Chilean Wines Classification based only on Aroma Information

Nicolás H. Beltrán, Manuel A. Duarte-Mermoud, Víctor A. Soto, Sebastián A. Salah and Matías A. Bustos

***Abstract*—**Results of Chilean wine classification based on the aroma information provided by an electronic nose are reported in this paper. The classification scheme consists of two parts; in the first stage, Principal Component Analysis is used as a feature extraction method to reduce the dimensionality of the original information, while in the second stage, Radial Basis Functions Neural Networks is used as pattern recognition technique to perform the classification. This study is aimed to classify wine samples from different years, valleys and vineyards of Chile, into one of the classes Cabernet Sauvignon, Merlot or Carménère.

***Keywords*—**Feature extraction techniques, Pattern recognition techniques, Principal component analysis, Radial basis functions neural networks, Wine classification

I. Introduction

DURING the last decade several authors have reported results concerning wine classification using information supplied by an electronic nose.

In [1] an aromatic classification of three wines of the same variety but different years (1995, 1996 and 1997) is presented. The input data for classification is obtained from an electronic nose [2] based on six sensors of conducting polymers. For classification purposes a Multilayer_Perceptron (MLP) trained with the backpropagation algorithm (BP) [3] and a Time Delay Neural Networks (TDNN) trained with the Levenberg-Marquadt algorithm [3] were used.

In [4] wine classification is done using a NN with data provided by an electronic nose built by the authors using sensors commercially available. These sensors are of tin oxide and use the principle of resistance variations due to the adsorption of gas molecules on its surface.

Recently, in [5] an electronic nose based on metal oxide semiconductor thin-film sensors has been used to characterize and classify four types of Spanish red wines of the same variety of grapes. Principal component analysis (PCA) and probabilistic neuronal network (PNN) were used to classify the wines.

This work was supported by CONYCIT- Chile, under the grant FONDEF D01-1016, "Chilean Red Wine Classification by means of Intelligent Instrumentation".

Authors are with the Electrical Engineering Department, University of Chile. Av. Tupper 2007, Casilla 412-3, Santiago, Chile.

Corresponding author Manuel A. Duarte-Mermoud, phone: (56-2) 978-4213; fax: (56-2) 672-0162; e-mail: mduartem@cec.uchile.cl .

In this paper a wine classification methodology based on a neural networks technique is proposed. We use as input data, gas chromatograms of Volatile Organic Components (VOC) of Chilean red wines samples belonging to the varieties Cabernet Sauvignon, Merlot and Carménère. The gas chromatograms are provided by an electronic nose Fast GC Analyzer 7100 supplied by Electronic Sensor Technology [13]. The sensor used by the instrument to detect the VOC is of the type surface acoustic wave (SAW).

The first stage of the proposed methodology is concerned with the dimension reduction of the patterns preserving the original information from classification viewpoint. This is done using the feature extraction method Principal Component Analysis (PCA) [6]. Once the dimension of the input data has been reduced the information is processed in a classification stage based on a radial basis functions neural networks (RBFNN) [6, 7] technique. The general scheme of the proposed methodology is shown in Figure 1.

In Section II a brief description of the feature extraction and the classification techniques used in this study are presented. Section III is devoted to explain the methodology used in this work. The results obtained together with a discussion are presented in Section IV. Finally some conclusions are drawn in Section V.

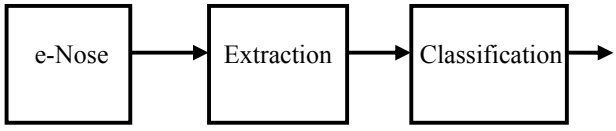

Fig. 1 Block diagram of the methodology used in this work for wine classification

II. Brief Description of the Feature Extraction and Classification Techniques

In this section we present a brief description of the feature extraction and classification (pattern recognition) techniques used in this study. The description of each technique is made only for completeness and for a more detailed description the reader is referred to the cited references.

A. Principal Components Analysis (PCA)

Principal Components Analysis (PCA) technique points to transform the original feature space into one in which the data is not correlated. This new space is obtained by projecting the original data onto a set of orthogonal axis in which the variance of the input data is maximized. This technique is

based on the following theorem of the principal component analysis [6].

Fundamental Theorem for PCA

Given a set of variables $x_i \in \Re^m$ for i=1,2,...n with a non singular covariance matrix $\Sigma_x \in \Re^{m \times m}$, it is always possible to define a subset of non correlated variables $y_i \; y_i \in \Re^p$ for i=1,2,…n, by means of a linear transformation W, corresponding to a rigid rotation, whose columns are eigen-vectors of Σ_x. The covariance matrix of the new set of variables $\Sigma_y \in \Re^{p \times p}$, is diagonal and contains the eigen-values λ_i for i = 1,2 … of Σ_x associated to the eigen-vectors which are columns of W.

Form the previous theorem the eigen-values λ_i can be seen as the variance of the patterns in the transformed space, which are related to the range of the patterns of each axis of this space. On the other hand, the eigen-vectors ϕ_i associated to the eigen-values λ_i, determine the direction of the axes of maximum variance. Thus selecting a subset of eigen-vectors a rotation will be performed that align the transformed axes with the direction of the maximum variance of data. The dimension reduction will be determined by the size of the set of eigen-vectors chosen.

In Section III E a detailed explanation is presented, as to how the PCA methodology is used in this particular case to reduce the dimension of the input data to the classifier.

B. Radial-Basis Functions Neural Networks (RBFNN)

Pattern recognition techniques based on neural networks have shown a great behavior for a wide range of applications [6, 8] and they are very attractive since a minimum knowledge on the patterns is required. The radial-basis functions neural networks (RBFNN) constitute the main alternative to the multi-layer perceptron (MLP) for data interpolation and pattern classification problems. They are characterized by using functions with symmetry around a center *c* in the *n*-dimensional space of the input patterns instead of using a linear activation function.

Each neuron in a RBFNN corresponds to a region in the *n* dimensional input space with center *c*. The activation level of a neuron in a RBFNN to an input *x* is a function of the Euclidean distance between *x* and the center *c* of the neuron. The output of the neuron in a RBFNN is given by the general equation

$$y_k(x) = \sum_{j=0}^{M} w_{kj} \phi_j(x) \tag{1}$$

where $\phi_j(\cdot)$ are the RBF and w_{kj} are the weights in the output layer (See Figure 3). The basis functions $\phi_j(x)$ can be interpreted as the a posteriori probabilities $p(j/x)$ indicating the presence of certain characteristics in the input space. Similarly, the weights w_{kj} can be interpreted as the a posteriori probabilities $p(C_k / j)$ of the members of a class given certain input characteristics. That is the reason why it is natural to apply RBFNN to pattern classification problems [8]. The basic unit of a RBFNN is shown in Figure 2.

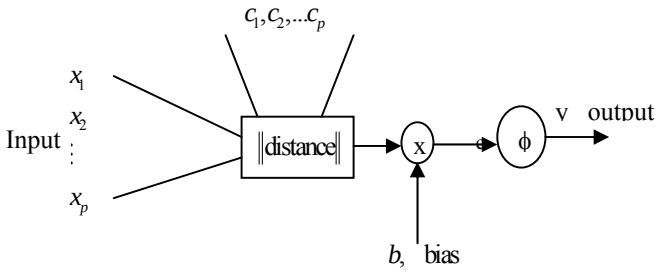

Fig. 2 Mathematical model of a neuron in a RBFNN

In this study basis functions $\phi(x)$ for one neuron is of Gaussian type

$$\phi(x) = \exp\left(\frac{\|x - c\|^2}{2\sigma^2} \right) \tag{2}$$

where *c* determines the center and parameter σ determine the size of the receptive field . σ is also known as *spread* and defines the selectivity of the neuron as

$$selectivity = \frac{1}{\sigma} \tag{3}$$

A small σ implies a high selectivity whereas a large value of σ makes the neuron less selective.

A RBFNN is composed of two layers one containing the RBF and a linear output layer summing the multiplication of the outputs of the RBF with the vector of weights , as shown in Figure 3.

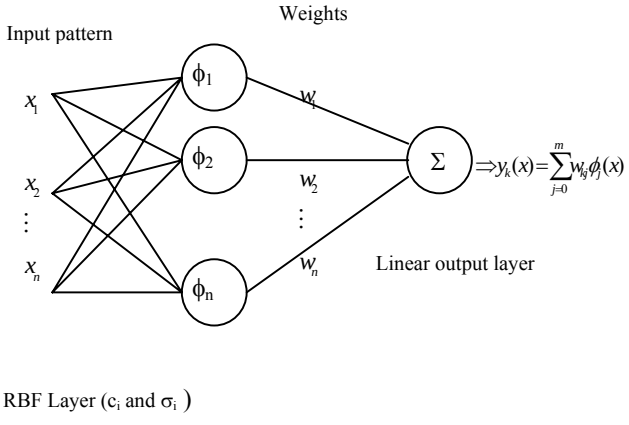

Fig. 3 Two layer architecture of a RBFNN

Then it is necessary to define the *spread* σ and the centers c of the neurons forming the receptive fields of the network. The usual way is to set the center *c* at each one of the training patterns of the problem. Thus, if we have *p* training patterns

the network has p neurons centered at each pattern. This strategy guarantee zero error in the training set and the freedom to choose σ that generates a controlled spatial overlapping to guarantee a good generalization. Depending on the computational implementation utilized, σ can be equal for all neurons or have different values for each unit.

The next step is to choose the weight vector $w \in \Re^m$. To this extent the RBFNN is evaluated at the p training patterns

$$\phi_{ji} = \phi(\|x_i - x_j\|) \quad \forall i.j = 1...m \tag{4}$$

where || || corresponds to the Euclidean norm between two vectors.

We define matrix Φ composed by all the ϕ_{ji}, as the *interpolation Matrix* of the problem [9], from which the weights can be obtained through the relationship

$$\Phi w = T \tag{5}$$

where $w \in \Re^m$ is the weight vector and $T \in \Re^m$ is the objective vector (*target)* containing the desired outputs.

Then if Φ is nonsingular the weights are obtained as

$$w = \Phi^{-1} T \tag{6}$$

The Michelli`s Theorem [9] guarantee that if all vectors x_i used to compute Φ are all different, then Φ will be nonsingular.

III. Experimental Setup

A. Electronic Nose

The electronic nose used in the study is the Fast GC Analyzer 7100 built by Electronic Sensor Technology [13] with surface acoustic wave (SAW) sensor. The most important operating parameters for the electronic nose are:

Sensor : Temperature of the SAW detector in ºC
Column : Temperature of the column GC in ºC
Valve : Temperature of the six position valve in ºC
Inlet : Temperature of the input gas in ºC
Trap : Temperature of the trap in ºC
Ramp : Value of the temperature ramp in ºC/sec
Acquisition Time: Data acquisition time in sec
Sampling Period : Rate at which the information is registered in sec

This set of parameters defines the method under which operates the instrument. After a series of tests and experiments it was determined that the best values of the parameters for our study are those shown in Table I. These values were set on the instrument to perform all the analyses of the wine samples to generate the databases used in this study.

Commercial Chilean wines samples, belonging to the varieties Cabernet Sauvignon, Merlot and Carménère were analyzed. Data was obtained from 40 [ml] wine samples that were introduced into a 60 [ml] vials with septa caps that avoided the contact with oxygen from the environment. The measurements were done immediately after the bottle was opened, maintaining the room temperature at 20 ºC. Figure 4 shows a photograph of the electronic nose during the measurement of a wine sample.

TABLE I
OPERATION PARAMETERS FOR THE ELECTRONIC NOSE

Parameter	**Value**	***Units***
Sensor	60	*ºC*
Column	40	*ºC*
Valve	140	*ºC*
Inlet	175	*ºC*
Trap	300	*ºC*
Ramp	10	*ºC/s*
Acquisition time	20	*s*
Sampling rate	0,01	*s*

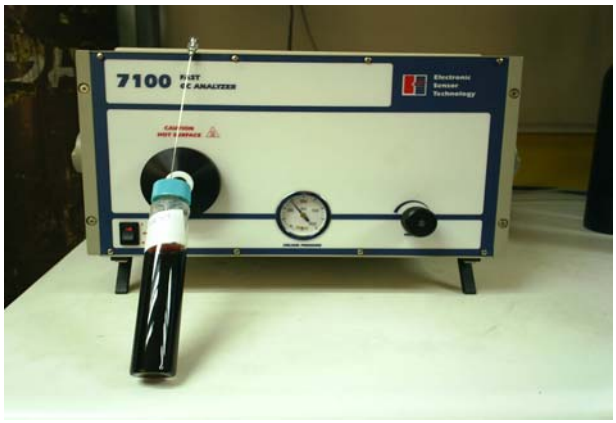

Fig. 4 Electronic nose model *Fast GC Analyzer 7100* from Electronic Sensor Technology

B. Database

The database used in the study is formed by 100 commercial samples of Chilean wines of the type Cabernet Sauvignon, Merlot and Carménère. These wines belong to the vintages 1997-2003 and come from different valleys of the central part of Chile. The distribution of the samples is shown in Table II.

TABLE II
DISTRIBUTION OF WINE SAMPLES

Class	Type	Number	Percentage
1	Cabernet Sauvignon	36	36%
2	Merlot	44	44%
3	Carménère	20	20%
	Total	**100**	**100**

The information from each sample was obtained setting the parameters shown in Table I for the instrument. 10 measurements were carried out for each one of the 100 samples obtaining in total 1000 profiles (chromatograms).

C. Data Pre-Processing

The chromatograms obtained from the electronic nose are curves of 12 sec. of duration with a sampling period of 0.02 sec., containing 600 points in total. Figure 5 shows a typical profile corresponding to a Cabernet Sauvignon sample properly normalized.

Normalization of each chromatogram was performed to normalize the signal amplitudes in the interval [-1,1] and in this way minimize differences coming from sample temperature, column lifetime, and sensor load. To this extent the maximum amplitude was used to normalize the signal according to the following relationship:

$$x_i' = \frac{x_i}{x_{\max}} \tag{7}$$

where x_{max} is the maximum amplitude of all profiles.

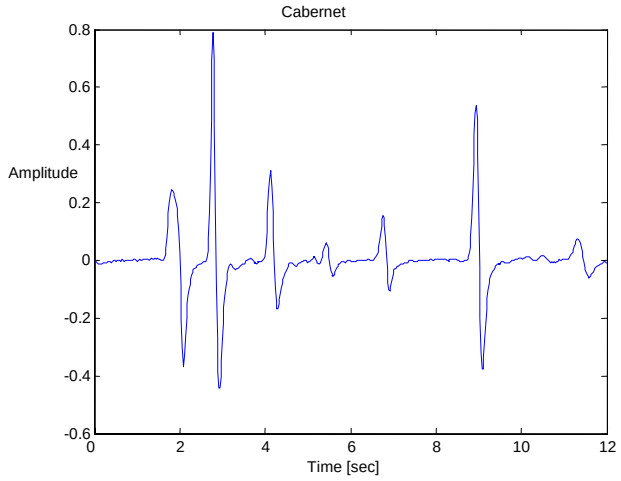

Fig. 5 Typical normalized chromatogram for a Cabernet Sauvignon (600 points and 12 s of duration)

D. Methodology

Profiles classification was done by using the classification technique Radial Basis Functions Neural Networks (RBFNN) described in Section II B. Because of the high data dimensionality, previous to the classification it was necessary to perform a feature extraction procedure of the original data using the Principal Component Analysis (PCA) technique described in Section II A.

The total database of 1000 profiles (360 Cabernet Sauvignon or Class 1, 440 Merlot or Class 2 and 200 Carménère or Class 3) was divided in two sets; one for training-validation (containing the 90 % of the samples) and other for Test (containing the 10% of the samples). The sample distribution is the following:

Training-Validation Set: 900 profiles corresponding to 90 wine samples, 330 profiles Cabernet Sauvignon (33 samples), 390 profiles Merlot (39 samples), and 180 profiles Carménère (18 samples).

Test Set: 100 profiles corresponding to 10 wine samples, 30 profiles Cabernet Sauvignon (3 samples), 50 profiles Merlot (5 samples), and 20 profiles Carménère (2 samples).

The samples for each set were randomly selected and based on the proportion of the samples of different kind contained in the original database.

As a measure of the behavior and to obtain the optimal values of the parameters for each method, Cross-Validation was used [8, 10, 11]. The database was divided into *n* sets, using *n-1* for training and the reminder for validation. The process is repeated *n* times so that all *n* sets are used once for validation.

In the training-validation process it is used cross-validation with the aim to measure the behavior and to tune the optimal parameters for each classification feature extraction method. Then each classifier is evaluated with the *Test* set, using the whole training-validation set and the optimal parameters determined by cross validation. It should be pointed out that the Test Set is never used in the training stage and therefore it is completely unknown to the classifier what is a good performance measurement of each method.

Since there are 10 profiles for each wine sample, the size of cross validation sets is 10 and then the training-validation base will be divided into 90 sub-sets of 10 elements, each one representing one wine sample. Thus, for each method the training is done using 890 profiles and one simulation for validation having 10 elements. The process is repeated 90 times so that each subset of 10 elements is used once to validate the method. The measure of the behavior will be the average and the standard deviation of the percentage of correct classification in validation

Finally, once cross validation is done and the optimal parameters are found for each method, one simulation is carried out with the Test Set to evaluate the performance of each method when unknown samples are presented. The behavior will be measured again in terms of the average and the standard deviation of the percentage of correct classification in the test set.

E. Feature Extraction using Principal Components Analysis (PCA)

The central idea of the PCA is to transform the input space of the variables *P* onto a space *P'* where the data is not correlated i.e. the variance of the data is maximum. This is achieved by computing the eigen-values and the eigen-vectors of the of covariance matrix of the initial data and selecting those eigenvectors that have the largest eigen-values. These components represent the axes of the new transformed space. By projecting the initial data onto these axes the largest data variance is obtained.

The profiles can be seen as characteristic vectors belonging to $\Re^{600}$ and the database as a matrix of *600* x *1000,* where the 1000 columns correspond to each profile and the 600 rows to the points that are going to be reduced.

Considering the training-validation set we have a matrix of *600* x *900* (900 profiles (columns) of 600 points (rows)), then

the covariance matrix of the training-validation set is

$$\Sigma_X = X * X^T \quad (8)$$

with X the training-validation matrix and Σ_X the covariance matrix of X of *600* x *600*. Then computing the eigen-values and the eigenvectors of X and selecting the eigen-vectors with the largest eigen-values the principal components transformation matrix will be determined. One way of choosing the eigen-values (and the eigenvectors associated) is considering the contribution to the global variance [8, 12] of each eigen-value, γ_i, as:

$$\gamma_i = \lambda_i / \sum_{j=1}^{N} \lambda_j \quad (9)$$

being N =*600* the total number of eigen-values of the covariance matrix *X*.

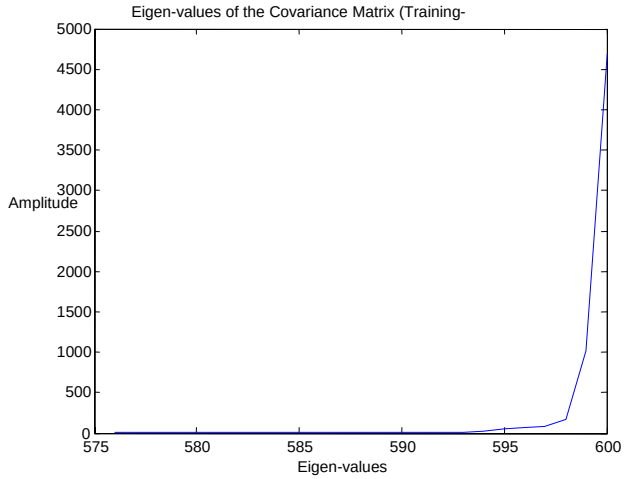

Fig. 6 Detail of the last 25 eigen-values of the training-validation covariance matrix

TABLE III
CLASSIFICATION RESULTS USING PCA AND RBFNN, OBTAINED IN VALIDATION AND TEST SETS, EMPLOYING 20 PRINCIPAL COMPONENTS

Selectivity	% Correct Classification in validation	Standard Deviation	% Correct Classification in test
10	39.8	0.4913	30
1	36.6	0,4845	50
0,1	60.8	0,4402	52
0,02	35.3	0,3557	63
0,01	53.5	0,3846	67
0,0078125	61.3	0,3641	65
0,00390625	66.1	0,3853	76
0,00195313	71.4	0,3776	60

γ_i associates to each eigen-value (and each eigen-vector or principal component) a factor of relative importance considering its contribution to the total variance.

When computing the eigen-values of the covariance matrix *X*, they are ordered in ascending order [11,19], thus the last components are those contributing the most to the information (in terms of the covariance) whereas the first can be considered as noise and therefore disregarded. In Figure 6 are plotted the last 25 eigen-values (of a total of 600) of the training-validation covariance matrix of *600* x *600*.

Figure 6 shows only the last 25 eigen-values of the values training-validation covariance matrix. It is clearly illustrated that practically all the information in terms of the covariance are kept in the last 25 eigen-values.. When computing the contribution of the last 20 eigen-values to the global covariance using (9), the contribution to the total information is 99.87% and the last 10 eigen-values contribute 99.46%. Therefore it was chosen the matrix transformation composed by the 20 eigen-vectors associated to the last 20 eigen-values, generating a *600*x*20* matrix (the 600 rows represent the initial characteristics or points and the 20 columns the eigen-vectors or new characteristics). Multiplying each original profile by the transformation matrix a low dimension profile is obtained (dimension 20) which is used in the classification procedure.

IV. RESULTS

As explained in Section II B, in a RBFNN is necessary to define the centers of the neurons and the parameters σ known as *spread,* which define the selectivity of the neuron. For all simulations the neurons were located at each training pattern [28], thus when cross validation is carried out the networks has 890 neurons corresponding to each profile. Recall that the NN has two layers; the first has radial basis activation functions and the second linear activation functions.

Simulations were carried out making cross validation with the training–validation set for different values of the selectivity σ, and computing the performance. The same was done for the test set.

For this method 20 principal components containing the 99.86% of the total information of the training-validation data were considered. That is to say the data dimension is reduced from 600 to only 20 points.

Different values of the selectivity were considered in the interval [2^{-9}, 10]. For higher values of the selectivity the results were poor in both validation and test sets. The results obtained are presented in Table III.

Classification results given in Table III show that, the best case is 76% of correct classification using the test set.

The simulations were carried out using Matlab 6.0, the Neural Network Toolbox and the Signal Processing Toolbox.

The average processing time for each simulation are shown in Table IV as a function of the number of principal components chosen. The average is computed over three runs performed for each simulation.

TABLE IV
AVERAGE PROCESSING TIME USED IN THE SIMULATION OF THE METHOD PCA-RBFNN FOR DIFFERENT PATTERN SIZES

Number of Principal Components	t[s]	Standard Deviation
5	891.82	4.517
10	905.64	4.498
15	968.37	4.209
20	998.22	4.008

The results shown in Table III were obtained with the aroma information of 100 wines coming from a database with an uneven class distribution. Increasing the aroma information, this is, having a larger database and improving the distribution of the classes, the percentage in correct classification in test set might improve.

For future studies it is suggested to analyze other feature extraction techniques like Wavelets and other classification techniques as Support Vector Machines (SVM). Currently we are working in incorporating more wine samples to the original database to complete 200 samples. The idea is to repeat this study with this new database these new techniques.

The results obtained in this study are promising and the first on Chilean wines using gas chromatograms supplied by an electronic nose. They provide the basis for future work on classification of Chilean wines. Other work developed by the authors on this subject can be found in [14, 15, 16].

V. CONCLUSIONS

The classification of red Chilean wines of the type Cabernet Sauvignon, Merlot and Carménère, from different vintages and different valleys, was successfully performed based on the aroma information (Volatile Organic Compounds chromatograms) supplied by an electronic nose.

Principal Component Analysis was used as feature extraction and Radial Basis Function Neural Networks as classification technique. The best parameters for each method were obtained from the cross validation process with the training-validation set.

The RBFNN showed a discrete performance in the training-validation set with classification rates about 71% (for σ=0.00195) and a 76% in the test set (for σ=0.00391).

Finally, it is important to point out that the results of this study are promising and corroborate that in spite of the reduced dimension of the database it is indeed possible to classify wines according to the varieties using aroma information coming from an electronic nose. It should be highlighted that once the system is trained, a measurement to classify a sample takes a few minutes making of this system attractive for quality control process: Usually a conventional technique for the same purpose takes days and requires specialized laboratories.

REFERENCES

[1] A.Yamazaki and T. B. Ludermir, "Classification of Vintages of Wine by an Artificial Nose with Neural Networks", *Proceedings of Tercer Encuentro Nacional de Inteligencia Artificial*, Fortaleza, Brasil, 2001.

[2] M.S. Santos, "Construction of an Artificial Nose using Neural Networks". Ph.D. Thesis, Centre of Informatics, Federal University of Pernambuco, Brazil, 2000.

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Macmillan College Publihing Company, 1994.

[4] J.P. Santos, J. Lozano, H. Vásquez, J.A. Agapito, M.A. Martín, J. González. "Clasificación e Identificación de Vinos Mediante un Sistema de Estado Sólido", *Proceedings of the XXI Jornadas de Automática*, Sevilla, 2000.

[5] M. García, M. Aleixandre, J. Gutiérrez and M.C. Horrillo, "Electronic nose for wine discrimination", *Sensors and Actuators B: Chemical,* vol. 113, pp. 911-916, Feb. 2006.

[6] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, News York, 2002.

[7] J. Ghosh, A. Nag, *An Overview of Radial Basis Functions Networks*. Physica-Verlag, 2000.

[8] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, 1996.

[9] H. Mhaskar, M. Michelli, "Approximation by Superposition of Sigmoidal and Radial Basis Functions", *Advances in Applied Mathematics*, vol 13, pp. 350-373, 1992

[10] J. Schurmann, *Pattern Classification: A Unified View of Statistical and Neural Approaches*, J. Wiley & Sons, 1996.

[11] K. Fukunaga and R. Hayes, "Estimation of Classifier Performance". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1087-1101, Oct. 1989.

[12] F. J. Cortijo Bon (2001). Selección y Extracción de Características. Available: http://www-etsi2.ugr.es/depar/ccia/rf/www/tema5_00-01_www/tema5_00-01_www.html.

[13] Electronic Sensor Technology, 7100 Fast GC Analyzer: Operation Manual, Electronic Sensor Technology, Bussines Center Circle, 1999.

[14] N.H. Beltrán, M.A. Duarte-Mermoud, S.A. Salah, M.A. Bustos, A.I. Peña-Neira, E.A. Loyola and J.W. Jalocha. "Feature selection algorithms using Chilean wine chromatograms as examples". *Journal of Food Engineering*, vol. 67, No. 4, pp. 483-490, Apr. 2005.

[15] N.H. Beltrán, M.A. Duarte-Mermoud, M.A. Bustos, S.A. Salah, E.A. Loyola, A.I. Peña-Neira and J.W. Jalocha, "Feature extraction and classification of Chilean wines". *Journal of Food Engineering*, vol.75, No. 1, pp. 1-10, Jul. 2006.

[16] M.A. Bustos, M.A. Duarte-Mermoud, N.H. Beltrán, S.A. Salah, E.A. Loyola, A.I. Peña-Neira and J.W. Jalocha, "Classification of Chilean wines using a Bayesian approach" (In Spanish). *Viticultura / Enología Profesional*, No. 90, pp. 63-70, Jan.-Feb. 2004.