and classification of adulterants in maple syrup with the use of infrared spectroscopic techniques. *Journal of the Science of Food and Agriculture*, 83: 714-721.
- [18] Jha SN, Gunasekaran S (2010). Authentication of sweetness of mango juice using Fourier transform infrared-attenuated total reflection spectroscopy. *Journal of Food Engineering*, 101: 337-342.
- [19] Nikbakht M, Tavakkoli HT, Malekfar R and Ghobadian B (2011). Nondestructive Determination of Tomato Fruit Quality Parameters Using Raman Spectroscopy. *Journal of Agricultural Science and Technology*, 13(4): 517-526.
- [20] Dinc E, Baydan E, Kanbur B and Onur F (2002). Spectrophotometric multicomponent determination of sunset yellow, tartrazine and allura red in soft drink powder by double divisor-ratio spectra derivative, inverse least-squares and principal component regression methods. *Talanta*, 58: 579-594.
- [21] Dinc E, Aktas AH, Baleanu D and Ustundag O (2006). Simultaneous Determination of Tartrazine and Allura Red in Commercial Preparation by Chemometric HPLC Method. *Journal of Food and Drug Analysis*, 14 (3): 284-291.

- [22] Karaoglan GK, Gumrukcu G, Ozgur MU, Bozdogan A and Asci B (2007). Abilities of Partial Least-Squares (PLS-2) Multivariate Calibration in the Analysis of Quaternary Mixture of Food Colours (E-110, E-122, E-124, E-131). *Analytical Letters*, 40: 1893-1903.
- [23] Arvantoyannis I, Katsota MN, Psarra P, Soufleros E and Kallinthraka S (1999). Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends Food SciTechnol*, 10: 321-336.
- [24] Buratti S, Bendetti S, Scampicchio M and Pangerod EC (2004). Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue. *Anal ChimActa*, 525: 133-139.
- [25] Cynkar WU, Cozzolino D, Damberg RG, Janik L and Gishen M. (2007). Feasibility study on the use of a head space mass spectrometry electronic nose (MS e-nose) to monitor red wine spoilage induced by *Brettanomyces* yeast. *Sensor Actuat*, 124: 167-171.
- [26] Alexandre M, Lozano J, Gutierrez J, Sayago I, Fernandez MJ, Horrillo MC (2008). Portable e-nose to classify different kinds of wine. *Sensors and Actuators*, 131: 71-76.
- [27] Rumelhart DE, Hinton GE and Williams RJ (1986). Learning representations by back-propagating error. *Nature*, 323: 533-536.
- [28] Kosko B (1992). *Neural Networks for Signal Processing*. Prentice-Hall, London, pp. 399.
- [29] David R, Sovan L, Ioannis D, Jean J and Jacques LSA (1997). Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes*, 40: 35-43.