



ELECTRONIC NOSE FOR ANCHOVY FRESHNESS MONITORING BASED ON SENSOR ARRAY AND PATTERN RECOGNITION METHODS: PRINCIPAL COMPONENTS ANALYSIS, LINEAR DISCRIMINANT ANALYSIS AND SUPPORT VECTOR MACHINE

A. Amari ¹⁾, N. El Bari ²⁾ and B. Bouchikhi ^{1)*}

¹⁾ Sensors, Electronic & Instrumentation Group, Faculty of Sciences, Physics Department, University Moulay Ismaïl, B.P. 11201, Zitoune, Meknes, Morocco

²⁾ Biotechnology, Agroalimentary and Biomedical Analysis Group, Faculty of Sciences, Biology Department, University Moulay Ismaïl, B.P. 11201, Zitoune, Meknes, Morocco

* Corresponding author; E-mail: benachir.bouchikhi1@caramail.com

Abstract: An electronic nose based system, which employs an array of six inexpensive commercial gas sensors based on tin dioxide (Figaro Engineering Inc., Japan), has been used to analyse the freshness states of anchovies. Fresh anchovies were stored in a refrigerator at $4 \pm 1^\circ\text{C}$ over a period of 15 days. Electronic nose measurements need no sample preparation and the results indicated that the spoilage process of anchovies could be followed by using this technique. Conductance responses of volatile compounds produced during storage of anchovy were monitored and the result were analysed by multivariate analysis methods. In this paper principal component analysis (PCA) and linear discriminant analysis (LDA) were used to investigate whether the electronic nose was able to distinguishing among different freshness states (fresh, moderated and non-fresh samples). The loadings analysis was used to identify the sensors responsible for discrimination in the current pattern file. Therefore, the support vector machines (SVM) method was applied to the new subset, with only the selected sensors, to confirm that a subset of a few sensors can be chosen to explain all the variance. The results obtained prove that the electronic nose can discriminate successfully different freshness state using LDA analysis. Some sensors have the highest influence in the current pattern file for electronic nose. Support vector machine (SVM) model, applied to the new subset of sensors show the good performance.

Keywords: Electronic nose; Anchovy spoilage; Principal component analysis; Linear discriminant analysis; Support vector machine.

1. INTRODUCTION

Analysis of odour and flavour in food has traditionally been performed either by a trained sensory panel or by headspace gas chromatography mass spectrometry. These methods are time consuming and costly and there is a need in the food industry for objective automated non-destructive techniques that can characterise odour and flavour in food [1, 2]. New methods should allow a high number of samples to be analysed within a short period of time with a sufficient reproducibility and accuracy. During recent years there has been a rapid development of a concept named "electronic nose" (artificial nose) based on chemical gas-sensor technology which seems to fulfil these requirements [3].

Export of fresh fish from Morocco to the markets in Europe has become increasingly important in recent years. Fish freshness control has received

very much attention in the past years because its relevance in food industry. Fish freshness studies have been performed using different approaches [3, 4]. One of the unique characteristics of fish as food is that it is a highly perishable commodity [5]. Consequently, there is a need for a simple technique to monitor freshness and quality of fish (Anchovies).

Electronic nose (e-nose) is more and more widely used in environmental monitoring, food production and medicine such as odour evaluation [4, 6, 7]. The electronic nose has proven to be a rapid, non-destructive technique for measuring volatile compounds which exhibit spoilage odours in fish [4, 6, 8]. The response of all sensors in the e-nose together constitutes a unique profile that gives the "fingerprint" of odour. The developed system was used to study the influence of several storage days on the freshness of Anchovy from the Moroccan Sea.

The objectives of this study were: (1) to evaluate

the capacity of the electronic nose for monitoring anchovy freshness, using a specific electronic nose device based on a 6-sensor array, (2) to study principal component analysis (PCA) and linear discriminant analysis (LDA) techniques to obtain whether the electronic nose would be able to distinguish different freshness stage, (3) to identify each sensor responsible for a discrimination in the current pattern file, using loading analysis, (4) to apply the support vector machine (SVM) with an array of only the selected sensors, to confirm the good generalisation using the new dataset.

2. EXPERIMENTAL METHOD

A. STORAGE CONDITION AND SAMPLING

The anchovies used in this study were caught from the coast of Mehdiya (Morocco) in July 2006. In order to maintain the fish in a fresh condition, portable cooler containing crushed ice was used to store the anchovies immediately after catch. The length of time from anchovy capture to receipt at the laboratory varied from 4 to 6 h. Freshly samples were, immediately after receiving, placed in plastic boxes (1 fish/box) and introduced in refrigerator kept at a constant temperature of $4 \pm 1^\circ\text{C}$ to be analysed at 1, 3, 5, 7, 9, 11, 13 and 15 days of storage. In total 53 anchovy samples were analysed.

The sampling system consists of a dynamic headspace sampling (Fig. 1). The sensor array comprises six Taguchi gas sensors (TGS) from Figaro Engineering, Japan, a capacitive humidity sensor PHILIPS H1 from Philips and a temperature sensor LM35DZ from National Semiconductor. The TGS sensors were chosen because of their large commercialisation, relatively low price and high sensitivity. The following TGS were applied in this work: TGS 823, TGS 825, TGS 826, TGS 831, TGS 832 and TGS 882. The experimental set-up was shown in Fig. 1. The sensors were chosen on the basis of sensitivity of the sensors to different gases which characterise the spoilage of fish [4, 6, 9]. The sensor array was mounted in a flow type gas chamber equipped with two ports for the gases (inlet and outlet) and a head with wires. The electrical conductance of the sensors varies in the presence of reducing/oxidizing gases. In general, the sensors conductance was found to increase as the anchovy 'age'.

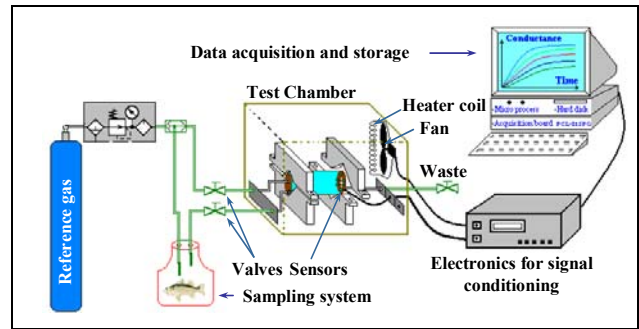


Fig. 1 –The experimental set-up used to analyse the headspace of anchovies.

B. FISH MANIPULATION

An anchovy sample is taken from the refrigerator before every experiment, weighted, placed in a Petri dish for 40 minutes and then introduced into the e-nose chamber. We characterise our sensors array by this repetitive procedure: first, the carrier gas (nitrogen) was allowed to flow through the sensor chamber for 50 minutes so that the sensors could stabilise their baseline resistance. Afterwards, a valve allowed the nitrogen to flow through the sampling vessel and the sensor chamber for another 50 minutes. The response of the sensors was collected and stored every 1 s. After each measurement, the sensor chamber was opened and cleaned with nitrogen gas to regenerate the sensors' baseline. Each time that a new set of anchovies was analysed, new glass vessels were employed.

C. INTERFACE ELECTRONICS

The data acquisition and storage system was controlled using LabVIEW© software (National Instruments Inc.). A PCL-812PG data acquisition card was used for online data gathering. Fig. 2 shows the LabVIEW© PC interface for data acquisition.

Sensors were held at the temperature of 33°C , it was maintained constant with an accuracy of $\pm 1^\circ\text{C}$. During the measurement phase, the computer records the conductance changes that the sensors experienced. When the measurement was completed, the acquired data were properly stored for later use.

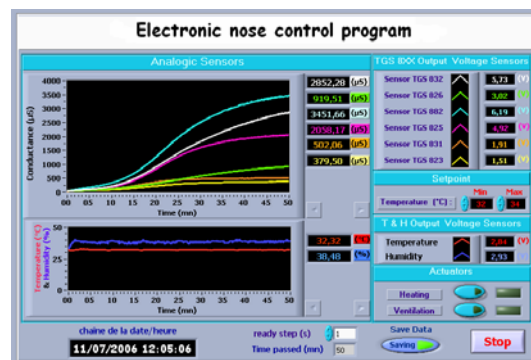


Fig. 2 –Typical Lab VIEW© PC interface for the electronic nose system.

D. PCA, LDA AND LOADING ANALYSIS

Most feature extraction techniques for electronic nose applications have been based on linear techniques, mainly Principal Components Analysis (PCA) and Fisher's Linear Discriminant Analysis (LDA).

Principal Components Analysis:

PCA is a powerful, linear, and 'unsupervised' pattern recognition technique used as a mathematical tool for analyzing, classifying and reducing the dimensionality of numerical datasets in a multi-variate problem [10]. PCA decomposes the primary data matrix by projecting the multi dimensional onto a new coordinate's base formed by the orthogonal directions with data maximum variance. The eigenvectors of the data matrix are called principal components and they are uncorrelated among them. The principal components are ordered so that the first one displays the greatest amount of variance, followed by the next greatest and so on. PCA was very widely used for gas sensors applications [11, 12].

Linear Discriminant Analysis:

LDA provides a linear projection of the data, by taking account of the scatter of data within each class and across classes. Projection directions are those that maximize the inter-class separation of the projected data. The LDA transformation matrix is given by:

$$T_{LDA} = S_{\omega} A_{\omega}^{-1/2} S_B \quad (1)$$

where S_{ω} and A_{ω} are, respectively, the eigenvectors matrix and the diagonal eigenvalues matrix of the within-class scatter ω . S_B is the eigenvectors matrix of the between class scatter B . LDA was previously used for gas detection application [13].

The difference between PCA and LDA is that PCA does not consider the relation of a data point to the specified classes, while the LDA calculation uses the class information that was given during training. The main features of both PCA and LDA analysis are the coordinates of the data in the new base (scores plot) and the contribution to each component of the sensors (loads plot). The scores plot is usually used for studying the classification of the data clusters; while the loads plot can be used for giving information on the relative importance of the sensors to each principal component and their mutual correlation.

The loadings analysis is well correlated to the LDA. Using this analysis, the sensors can be investigated for their responsibility for the discrimination given by the trained patterns.

Sensors, located near to zero, have a minor responsibility for the distribution of pattern in the LDA plot. They may be switched off because they may have a negative influence on the pattern resolution, when particular normalisations are selected. The loadings analysis helps to identify the sensors responsible for discrimination in the current pattern file [13-15].

E. SUPPORT VECTOR MACHINE (SVM)

Support vector machine (SVM) is a statistical classification method proposed by Vapnik in 1995 [16]. The main idea of SVM is to separate the classes with a particular hyperplane, which maximises a quantity called margin. The margin is the distance from a hyperplane separating the classes to the nearest point in the dataset [6, 17].

SVM were originally designed for binary classification. Currently there are two types of approaches for multi-class SVM. One is by constructing and combining several binary classifiers "one against one or one against all methods", while the other is by directly considering all data in one optimisation formulation [18]. The direct approach consists to construct a classifier recognising the set of the classes: the determination of the hyperplane between these different classes permits to choose a class among the K when a new input is presented.

In order to check whether the SVM algorithm can provide a good identification with only the selected sensors, we have applied the direct approach that is considered a suitable method for practical use. This algorithm is clearly described by J. Weston in [19]. The kernel function used in this study was an order 3 polynomial with an optimal value of the penalty coefficient $C = 2^{17}$.

3. RESULTS AND DISCUSSION

A. ELECTRONIC NOSE RESPONSES AND SIGNAL ANALYSIS

Fig. 3 shows the evolution of the signals generated by the 882 TGS sensor during measurement of anchovy for all storage days; we show the same behaviour for all the other sensors. Each line represents the average signal variation of all anchovies measured in the same day respectively for each sensor of the array. It can be shown that the sensor signals reach a plateau. To analyse the response of the electronic nose, three parameters were extracted from each sensor conductance response: The initial conductance, dynamic conductance calculated within the interval [15 - 35 min] and the steady-state conductance [6, 9].

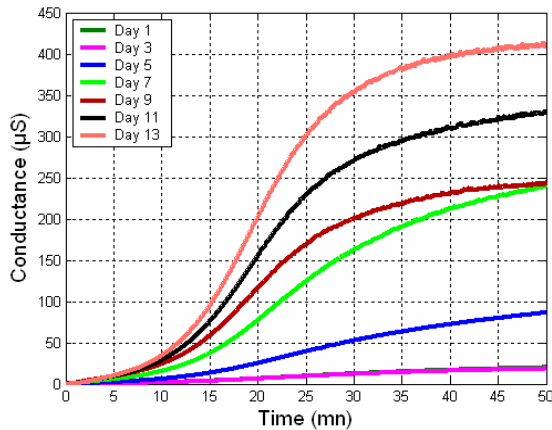


Fig. 3 –Time conductance evolution of the TGS 882 Sensor during all storage days.

Fig. 4 shows the evolution of the steady-state conductance. It can be shown that this characteristic parameter dramatically increases with storage days. This indicates that the concentration of volatile compounds, which characterise the spoilage odour of fish, increases with the number of storage days [4, 6, 9]. It can be inferred that the 825, 832 and 882 TGS sensors have higher values, which may implied that those are important on the current pattern file and evaluated the picking-date (this is presented in paragraph C).

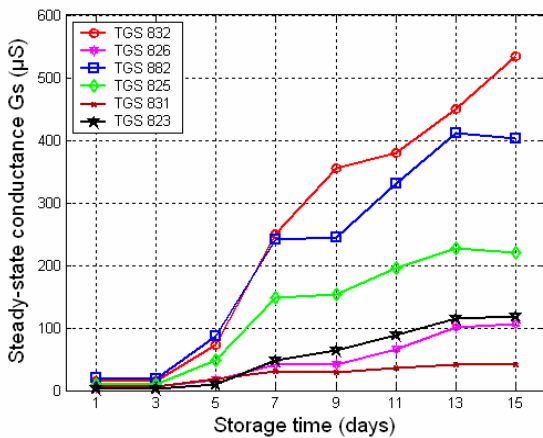


Fig. 4 –The steady-state conductance evolution during all storage days.

B. CLASSIFICATION OF ANCHOVY USING PCA AND LDA

In order to investigate whether the electronic nose was able to distinguish between different freshness state, PCA and LDA analysis were applied to 53 samples. Figs. 5 and 6 show the analysis results on a two-dimensional plane, principal component 1 (PC1) and principal component 2 (PC2) in Fig. 5 and first and second linear discriminant LD1 and LD2 in Fig. 6.

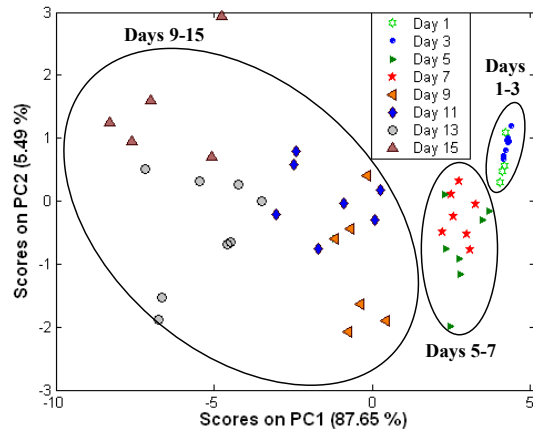


Fig. 5 –PCA analysis in the PC1–PC2 plane.

With PCA, the two first principal components allow us to well represent 93.14 % of the information in the database. We obtain a representation of eight storage days: 1, 3, 5, 7, 9, 11, 13 and day 15. The results obtained indicate that the sets of the measures are organized according to three classes: the class 1; from day 1 to day 3, the class 2; from day 5 to day 7 and the class 3; from day 9 to day 15. These three classes are related to three states of freshness of the anchovies: fresh, medium and aged.

The LDA analysis was applied to the same dataset, and it showed a very clear discrimination among the various classes representing different anchovy freshness state (Fig. 6). In this plot, about 99.08 % of the total variance of the data is displayed. LDA function 1 (LD1) and function 2 (LD2) accounted for 95.25 and 3.83 % of the variance, respectively.

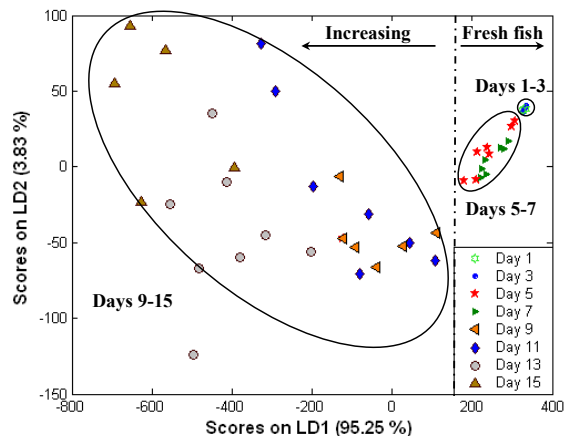


Fig. 6 –LDA analysis in the LD1–LD2 plane.

Using PCA analysis, it is possible to classify the fish into three freshness states. When the electronic nose was performed with LDA, better classification rates were observed.

As a supervised classification method, we have applied the PNN approach, which gave a success-rate in classification of 91.07 % [20].

C. LOADINGS PLOT ANALYSIS

Using only the steady-state conductance change, the so-called loading analysis was performed to identify the important of sensors responsible for discrimination in the current pattern file. Single sensors may be switched off for analysis if they have rather smaller influence on the identification process. Sensors with loading parameters near to zero for a particular linear discriminant have a low contribution to the total response of the array, whereas high values indicates a discriminating sensor. Fig. 7 shows a loading plot of the loading factors associate to LD1, LD2 and LD3 for anchovy. Fig. 8 shows the projections of the experimental results on a two-dimensional (2D) plane LD1-LD2. The figures show the relative importance of the sensors in the array.

It was shown that the 825, 832 and 882 TGS sensors have higher influence in the current pattern file, while the 823, 826 and 831 TGS sensors have low influence. This is identical with the result showed in Fig. 4.

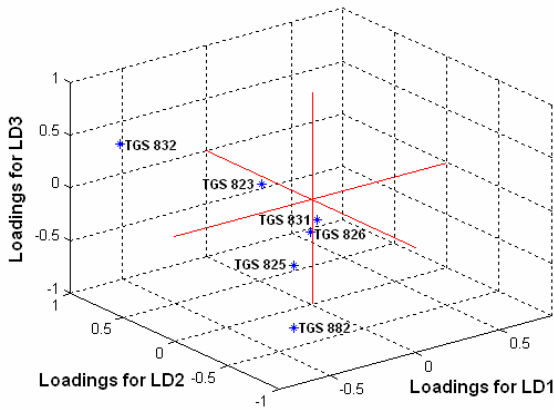


Fig. 7 –Loadings analysis related to LD1, LD2 and LD3 for anchovy.

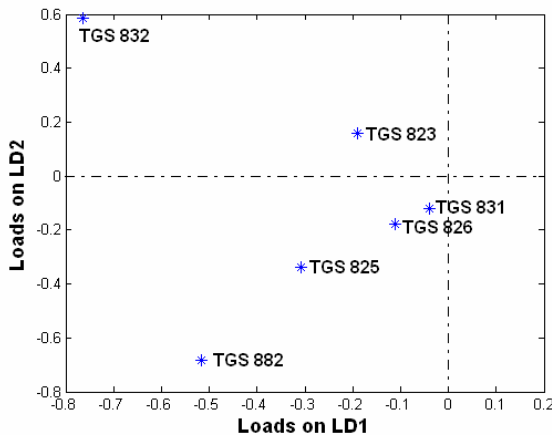


Fig. 8 –Loadings analysis projection related to LD1 and LD2 .

D. VALIDATION ANALYSIS USING SVM

The classification of the tested anchovies has been tackled with a pattern recognition method based on an artificial neuronal network providing nonlinearity in the multivariate classification performance.

The purpose of this study was to evaluate the performance of the SVM using only the responses of the optimized sensors, 825, 832 and 882 TGS sensors, which have a high contribution to the total response of the array. SVM was used to sort anchovy samples in three classes: fresh, medium and aged samples (Fig.6).

Table 1. Confusion matrix for SVM in the classification of anchovy samples according to three classes . Rows and columns are for true and predicted values, respectively.

SVM	fresh	medium	aged
fresh	12	0	0
medium	0	15	1
aged	0	2	24

Table 1 summarizes the confusion matrix that corresponds to the SVM classifier. It gives a good generalisation with a success-rate in identification of 94.34 %, which was estimated using the leave-one-out (LOO) method [21].

4. CONCLUSION

The developed electronic nose system allows determining the quality of anchovy fish samples. The experimental approach has an excellent potentiality for applications, such as in situ quality control of fish freshness at fish markets. In fact, the results show the practical use of a very simple instrument for real applications in the food sector. The freshness of anchovy depends on storage time and storage condition. Slower spoilage rate reflected by lower conductance of sensor’s responses

Principal component analysis and linear discriminant analysis were used to investigate whether the electronic nose was able to distinguishing among different freshness states. The electronic nose was able to detect a clear difference in volatile profile on the fish freshness using linear discriminant analysis than using Principal component analysis.

The loadings analysis was used to identify the sensors responsible for discrimination in the current pattern file. The 825, 832 and 882 TGS sensors in the electronic nose have the highest influence in the current pattern file. Hence, nearly a subset of few sensors can be chosen to explain all the variance. To confirm that, the SVM approach was applied to the new subset. The results showed a good performance

of the support vector machine model, applied to the new subset, with a classification success rate of 94.34 %

5. REFERENCES

- [1] G. Olafsdottir, E. Martinsdottir, J. Oehlenschlager, P. Dalgaard, B. Jensen, I. Undeland, I. Mackie, G. Henehen, J. Nielsen, H. Nilsen, "Methods to evaluate fish freshness in research and industry," *Trend Food Sci. technol.* (8) (1997). p. 258–265
- [2] N. Kocsis, M. Amtmann, Z. Mednyanszky, K. Korany, "GC-MS investigation of the aroma compounds of hungarian red paprika (*Capsicum annum*) cultivars," *Journal of Food Composition and Analysis* (15) (2002). p. 195–203
- [3] S. Garrigues, T. Talou, D. Nesa, "Comparative study between gas sensors arrays device, sensory evaluation and GC/MS analysis for QC in automotive industry," *Sens. Actuators B* (103) (2004). p. 55–68
- [4] M. O'Connell, G. Valdora, G. Peltzer, R.M. Negri, "A practical approach for fish freshness determinations using a portable electronic nose," *Sens. Actuators B* (80) (2001). p.149–154
- [5] J. Hammond, B. Marquis, R. Michaels, B. Oickle, B. Segee, J. Vetelino, A. Bushway, M.-E. Camire, K.-D. Dentici, A "semiconducting metal-oxide array for monitoring fish freshness," *Sens. Actuators B*, (84) (2002). p. 113–122
- [6] A. Amari, N. El Barbri, E. Llobet, N. El Bari, X. Correig, B. Bouchikhi, "Monitoring the freshness of Moroccan sardines with a neural-network based electronic nose," *Sensors* (6) (2006). p. 1209–1223
- [7] A.-C. Romain, J. Nicolas, P. Andre, "In situ measurements of olfactive pollution with inorganic semiconductors: limitations due to the influence of humidity and temperature," *Seminars Food Anal.* (2) (1997). p. 283–296
- [8] G. Olafsdottir, P. Nesvadba, C. Di Natale, M. Careche, J. Oehlenschlager, S. V. Tryggvadottir, R. Schubring, M. Kroeger, K. Heia, M. Esaiassen, A. Macagnano, B. M. Jorgensen, "Multisensor for fish quality determination," *Food Science and technology* (15) (2004) p. 86–93
- [9] N. El Barbri, A. Amari, M. Vinaixa, B. Bouchikhi, X. Correig, and E. Llobet, "Building of a metal oxide gas sensor-based electronic nose to assess the freshness of sardines under cold storage", *Sensors and Actuators B: Chem.* (2007), doi:10.1016/j.snb.2007.06.007, accepted
- [10] J. Brezmes, E. Llobet, X. Vilanova, J. Orts, G. Saiz, X. Correig, "Correlation between electronic nose signals and fruit quality indicators on shelf-life measurements with pinklady apples," *Sens. Actuator B* (80) (2001). pp. 41–50
- [11] M. Penza, G. Cassano, "Application of principal component analysis and artificial neural networks to recognize the individual VOCs of methanol/2-propanol in a binary mixture by SAW multi-sensor array," *Sens. Actuator B* (89) (2003). p.269–284
- [12] A.M. Taurino, C. Distanto, P. Sicilano, L. Vasanelli, "Quantitative and qualitative analysis of VOCs mixtures by means of microsensors array and different evaluation methods," *Sens. Actuators B Chem.* (93) (2003). p. 117–125
- [13] A.H. Gomez, J. Wang, G.X Hu, A.G. Pereira, "Electronic nose technique potential monitoring mandarin maturity," *Sens. Actuator B* (113) (2006). p. 347–353
- [14] J. Brezmes, B. Ferreras, E. Llobet, X. Vilanova, X. Correig, "Neural network based electronic nose for the classification of aromatic species," *Analytica Chimica Acta*, (348) (1997). p. 503–509
- [15] E. Llobet, J. Brezmes, X. Vilanova, J.E Sueiras, X. Correig, "Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick film tin oxide gas sensor array," *Sens. Actuator B*, (41) (1997). p. 13–21
- [16] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery* 2 (1998). p. 121–167
- [17] O. Gualdrón, J. Brezmes, E. Llobet, A. Amari, X. Vilanova, B. Bouchikhi, X. Correig, "Variable selection for support vector machine based multisensors systems," *Sens. Actuator B*, (122) (2007). p. 259–268
- [18] C.-W. Hsu, C.-J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Transactions on Neural Networks*, (13) (2002). p. 415–425
- [19] J. Weston, C. Watkins, "Multi-class support vector machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London. UK, 1998.
- [20] A. Amari, E. Llobet, N. El Bari, X. Correig, B. Bouchikhi, Application of an 'Electronic Nose' device to monitor the spoilage process of anchovies, *Proceedings of the Workshop "International Symposium on Olfaction and Electronic Nose, (ISOEN'2007)"*, St. Petersburg, Russia, 3-5 May 2007, pp.60-61.
- [21] R.O. Duda, P.E. Hart, D.G. Store, *Pattern Classification*, John Wiley & Sons, 2001.



Aziz Amari, received his master degree from the University of Moulay Ismail, Morocco in 2004. He is currently a PhD student at the Sensors, Electronic and Instrumentation Group, University Moulay Ismail, Faculty of Sciences (Meknes, Morocco). His

research interests include gas sensors and electronic noses for food analysis.



Nazha El Bari, received her PhD degree in 1989 from the University of Nancy I (France). She joined the University of Moulay Ismail in 1990, and she was awarded a Doctor of Sciences degree in 1995 in this university. She has been Professor since 1999, and she created a

Biotechnology research group in 2005. Her research interests include quality control of milk and their products.



Benachir Bouchikhi, received his PhD degree in 1982 from the University of Aix Marseille (France). He joined the University of Nancy I in the same year, and he awarded a Doctor of Sciences degree in 1988. In 1994, he has been Professor at the Faculty of Sciences of

Moulay Ismail University (Morocco), and he created a Sensors, Electronic and Instrumentation Group in 2005. His current major research interests involve electronic nose and characterisation of thin film chemical sensors.