Chilean Wine Classification Using Volatile Organic Compounds Data Obtained With a Fast GC Analyzer

Nicolás H. Beltrán, *Senior Member, IEEE*, Manuel A. Duarte-Mermoud, *Member, IEEE*, Víctor A. Soto Vicencio, Sebastián A. Salah, and Matías A. Bustos

***Abstract*—The results of Chilean wine classification based on the information contained in wine aroma chromatograms measured with a Fast GC Analyzer (zNose^{TM}) are reported. The aroma profiles are the results of the derivative of frequency change responses of a surface acoustic wave (SAW) detector when it is exposed to a flux of wine volatile organic compounds (VOCs) during aroma measurement. Classification is done after two sequential procedures: first applying principal component analysis (PCA) or wavelet transform (WT) as feature extraction methods of the aroma data, which results in data dimension reduction. In the second stage, linear discriminant analysis (LDA), radial basis function neural networks (RBFNNs), and support vector machines (SVMs) are used as pattern recognition techniques to perform the classification. This paper compares the performance of three classification methods for three different Chilean wine varieties (Cabernet Sauvignon, Merlot, and Carménère) produced in different years, in different valleys, and by different Chilean vineyards. It is concluded that the highest classification rates were obtained using wavelet decomposition together with SVM with a radial base function (RBF) type of kernel.**

***Index Terms*—Aroma measurement, electronic nose, feature extraction techniques, pattern recognition techniques, statistical classification, support vector machines (SVMs), wine classification.**

I. Introduction

SEVERAL wine classification results using different approaches and methodologies have been reported during the last two decades. In [1]–[3], wine classification was done by using the concentration of some wine chemical compounds obtained through liquid or gas chromatograms.

A different approach has been taken by the authors in [4]–[6], where all of the information contained in wine liquid chromatograms is used, and where it was shown that it was not necessary to perform an identification of each peak of the chromatograms for classification purposes. The database reported in [4]–[6] is composed of 111 high-performance liquid chromatography (HPLC) wine chromatograms distributed as 25 samples of Cabernet Sauvignon, 37 samples of Merlot, and 37 samples of Carménère. Each 90-min profile containing 6751 points, which is a large number of points (high dimensionality) of input data to a classifier, was reduced by using resampling and feature extraction techniques. Feature extraction was done with discrete wavelet transform (DWT) [8], fast Fourier transform (FFT) [9], Fisher transform (FT) [9], and typical profiles (TPs) [5], [6]. Then, several classification methods were analyzed and compared, including linear discriminant analysis (LDA) [9], quadratic discriminant analysis (QDA) [7], K-nearest neighbors (KNNs) [9], and the probabilistic neural network (PNN) [7]. Wavelet extraction with a classifier based on PNN gave the best classifier performance. With a cross-validation process of the type leave-one-out (LOO), an average percentage of correct classification (92.5%) was reported.

Electronic noses [10] have been used in several applications ranging from food quality [11], [12] to medicine [13]. Special emphasis has been placed on the classification of alcoholic beverages using electronic noses. Recently, in [14], 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 PNNs for pattern recognition were used with good results. In [15], it is stated that the effect of ethanol, i.e., the major constituent of the head space of alcoholic beverages, generates a strong signal on the sensor arrays used in electronic noses, impairing aroma discrimination.

In [16], an aromatic classification of three wines of the same variety but different years (1995, 1996, and 1997) is reported. The input data for classification are obtained from an electronic nose based on six sensors of conducting polymers; therefore, each pattern generated by the electronic nose has six points. For classification purposes, a multilayer perceptron (MLP) trained with the backpropagation (BP) algorithm and a time delay neural network (TDNN) trained with the Levenberg–Marquadt algorithm [17] were used. The generated database contained 5400 patterns, divided in sets for training (50%), validation (25%), and test (25%). It was shown that incorporating temporal processing improved the classification rate, i.e., the TDNN had better performance than the MLP.

A classification of seven wines of three different varieties (three white, three red, and one rose) was done in [18]. The input data for the classification process were obtained from a spectrophotometer that performs spectral analysis in the ultraviolet visible (VIS–UV) and near-infrared (NIR) ranges. As

Manuscript received June 4, 2007; revised December 7, 2007. First published June 6, 2008; current version published October 10, 2008. This work was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONYCIT), Chile, under Grant FONDEF D01-1016.

N. H. Beltrán and M. A. Duarte-Mermoud are with the Electrical Engineering Department, University of Chile, Santiago 1058, Chile (e-mail: Nicolas.Beltran@die.uchile.cl).

V. A. Soto Vicencio was with the University of Chile, Santiago 1058, Chile. He is now with AMEC-CADE, Santiago, Chile.

S. A. Salah was with with the University of Chile, Santiago 1058, Chile. He is now with Inversiones Ultra, Santiago, Chile.

M. A. Bustos was with the University of Chile, Santiago 1058, Chile. He is now with Mining and Metallurgical Research Center, Santiago, Chile.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2008.925015

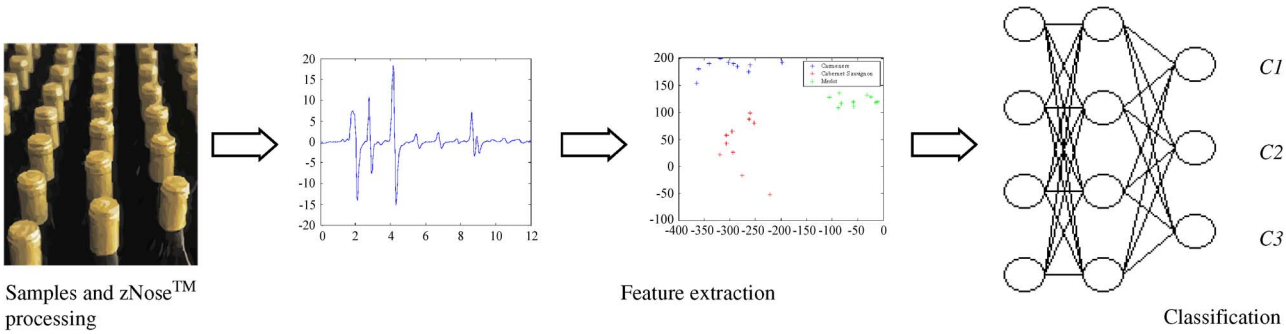


Fig. 1. Block diagram of the proposed methodology for wine classification.

the classifier, an MLP trained with BP was used [17]. Two databases (one with 100 patterns for VIS–UV analysis and the other with 120 patterns for NIR analysis) were used. In both cases, the data were divided into training and validation sets, obtaining a better performance with the NIR analysis reaching classification rate near 100% in validation. In [19], wine classification is done using neural networks (NNs) with data provided by an electronic nose built by the authors using commercially available sensors. These sensors are made of tin oxide and use the principle of resistance variations due to the adsorption of gas molecules on its surface.

The results reported above give a general overview of wine classification. In this paper, a comparison of wine classification methods based on neural and statistical techniques, using Chilean red wine aroma information of the varieties Cabernet Sauvignon, Merlot, and Carménère, is presented. Aroma chromatograms were obtained from a Fast GC Analyzer zNose™ model 7100 built by Electronic Sensor Technology [20]. The aroma profiles are the results of the derivative of frequency change responses of a surface acoustic wave (SAW) detector, when it is exposed to a flux of wine volatile organic compounds (VOCs) during aroma measurement. The gas detector used by the system is a proprietary SAW device. Although the GC technique to analyze the gas mixture used by zNose™ is a well-known technique, the instrument highlights the fact that the measurement of odor is performed in a short time (usually less than a minute) and is suitable for routine analysis. In comparison, techniques for the identification of VOC in wine, like a mix of GC and mass spectroscopy (MS) or UV–NIR or e-noses, are not usually used for routine analysis since the results are time consuming and require expertise for the right interpretation.

The first step in the proposed methodology is concerned with dimensional reduction of the patterns while preserving the original information. This is done using feature extraction methods like PCA [21] and DWT [8]. Once the dimension of the input data has been reduced, a classification stage follows, where classical statistical techniques like LDA [21], [22] and a technique based on radial basis function neural networks (RBFNNs) [21], [23] are used. Finally, a classifier based on support vector machines (SVMs) [24]–[27] was also studied. In Fig. 1, a block diagram illustrates the methodology.

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

II. Brief Description of the Feature Extraction and Classification Techniques

A. Feature Extraction Techniques

The main goal of the feature extraction techniques is to reduce the dimension of data input to make the data analysis simpler. These techniques are usually based on transformations from the original data space into a new space of lower dimension. In the linear case, we look for a matrix transformation W such that

$$y = W \cdot x \tag{1}$$

where x is the original feature vector of dimension m, W is the transformation matrix of $p \times m$, and y represents the new vector of transformed features of dimension p, with $p < m$. In this paper, we use DWT and PCA as feature extraction techniques. Details on the application of these methods are given in Sections III-D and -E, respectively.

B. Pattern Recognition Techniques

Pattern recognition covers a wide spectrum of problems from many scientific disciplines, particularly in the engineering area and its applications. Currently, the classification and pattern recognition techniques show a strong development motivated by the new advances in robotics and artificial intelligence. Some of the most important problems in the area of classification and pattern recognition correspond to artificial vision [37], character recognition [38], computerized medical diagnosis [39], and speech recognition [40].

Mathematically speaking, a pattern is an n-dimensional vector having in its components the characteristics of determined phenomenon or object. Generally, this vector is obtained as a response of a sensor with a significant sensitivity to some species. The design of a pattern classification system consists of three stages: data acquisition, preprocessing and data representation, and classification and decision making.

The first stage is performed by a set of sensors, whereas the aim of the second stage is to eliminate the noise measurement and to make the data ready for the classification stage, where

a suitable method is chosen so that the patterns are recognized and associated with a well-defined class.

The classification and pattern recognition techniques can be classified into the following three groups:

1) Statistical: The patterns are classified according to a model or statistical distribution of the characteristics.
2) Neural: The classification is done by means of a network formed by basic units (neurons) responding to an input stimulus or pattern.
3) Structural: The patterns are classified based on a measure of structural similarity.

Techniques based on neural networks have shown good performance for a wide range of applications [21], [22], and they have become attractive for instrumentation developers, because a minimum knowledge of the patterns is required, giving some simplicity overcoming the mathematical rigors of other techniques [17].

In this paper, the pattern recognition techniques LDA, RBFNN, and SVM are analyzed and compared. Since the LDA and RBFNN methods are well known, in the next section, a general and brief description of SVM is given.

C. SVM

SVM is a technique introduced by Vapnik [25] and his collaborators as a powerful classification and regression method. The main idea is that SVM minimizes the empirical risk (defined as the error on the training set) and minimizes the generalization error. The main advantage of the SVM applied to classification is that the classifier has a minimum Vapnik–Chervonenkis dimension [25], which implies a small error probability in the generalization. The other characteristic is that SVM allows classifying nonlinearly separable data, since the technique makes a mapping from the input space onto the characteristic space of higher dimension. Here, the data are linearly separable by a hyperplane, introducing the concept of optimal hyperplane.

Usually, for a better understanding of the SVM methodology, the case of linearly separable data is first analyzed, followed by the case when data are not linearly separable.

Linearly Separable Data Set: Let x be an input vector in $\Re^m$, and let $\mathbf{y}$ be its corresponding output, belonging to the set $\{-1, 1\}$. Let us consider n pairs of input–output vectors (x_i, y_i) with $x_i \in \Re^m$ and $i = 1, \ldots, n$. A hyperplane, which is able to perform a linear separation of the input data, will have the form

$$w^T x + b = 0 \tag{2}$$

where $w = [\, w_1 \quad w_2 \quad w_3 \quad \cdots \quad w_m \,]^T \in \Re^m$ represents the weighting vector, and b is a constant term.

According to the concept of optimal hyperplane proposed by Vapnik for linearly separable input data $x_i \in \Re^m$, the distance between the plane and the closest input vectors is maximum, and the optimal hyperplane separates the data without error. Then, the optimal hyperplane is that with the largest margin, defined as the minimum distance of a training vector to the decision surface (boundary), i.e., to the separator hyperplane.

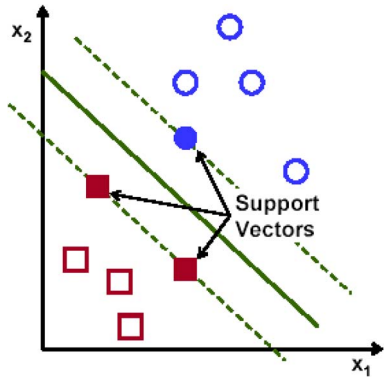


Fig. 2. Graphic representation of the separator hyperplane for a 2-D problem. The support vectors that define the maximum margin is explicitly shown.

Vapnik also proved that maximizing the margin also means the minimization of the Vapnik–Chervonenkis dimension [25]. From this, we can conclude that the separator hyperplane with maximum margin will also minimize the structural risk.

Fig. 2 shows the support vectors and the maximum margin with respect to the optimal separator hyperplane for a simple 2-D case. The separator hyperplane is defined by three vectors (support vectors), which replace the whole data set and allow all the patterns for each class (two in this case) to be correctly classified.

Nonlinearly Separable Data: In most of the classification problems, there is no linear function able to separate the data for each class. A solution is to map the input data onto a higher dimensional space, where the classification problem could be faced using a linear separator (hyperplane). In the case of nonlinearly separable data, the idea is to introduce new variables ξ_i (called slack variables) that relax the constraints on the canonical hyperplane [see (2)].

The slack variable measures the deviation of the data from the ideal case (linear case). For $0 \leq \xi_i \leq 1$, the data are located at the right side of the separator hyperplane, i.e., are correctly classified but positioned within the region of the maximum margin. For $\xi_i > 1$, the data are located at the incorrect side of the separator hyperplane and, therefore, are wrongly classified.

Additionally, a parameter C is introduced representing the tradeoff between the incorrect classification rate and the capacity of the model. Large values of C favor solutions with a small rate of incorrect classification. Small values of C give rise to models with low complexity. In addition, this parameter can be seen as a regularization parameter that can be determined by cross validation [9].

Based on the Cover theorem [28], which states that, in a nonlinearly separable classification problem, it is easier to find a linear separator hyperplane in a higher dimensional space than in a lower dimensional space, SVM operates in two stages. First, it performs a nonlinear map of the characteristic vectors onto a higher dimensional space. Second, it builds an optimal separator hyperplane in the high-dimensional space.

Some evident problems arise from this methodology, such as high dimension of data and large computational cost. These problems are faced in SVM in a very efficient way, since the generalization in the higher dimensional space is guaranteed

by designing a classifier with a maximum margin and since the projection onto the new space is only implicit [25]. To this respect, it is important to recall that the solution of the SVM only depends on the product $\langle x_i, x_j \rangle$ among the training patterns. However, this operation in the high-dimensional space $\boldsymbol{\varphi}(x)$ does not have to be explicitly done if we have a function $K(x_i, x_j)$ (so-called kernel [34]) such that

$$K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle. \tag{3}$$

Let us assume that the original characteristic vectors x belong to $\Re^m$ and that the nonlinear projection onto the space is $\boldsymbol{\varphi}(x) \in \Re^n$ with $n > m$, where the classes are linearly separable. The separator hyperplane in $\Re^n$ is given by

$$\sum_{j=1}^{n} w_j \varphi_j(x) + b = 0. \tag{4}$$

Using vector notation and eliminating b by imposing that $\varphi_1(x) = 1$, (4) can be written as

$$w^T \varphi(x) = 0 \tag{5}$$

and the optimal hyperplane, i.e., that with maximum margin, is given by

$$w = \sum_{i=1}^{n} \alpha_i y_i \varphi(x_i). \tag{6}$$

The kernel should be expressed as the inner product of two vectors. In this paper, we have chosen the radial basis kernel defined as

$$K(x, x') = \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right). \tag{7}$$

The parameter σ is determined by the user, but the number of radial base functions (RBFs) and their centers are automatically determined by the number of support vectors and their values.

In summary, the degrees of freedom of the SVM technique are basically three, i.e., the choice of the kernel, the choice of the kernel parameters, and the choice of the regularization parameter C, which penalizes the training errors.

III. Experimental Setup

Wine VOCs were detected with zNose$^{\mathrm{TM}}$ model 7100. The zNose$^{\mathrm{TM}}$ is a system that features high-speed gas chromatograph with a gas chromatograph sensor. The GC sensor includes a six-port valve and oven, a preconcentrating trap, a GC column of 1-m length, and a SAW detector [41]. The SAW detector is a thermally stabilized very high Q quartz resonator, which according to the manufacturer's specifications provides per billion sensitivity to VOCs. The sensitivity for zNose$^{\mathrm{TM}}$ is comparable to a long-column GC equipped with an electron-capture detector. High speed (in seconds) requires a tradeoff in resolving power; however, sufficient resolution is retained for many quality control applications. The vapor trap can sample weak vapors and preconcentrate before analysis in the GC column.

TABLE I
zNose$^{\mathrm{TM}}$ Setting Parameters Used to Perform Measures on the Wine Samples [20]

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

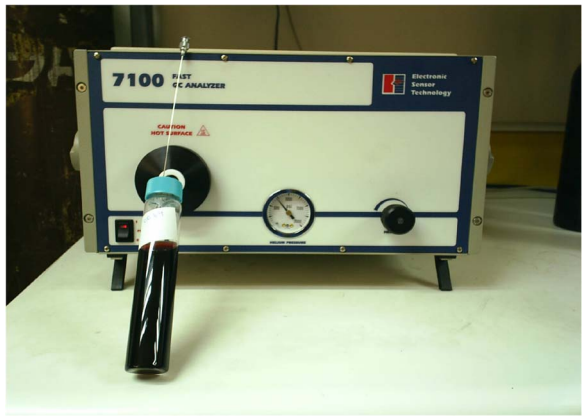


Fig. 3. Fast GC Analyzer zNose$^{\mathrm{TM}}$ model 7100 from Electronic Sensor Technology.

The operating parameters that can be set up in zNose$^{\mathrm{TM}}$ are the following:

- sensor — temperature of the SAW detector in degrees Celsius;
- column — temperature of the GC column in degrees Celsius;
- valve — temperature of the six position valves in degrees Celsius;
- inlet — temperature of the input gas in degrees Celsius;
- trap — temperature of the trap in degrees Celsius;
- ramp — value of the temperature ramp in degrees Celsius per second;
- acquisition time — duration time of the analysis in seconds;
- sampling period — rate at which the information is registered in seconds.

This set of parameters defines the method under which the instrument operates. After performing a series of tests and experiments, it was determined that the setting parameters for measuring wine VOC are those shown in Table I.

To obtain an aroma profile, 40 ml of each wine sample was introduced into 60-ml vials with septa caps, avoiding contact of the sample with oxygen in the air. The measurements were done immediately after the bottle was opened, maintaining the room temperature at 20 °C. Fig. 3 shows a photograph of zNose$^{\mathrm{TM}}$ during the measurement of a wine sample.

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

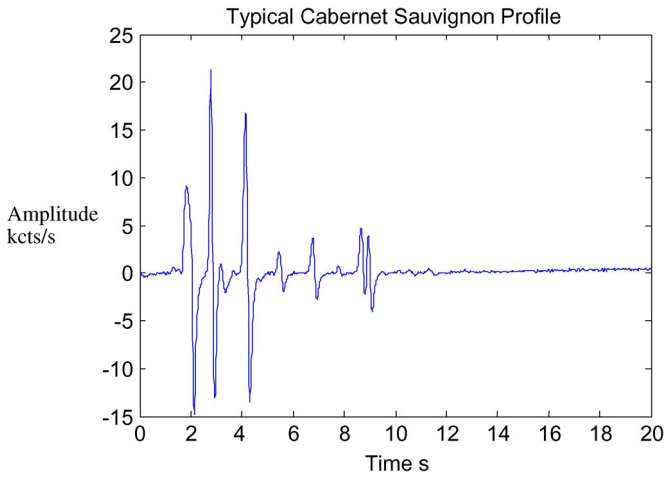


Fig. 4. Typical 20-s chromatogram of a Cabernet Sauvignon with sampling period of 0.01 s.

A. Database

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

B. Data Preprocessing

Chromatograms for each sample were obtained by setting the zNose$^{\mathrm{TM}}$ parameters as given in Table I. Ten measurements were done for each one of the 100 samples (one sample per bottle), obtaining in total 1000 profiles (chromatograms).

After 12 s, the chromatogram signal profile, shown in Fig. 4, is essentially a small amount of noise. This is because the 10 °C/s ramp temperature is applied between 2 and 12 s, spanning in temperature from 40 °C to 140 °C. Clearly, after 12 s, no more VOC wine components are released from the chromatograph column. Preliminary tests indicated that using the information contained in the first 12 s was equivalent to considering the completely 20-s profile. In addition, it was determined that by using a sampling period of 0.02 s, which is twice the original sampling period, similar results in terms of information content are obtained. Therefore, classification was performed by using chromatograms taken during 12 s and composed of 600 points. Another step on the data preprocessing is profile normalization. As shown in Fig. 4, the profile amplitude is variable around zero with positive and negative values. A scale factor was applied to normalize the amplitude in the interval $[-1, 1]$. To this extent, the maximum amplitude

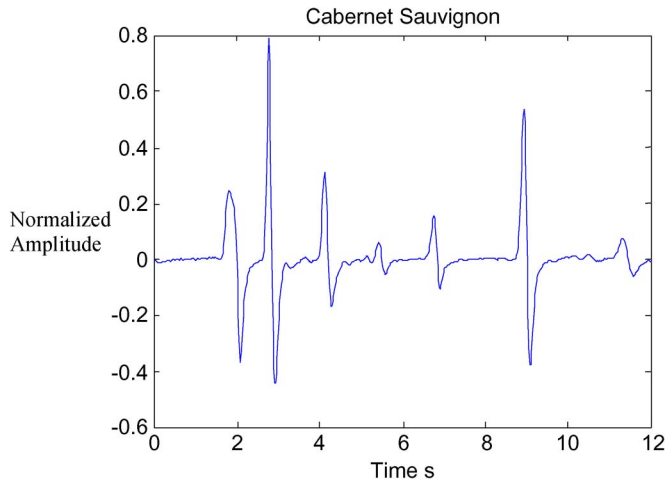


Fig. 5. Typical normalized chromatogram for a Cabernet Sauvignon (600 points).

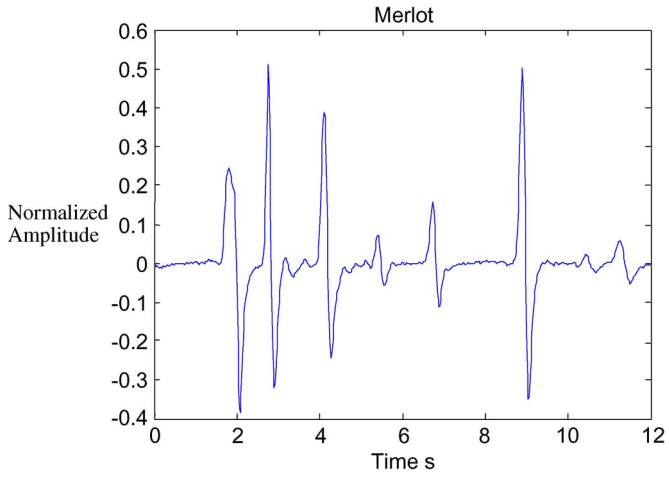


Fig. 6. Typical normalized chromatogram for a Merlot (600 points).

was used to normalize the signal according to the following relationship:

$$x_i' = x_i / x_{\max} \tag{8}$$

where $x_{\max}$ is the maximum signal amplitude of all profiles. The normalized profiles that followed this procedure are shown in Figs. 5–7 for Cabernet Sauvignon, Merlot, and Carménère, respectively.

The chromatograms shown in Figs. 5–7, taken from Cabernet Sauvignon, Merlot, and Carménère samples, look very much alike. However, the methodology that we developed in this paper does not take into account the wine profile shape but instead extracting their differences by means of feature extraction techniques, allowing their classification in the next stage. That is probably one of the most challenging aspects of this study, since starting from apparently indistinguishable information contained in the chromatograms, the classification system will be able to discriminate between wine classes with a certainty level of about 90%.

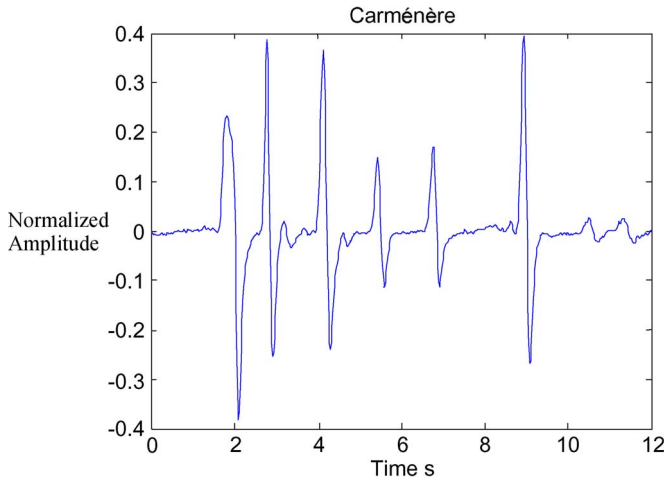


Fig. 7. Typical normalized chromatogram for a Carménère (600 points).

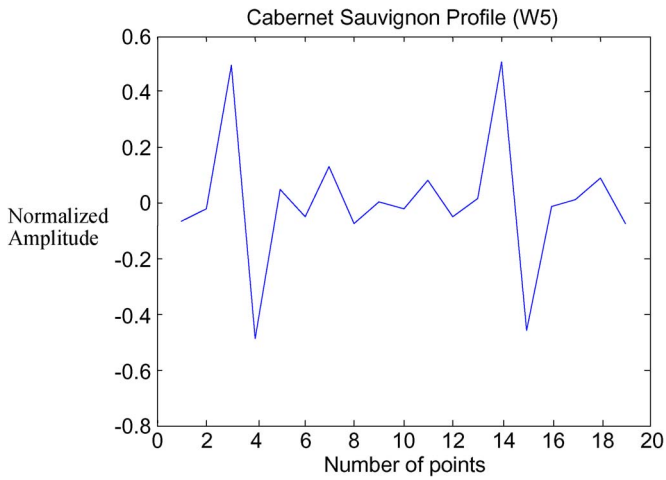


Fig. 8. Cabernet Sauvignon profile after WA for decomposition level 5 containing 19 points.

C. Methodology

To classify the profiles described in Section III-B, three classification techniques were used, i.e., LDA, RBFNN, and SVM. Following data dimension reduction, a feature extraction procedure was performed using the techniques of PCA and wavelet analysis (WA).

Once the data dimension was reduced by using the feature extraction method, the total databases of 1000 profiles (360 Cabernet Sauvignon or Class 1, 440 Merlot or Class 2, and 200 Carménère or Class 3) were divided in two sets: one for training–validation (containing 90% of the samples) and the other for test (containing 10% of the samples).

The sample distribution is as follows:

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 ten wine samples, 30 profiles Cabernet Sauvignon (three samples), 50 profiles Merlot (five samples), and 20 profiles Carménère (two samples).

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

As a measure of the behavior and to obtain the optimal values of the parameters for each method, cross validation was used [22], [29]. The database is divided into n sets using $n-1$ for training and the remainder for validation. The process is repeated n times so that all n sets are used once for validation.

In training–validation (cross validation with the aim to measure the behavior and to tune the optimal parameters for each classification), feature extraction methods are used. Then, each classifier is assessed with the test set using the whole training–validation set and the optimal parameters determined by cross validation. It is important to notice that the test set is never used in the training stage, and therefore, it was completely unknown to the classifier, becoming a good performance measure of each method. Each one of the 90 experiments was independently performed, choosing a different partition of the data set.

Finally, once cross validation is done and the optimal parameters are found for each method, a simulation is carried out with the test set to evaluate the performance of each method when unknown samples are presented. The classifier behavior is assessed in terms of the percentage of correct classification in the test set.

D. Feature Extraction Using WA

When DWT is applied to a signal of n points, it is equivalent to filtering the signal by low-pass and high-pass filters, which leads to getting the $n/2$ detail and approximation coefficients each time the filtering process is applied. The former is obtained from high-pass filters and the latter from low-pass filters. In this paper, only the approximation coefficients were considered for the classification stage since they contain most of the energy signal [30].

It is possible to choose the analysis level (or decomposition level of the signal) by defining the number of successive times that the wavelet transform (WT) will be applied. If the decomposition level is set as p for a signal with n original points, in stage p of the filtering process, we will have a signal of $n/2^p$ points represented by the detail and approximation coefficients. In this paper, different decomposition levels were examined by choosing the values 2–5. Since the chromatograms include 600 points, the first decomposition level generates a curve with 300 points with the approximation coefficients of the initial profile. At the second level, it has 150 points, and so on. As a mother wavelet $\psi(t)$ function, the orthonormal Haar basis was chosen due to its simplicity and because there are no free design parameters to be chosen by the user, as suggested in [30] and [31].

Figs. 8–11 show the profiles obtained after WA for decomposition levels 2–5, which correspond to a Cabernet Sauvignon sample. For example, for the fifth decomposition level, the profile contains only 19 points, $(600/2^5 \approx 19)$. From these results, the compression effects exhibited by the WT are clearly seen since in the new profile of 19 points, each point represents

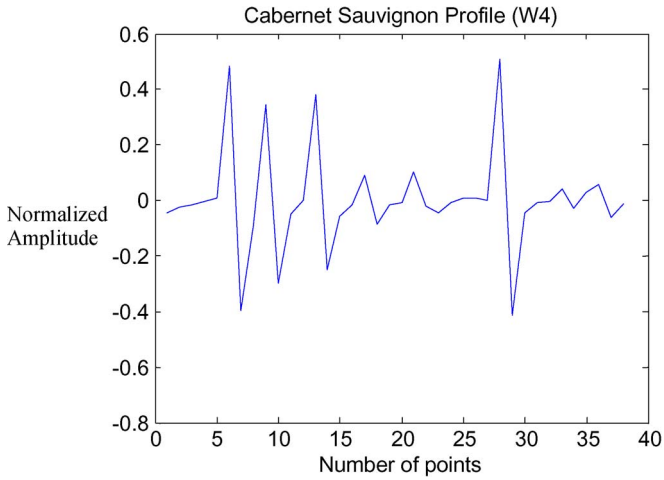


Fig. 9. Cabernet Sauvignon profile after WA for decomposition level 4, containing 38 points.

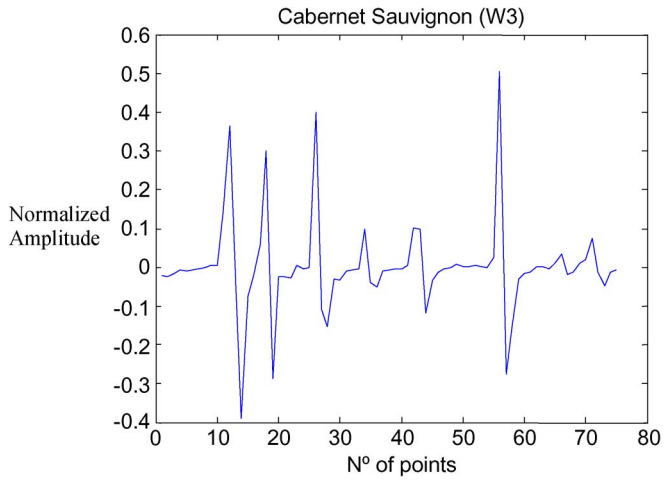


Fig. 10. Cabernet Sauvignon profile after WA for decomposition level 3, containing 75 points.

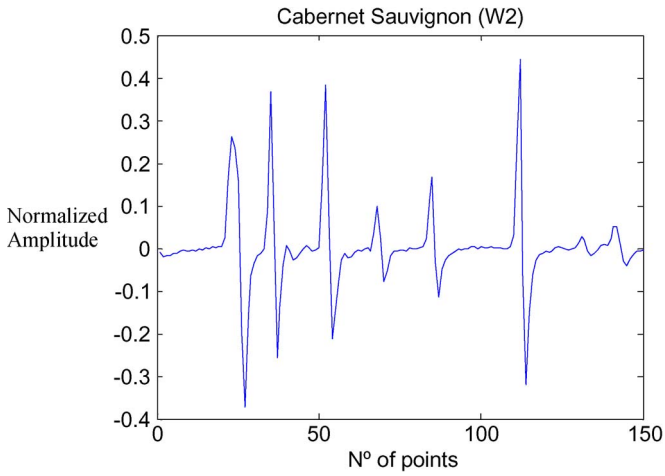


Fig. 11. Cabernet Sauvignon profile after WA for decomposition level 2, containing 150 points.

approximately 30 original profile points. This is achieved because the approximation coefficients were chosen for the analysis, and they keep most of the energy of the original signal.

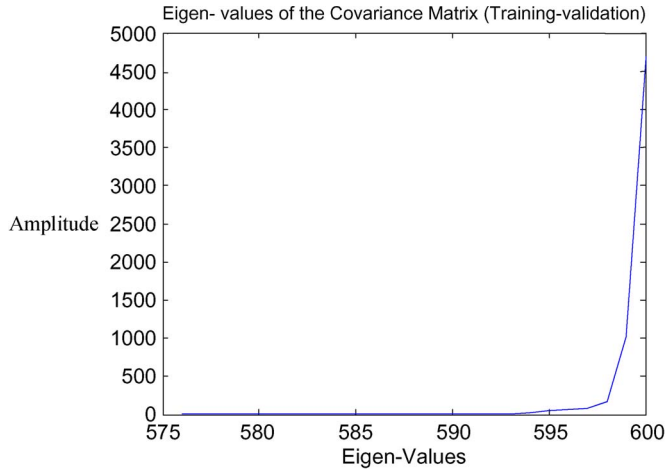


Fig. 12. Detail of the last 25 eigenvalues of the training–validation covariance matrix.

Therefore, low-frequency fluctuations retain the shape or envelope of the original signal, resulting in the space-temporal variations of the signal, which is a special characteristic particular to WA. Thus, these new profiles of lower dimension formed by the approximation coefficients are used in the classification process instead of the original profile. Notice that DWT is applied to the whole data set, including both training–validation and test sets.

The differences in samples after wavelet decomposition are shown in Figs. 8–11 for four decomposition levels. For the sake of space, this is shown only for a Cabernet Sauvignon wine sample. The other classes of wine (Merlot and Carménerè) have similar waveforms for each decomposition level as those shown in Figs. 8–11.

E. Feature Extraction Using PCA

PCA transforms the input space of variables P onto space P', where the data are not correlated, i.e., the variance of the data is maximum. This is achieved by computing the eigenvalues and eigenvectors of the covariance matrix of the initial data and selecting those eigenvectors with the largest eigenvalues. 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 $\times$ 1000, where each of the 1000 columns correspond to one profile, and the 600 rows correspond to the points to be reduced. Considering the training–validation set, we have a matrix of 600 $\times$ 900 [900 profiles (columns) of 600 points (rows)]; then, the covariance matrix of the training–validation set is

$$\Sigma_x = xx^T \tag{9}$$

where x is the training–validation matrix, and Σ_x is the covariance matrix of x (600 $\times$ 600). Then, computing the eigenvalues λ_i and eigenvectors of Σ_x and selecting the eigenvectors associated to the largest eigenvalues, the principal component transformation matrix will be determined. One way of choosing the eigenvalues (and the eigenvectors associated) is

TABLE III
CLASSIFICATION RESULTS IN VALIDATION FOR THE TRAINING–VALIDATION SET USING WAVELET AND LDA

wavelet decomposition level	Average % of correct classification in validation	Standard deviation	Total number of patterns wrongly classified
5 (19 points)	80.2	0.304	178
4 (38 points)	86.8	0.290	118
3 (75 points)	87.7	0.278	110
2 (150 points)	87.0	0.290	117

considering the contribution to the global variance [22] of each eigenvalue γ_i as

$$\gamma_i = \lambda_i / \sum_{j=1}^{N} \lambda_j \tag{10}$$

where $N = 600$ is the total number of eigenvalues of the covariance matrix Σ_x. γ_i associates with each eigenvalue (and each eigenvector or principal component) a factor of relative importance, considering its contribution to the total variance. When computing the eigenvalues of the covariance matrix Σ_x, these are ordered in ascending order; thus, the last components are those contributing the most to the information (in terms of covariance), whereas those at the beginning can be considered as noise and, therefore, disregarded. In Fig. 12, the last 25 eigenvalues (of a total of 600) for the training–validation covariance matrix of 600×600 are plotted.

It is interesting to notice that these eigenvalues practically contain all the information in terms of covariance. When computing the contribution of the last 20 eigenvalues to the global covariance using (10), the contribution to the total information is 99.87%, whereas the last ten eigenvalues contribute with 99.46%. The matrix transformation composed by the 20 eigenvectors associated with the last 20 eigenvalues was chosen, generating a 20×600 matrix (the 600 columns represent the initial characteristics or points and the 20 rows the eigenvectors or new characteristics). Multiplying each original profile by the transformation matrix, a low-dimension profile is obtained (dimension 20), which will be used in the classification procedure. All profiles in the data set (training–validation and test sets) were subjected to the extraction procedure before classification, ending with 900 patterns of dimension 20 for training–validation and 100 patterns of dimension 20 for test.

IV. RESULTS

In this section, the results using all the combinations of feature extraction and classification methods are presented. All the simulations were performed on a PC Pentium 4 of 1.7 GHz and 512-MB RAM using Matlab 6.0 together with several toolboxes: *Discriminant Analysis Toolbox* [33], *Signal Processing Toolbox*, *Neural Network Toolbox*, *Wavelet Analysis Toolbox*, and the *OSU Support Vector Machines Toolbox* version 2.33 [36].

A. Classification Results Using LDA

One of the advantages of using PCA is that there are no free design parameters to be defined by the user. All necessary information is contained in the input data vectors (arranged as a matrix) and the target vector. In what follows, we show the classification results of the wine samples into the three classes, i.e., Cabernet Sauvignon, Merlot, and Carménère, using the different methodologies studied in this paper.

TABLE IV
CLASSIFICATION RESULTS WITH TEST SET USING WAVELET AND LDA

wavelet decomposition level	Average % of correct classification in test	Total number of patterns wrongly classified
5 (19 points)	80	20
4 (38 points)	74	26
3 (75 points)	62	38
2 (150 points)	58	42

1) Results With Wavelet Extraction: Several simulations using wavelet analyses were performed with decomposition levels 2–5, which reduced the dimension of the input data to 150, 75, 38, and 19 points, respectively (see Figs. 9–12).

Table III shows the results obtained for the training–validation set by crossed validation for different decomposition levels. The behavior was assessed as the average percentage and standard deviation of correct classification in the validation set. Recall those 900 patterns: 330 Cabernet Sauvignon (Class 1), 390 Merlot (Class 2), and 180 Carménère (Class 3) is the training–validation set.

From Table III, the best classification rate resulted for a decomposition level 3 that generates profiles of 75 points. Table IV shows the results obtained using the test set formed by 100 profiles: 30 Cabernet Sauvignon, 50 Merlot, and 20 Carménère.

The best classification rate for the test set in Table IV corresponds to a level decomposition 5 that generates a profile of 19 points, whereas the worst result is for a decomposition level 2 (profiles with 150 points).

From the compared results shown in Fig. 13, it is apparent that there is no decomposition level, giving good classification results for both training–validation and test sets. Notice the high percentages obtained for levels 2 and 3 in validation (87.0% and 87.7%, respectively) and the poor results for the test set (generalization) (58% and 62%, respectively). Level 5 gives the best classification results for the test set (80%), and in validation, its performance is only 80.2%.

The average processing time (over three runs) for each simulation is a function of pattern size (wavelet level), which varies from 3.81 s for level 5 to 52.43 s for level 2. The WA that was separately carried out is not included in this processing time.

2) Results Obtained Using PCA Extraction: Feature extraction using PCA was carried out for 5, 10, 15, and 20 principal

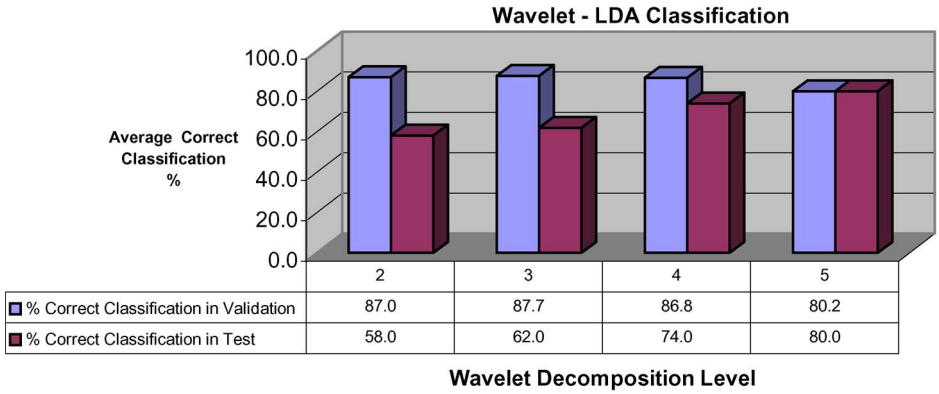


Fig. 13. Summary of the classification results using wavelet and LDA. Average percentage of correct classification for validation and test as a function of decomposition level.

TABLE V
CLASSIFICATION RESULTS FOR VALIDATION SET USING PCA AND LDA

Number of Principal Components	Contribution to the total variance	Average % of correct classification in validation	Standard deviation	Total number of wrongly classified
5	97.7%	70.1	0.406	269
10	99.5%	74.2	0.401	232
15	99.7%	75.5	0.388	220
20	99.8%	83.2	0.332	151

TABLE VI
CLASSIFICATION RESULTS IN THE TEST SET USING PCA + LDA

Number of Principal Components	% Correct classification in test	Total number of patterns wrongly classifies
5	37.0	63
10	39.0	61
15	44.0	56
20	59.0	41

components, whose contributions to the total variance and their classification results for training–validation are shown in Table V.

Table V clearly shows that from a set of only ten components (of a total of 600), a contribution to the total variance of over 99% is obtained. The best classification rate resulted with 20 point profiles.

Table VI shows the classification results obtained for the test set, where, as expected, a clear trend following the number of principal components is observed. The best classification rate (59%) is achieved with 20 principal components, which is much lower than the best classification rate obtained in validation (83.2%).

For comparison purposes, Fig. 14 shows the results obtained in validation and test. Notice the increase in classification rate for validation and test when the number of principal components increases. However, the results with the test set were much lower.

Regarding classification using LDA (details given in Figs. 13 and 14 and in Tables III–VI), we can comment that although LDA is a simple pattern recognition method, the percentages of correct classification for testing are under 80% when using either WT or PCA. There is an evident superiority on using WT over PCA as in the feature extraction stage, which is due to the capacity of wavelet technique to represent signals containing sharp peaks, like the ones shown in Figs. 5–7 for the three wine varieties. In contrast, LDA shows its poor generalization capacity, which is clearly observed since in all the results for the training–validation set, the percentages of correct classification are higher than those obtained when using the test set. This is, however, a characteristic of the LDA technique when it is used in problems where data are not easily separable.

The simulation average processing time (over three runs) depends on the number of principal components, on the time employed by the feature extraction method PCA, and on the time used in the simulation of LDA.

B. Classification Results Using RBFNN

The RBF used in the NN technique has the same form of kernel used for the SVM method given by (7). To apply RBFNN, the centers of the neurons and the parameter σ, which is known as *spread*, should be defined. The selectivity s of the neuron is then $s = 1/\sigma$. For all simulations, the neurons were located at each training pattern [23]; thus, when cross validation is carried out, the networks have a number of neurons corresponding to the dimension of each profile after the feature extraction technique has been applied (number of approximation coefficients or number of PCs depending on the feature extraction method used). Recall that the RBFNN has two layers; the first has radial basis activation functions and the second linear activation functions. Simulations were carried out

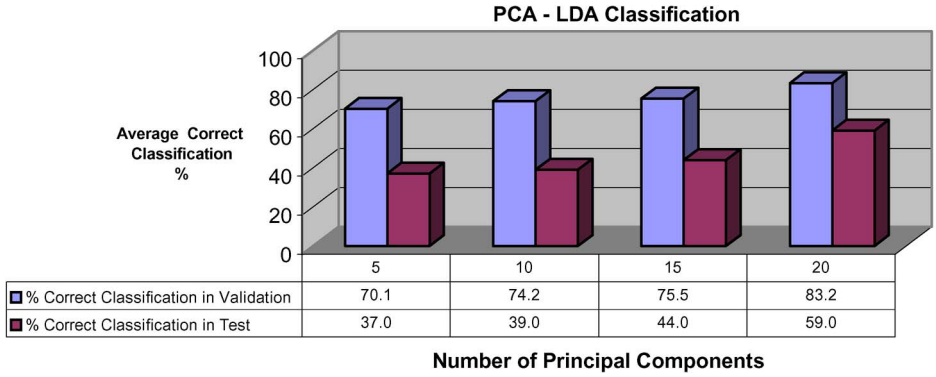


Fig. 14. Summary of the classification results for validation and test using PCA-LDA. The table below the graph shows the average percentage of correct classification for validation and test as a function of the number of principal components.

TABLE VII
RESULTS OF AVERAGE PERCENTAGE OF CORRECT CLASSIFICATION IN VALIDATION AS A FUNCTION OF WAVELET DECOMPOSITION, LEVEL, AND NEURON SELECTIVITY USING WAVELET EXTRACTION AND RBFNN

wavelet decomposition level	Selectivity=0.1		Selectivity=0.02		Selectivity=0.01	
	% Correct Classification in validation	Standard Deviation	% Correct Classification in validation	Standard Deviations	% Correct Classification in validation	Standard Deviation
5 (19 points)	75.6	0.339	76.8	0.329	76.5	0.320
4 (38 points)	75.6	0.339	76.5	0.339	72.7	0.373
3 (75 points)	74.8	0.340	76.6	0.346	75.4	0.332
2 (150 points)	72.3	0.359	75.4	0.354	74.1	0.335

TABLE VIII
AVERAGE PERCENTAGE OF CORRECT CLASSIFICATION FOR THE TEST SET AS A FUNCTION OF WAVELET DECOMPOSITION LEVEL AND NEURON SELECTIVITY USING WAVELET EXTRACTION AND RBFNN

	Selectivity=0.1	Selectivity=0.02	Selectivity=0.01
wavelet Decomposition level	% Correct Classification in validation	% Correct Classification in validation	% Correct Classification in validation
5	88.0	82.0	83.0
4	77.0	78.0	79.0
3	71.0	79.0	78.0
2	63.0	60.0	69.0

by making cross validation with the training–validation set for different values of selectivity s and computing the performance. A similar procedure was done for the test set.

1) Results Using Wavelet Extraction: To study the classifier performance based on RBFNN, different wavelet decomposition levels were used, and the method was assessed with the training–validation set doing a cross validation for different values of s. Table VII shows results where the average percentage of correct classification and its standard deviation are tabulated as a function of selectivity and wavelet decomposition level. From Table VII, it is seen that the percentage of correct classification is similar in all cases and above 70%. The best classification rate is obtained for a wavelet decomposition level equal to 5 (profiles of 19 points) and for a selectivity equal to 0.02 ($\sigma = 50$). The results obtained using the test set are shown in Table VIII.

From the results shown in Table VIII, the best classification rate in the test set (88%) is obtained for selectivity equal to 0.1 ($\sigma = 10$) and decomposition level equal to 5. For comparison purposes and as a summary, the classification results in validation and test are shown in Figs. 15 and 16.

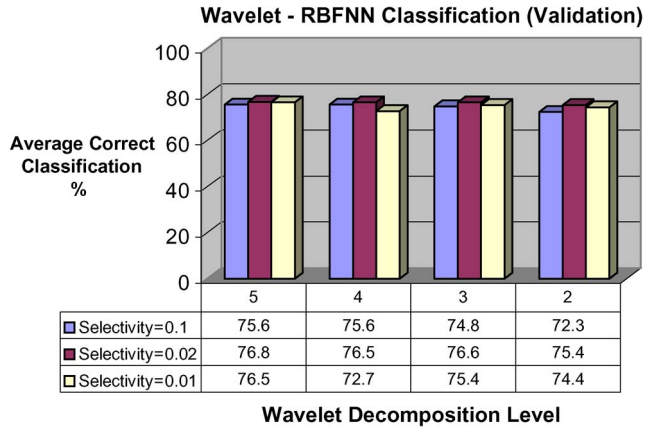
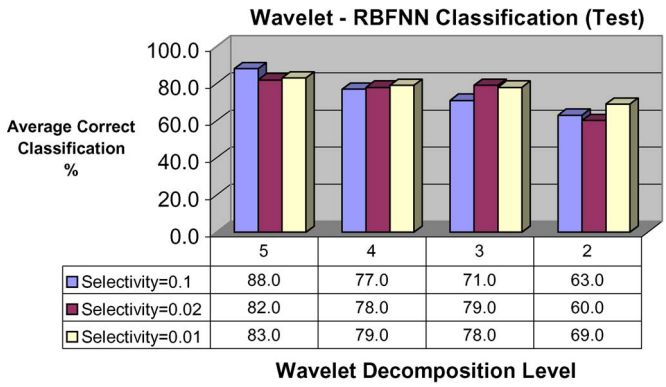


Fig. 15. Average percentage of correct classification in validation using cross validation as a function of selectivity and decomposition level for the method wavelet RBFNN.

These results show the high percentage of correct classification reached for the test set in spite of the reduced database used, which clearly shows the good generalization capacity of RBFNN.



	5	4	3	2
Selectivity=0.1	88.0	77.0	71.0	63.0
Selectivity=0.02	82.0	78.0	79.0	60.0
Selectivity=0.01	83.0	79.0	78.0	69.0

Fig. 16. Average percentage of correct classification in the test set as a function of selectivity and decomposition level for the method wavelet RBFNN.

The average processing time (over three runs) for this method varied from 554 s (for level 5) to 1237 s (for level 2).

2) Results Using PCA Extraction: For this method, the use of 20 principal components containing 99.86% of the total information of the training–validation data was considered, which is equivalent to the data dimension being reduced from 600 to only 20 points. Different values of selectivity were considered in the interval $[2^{-9}, 10]$. For higher selectivity values, the results of correct classification are unsatisfactory in both the validation and test sets. Table IX summarizes these results.

Fig. 17 shows a comparison of the classification results obtained by using RBFNN and PCA extraction.

From Fig. 17, it is clear that the classification results do not reach 80% correct classification and are lower than those obtained using wavelet extraction. The best results are obtained for the test set, and the average processing time (over three runs) goes from 892 s (for 5 PC) to 998 s (for 20 PC).

Wavelet extraction gives better results for testing (60%–88%) than those obtained from PCA extraction (smaller than 76%). This is due to the property of the wavelet to suitably represent signals with sharp peaks, which are common in the chromatograms used in this paper. The results plotted in Figs. 15–17 indicate the good generalization capacity of RBFNN compared to the LDA technique.

C. Classification Results Using SVM

For a classifier based on SVM, it is necessary to choose a kernel to carry out the mapping of the input data space. In [28] and [32], it is suggested to use the radial basis function type of kernel defined in (7). In [35], it is shown that the RBF kernel has the same behavior as the linear kernel with the parameters C and σ (C is the penalizing factor, and σ is the spread of the RBF). In this paper, a kernel described by (7) was used.

To determine the best values of C and σ, the methodology proposed in [32] was used, i.e., choosing values for C as well as $\boldsymbol{\sigma}$, which are power of 2. For example, $C = 2^0, 2^1, \ldots, 2^n$, etc., and for σ with $\sigma = 2^{-2}, 2^{-1}, 2^0, 2^1, \ldots, 2^n$, etc. Once the parameter coarse values are found, a refined analysis of them allows the best values to be chosen for classifier performance.

1) Results Using Wavelet Extraction: Several tests with different values of C and σ, for different decomposition levels, were carried out. The decomposition levels studied were 2, 3, 4, and 5, generating profiles of 150, 75, 38, and 19 points, respectively. After a series of simulations, it was determined that the best value for σ was $2^{-3.2} \approx 0.1$. In addition, it was found that the best classification rates in validation as well as in test were achieved for decomposition level 5.

Table X shows the results obtained for the SVM classifier, using an RBF kernel, for different values of C, using $\sigma = 0.1$, and wavelet decomposition level 5.

By using this technique, the best classification rate in validation was 84.8%, whereas in test, this rate was 90%, highlighting the good generalization property of the SVM. Notice that for wavelet decomposition level 5, the classification rate was quite high. However, as the decomposition level decreases (and therefore the dimension of the profile increases), the classification rate decreases in the test set, whereas in the validation, the classification rate is over 80% for all decomposition levels.

From the theoretical viewpoint by using SVM, the larger the training database, the smaller generalization error could be achieved [28], [35]. In spite of the reduced size of the databases used in this paper, the results obtained with SVM and wavelet extraction are quite promising.

Fig. 18 shows a summary of the best results obtained using SVM and wavelet extraction, where a decomposition level 5 by far has the best classification rate in the test set, and a consistent decreasing of the rate can be observed when the number of points increases.

The average processing time in this case goes from 34 s (for level 5) to 66 s (for level 2).

2) Results Using PCA Extraction: With PCA extraction, the performance was measured as a function of the parameters C and σ of the RBF kernel. By setting one of them to its nominal value ($C = 8192$ and $\sigma = 0.1$) and the other varying around the nominal value, Tables XI and XII summarize the results obtained for each case.

By comparing the results given in Table X (obtained using wavelet extraction) with those in Tables XI and XII, it is clear that by using PCA extraction, correct classification appears to be closer (but on the lower side) to the results obtained with wavelet extraction. For classifiers RBFNN and LDA, the difference between PCA and wavelet extraction is remarkable, whereas with SVM, both extraction methods give similarly acceptable results that are higher than 75% in the validation case. However, in the test set, the results of the SVM with PCA extraction are less satisfactory compared with those obtained with wavelet (level 5) and SVM.

The average processing time for each PCA–SVM simulation, considering three runs, varies from 1121 s (for 5 PC) to 1150 s (for 20 PC). This time corresponds to the time employed by the feature extraction method PCA together with the simulation of the SVM.

As with the two previous classifiers (LDA and RBFNN) and based on Fig. 18 and the results given in Tables X–XII, we here again obtain that wavelet extraction gives better results for the test set (80%–90%) than those using PCA extraction (lower than 60%). The reason is the same as mentioned in the LDA and RBFNN cases. Furthermore, the generalization

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

Selectivity	Average % of Correct Classification in validation	Standard Deviation	% Correct Classification in test
10	39.8	0.491	30.0
1	36.6	0.485	50.0
0.1	60.8	0.440	52.0
0.02	35.3	0.356	63.0
0.01	53.5	0.385	67.0
0.0078125	61.3	0.364	65.0
0.00390625	66.1	0.385	76.0
0.00195313	71.4	0.378	60.0

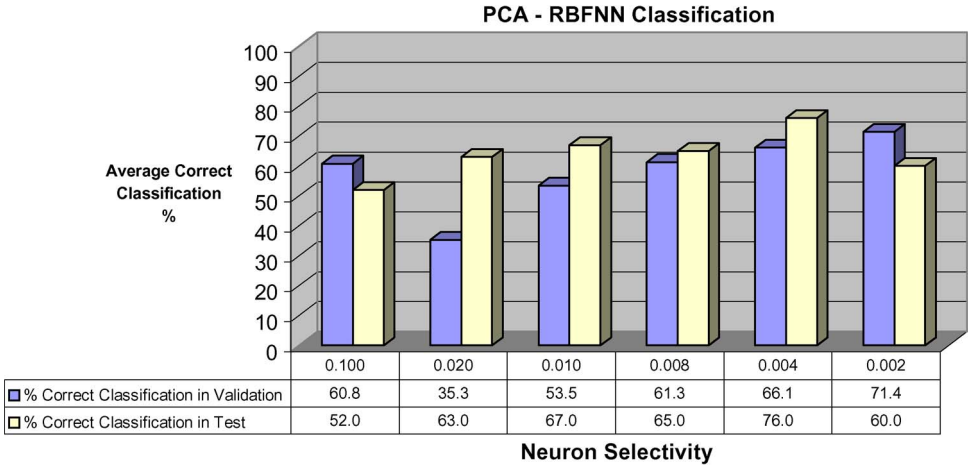


Fig. 17. Summary of the classification results in validation and test using RBFNN with PCA extraction for different values of neuron selectivity.

TABLE X
CLASSIFICATION RESULTS FOR WAVELET DECOMPOSITION LEVEL 5 (19 POINTS) AS A FUNCTION OF C FOR $\sigma = 0.1$ USING WAVELET WITH SVM IN VALIDATION AND TEST

C	σ	% Correct Classification in validation	Standard Deviation	% Correct Classification in test
128	0.1	76.6	0.383	80.0
256	0.1	77.3	0.369	83.0
512	0.1	79.5	0.343	84.0
1024	0.1	81.3	0.323	88.0
2048	0.1	82.2	0.313	90.0
4096	0.1	83.3	0.304	89.0
8192	0.1	84.0	0.300	89.0
16384	0.1	84.8	0.296	88.0
32768	0.1	84.8	0.290	84.0

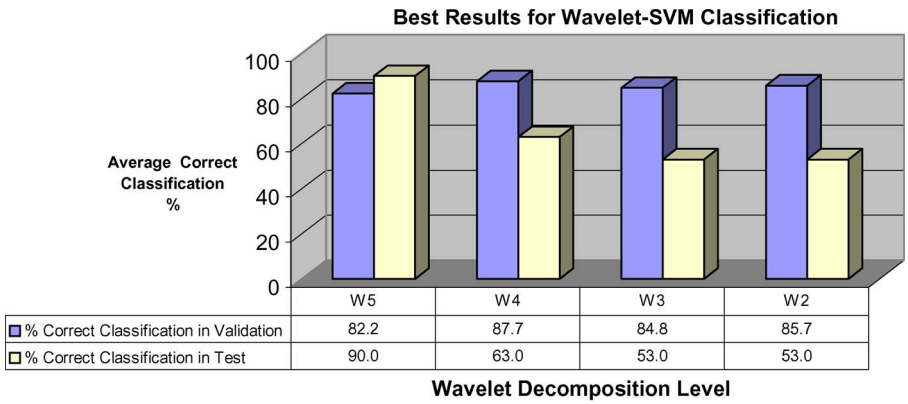


Fig. 18. Best classification results using SVM with wavelet extraction for different decomposition levels (W2, W3, W4, and W5).

TABLE XI
CLASSIFICATION RESULTS OBTAINED IN THE VALIDATION AND TEST SETS USING SVM AND PCA EXTRACTION FOR $C = 8192$ AND DIFFERENT VALUES OF σ

C	σ	% Correct Classification in validation	Standard Deviation	% Correct Classification in test
8192	5.00E-01	81.3	0.335	49.0
8192	2.50E-01	81.3	0.343	40.0
8192	1.25E-01	80.1	0.355	44.0
8192	6.25E-02	80.4	0.350	54.0
8192	3.13E-02	82.0	0.337	54.0
8192	1.56E-02	83.2	0.319	55.0
8192	7.81E-03	83.5	0.320	55.0
8192	3.91E-03	82.4	0.327	59.0
8192	1.95E-03	75.8	0.366	52.0
8192	9.77E-04	77.4	0.365	47.0
8192	4.88E-04	80.1	0.343	48.0
8192	2.44E-04	81.7	0.337	42.0
8192	1.22E-04	80.4	0.352	40.0

TABLE XII
CLASSIFICATION RESULTS OBTAINED IN THE VALIDATION AND TEST SETS USING SVM AND PCA EXTRACTION FOR $\sigma = 0.1$ AND DIFFERENT VALUES OF C

C	σ	% Correct Classification in validation	Standard Deviation	% Correct Classification in test
128	0.1	83.5	0.325	58.0
256	0.1	82.4	0.340	54.0
512	0.1	80.8	0.352	49.0
1024	0.1	79.8	0.359	43.0
2048	0.1	80.0	0.362	42.0
4096	0.1	80.3	0.351	46.0
8192	0.1	80.4	0.353	47.0
16384	0.1	79.7	0.354	46.0
32768	0.1	79.7	0.354	46.0

capacity of SVM is remarkable; it is even better than that of the RBFNN. This is probably one of the most important characteristics of SVM, which tmakes it one of the best pattern recognition techniques for a wide range of problems. In part, this is due to the fact that the separator hyperplane is built in an optimal way.

D. Summary of the Results

Figs. 19 and 20 summarize the best classification rates obtained with the test set for all the combinations of classifiers (LDA, RBFNN and SVM) and the two feature extraction methods used in this paper, i.e., wavelet (Fig. 19) and PCA (Fig. 20).

The results plotted in Fig. 19 show a good performance of the RBFNN and SVM classification methods for the test set, yielding, for both techniques, results of over 88% correct classification. Considering the test set correct classification, the best result was reached with wavelet extraction and SVM as a classifier (see Table X).

Fig. 20 shows the classification results when PCA extraction was used, which are less satisfactory than those obtained using wavelet extraction. Notice the good performance of the LDA and SVM classifiers in the validation set (both over 82.0% of correct classification), although they give poor results for the test set. It is interesting to observe that the best performance for testing using PCA is obtained for RBFNN (76.0%).

The best classification results in training–validation (the details of these results are not shown here for the sake of space) were obtained by combining wavelet extraction with classifiers LDA and SVM. Tables XIII and XIV show the confusion matrices for both cases. Confusion matrices are 3×3 matrices, where columns indicate the values predicted by the classifier, and rows indicate the real values contained in the objective vector (target) of each simulation. Therefore, the diagonal elements indicate the percentage of patterns correctly classified, and the off-diagonal elements indicate the patterns wrongly classified (classifier confusion).

From Table XIII, the best classification rate was obtained for Merlot (90.76%), whereas the worst classification rate was for the variety Carménère (80%). This is an expected result due to the large presence of Merlot in the training–validation database (43.3%) as compared with 20% of Carménère. In addition, the larger confusion occurs with Carménère, where 11.6% of the samples are confused with Merlot.

The confusion matrix in Table XIV shows that the best classification rate was achieved for the Merlot variety. By comparing with the confusion matrix in Table XIII, it is observed that LDA can better classify the variety Carménère but shows

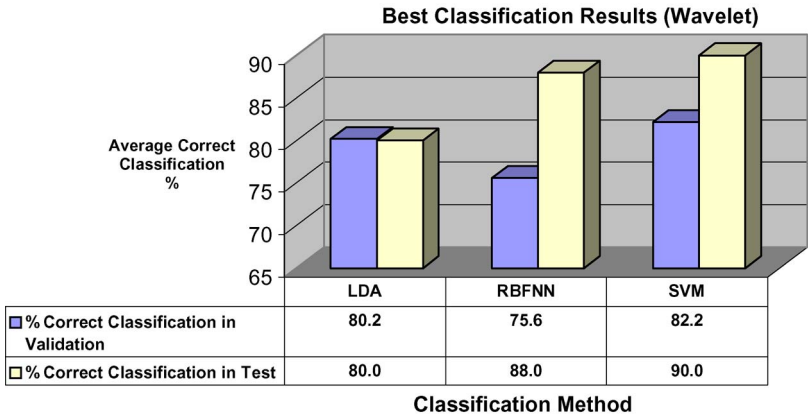


Fig. 19. Summary of the best classification results using wavelet as feature extraction.

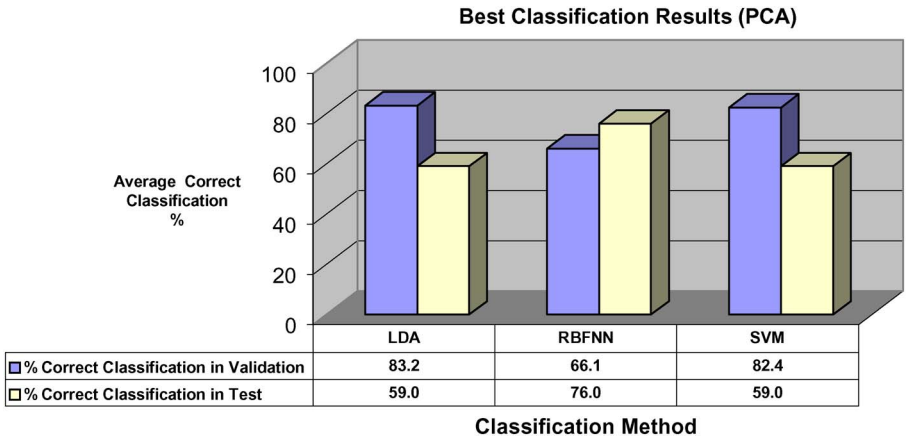


Fig. 20. Summary of the best classification results obtained using PCA as feature extraction.

TABLE XIII
CONFUSION MATRIX FOR THE BEST PERFORMANCE OF SVM CLASSIFIER IN VALIDATION. VALUES EXPRESSED IN PERCENTAGE WITH RESPECT TO THE TOTAL OF EACH CLASS

wavelet	Cabernet	Merlot	Carménère
Cabernet	88.48	9.39	2.12
Merlot	8.71	90.76	0.51
Carménère	8.30	11.60	80.00

TABLE XIV
CONFUSION MATRIX OBTAINED FOR THE BEST PERFORMANCE OF THE LDA CLASSIFIER IN VALIDATION. VALUES EXPRESSED WITH RESPECT TO THE TOTAL PATTERNS OF EACH CLASS

wavelet	Cabernet	Merlot	Carménère
Cabernet	86.60	12.12	1.21
Merlot	10.76	89.23	0.00
Carménère	10.00	3.30	86.67

more confusion between the Merlot and Cabernet Sauvignon varieties. The LDA classifier is confused between Cabernet Sauvignon and Merlot in 12.12% of the total patterns.

We believe that these results could be improved with a larger database featuring a uniform distribution of classes. Nevertheless, it is important to point out that the results are promising and demonstrate that it is feasible to classify Chilean wines according to varieties in spite of the reduced database dimension. Furthermore, the classification can be done by just using the aroma information of the wines, and the developed methodology highlights that classification is completed on the order of a couple of minutes, which makes it practically an online classification system.

V. CONCLUSION

The classification of Chilean wines of the varieties Cabernet Sauvignon, Merlot, and Carménère, from different vintages and different valleys, based only on the aroma information (gas chromatograms) detected by zNose^{TM}, has been successfully performed.

Two feature extraction techniques were applied (PCA and WA), together with three classification techniques (LDA, RBFNN, and SVM). For all six combinations, the performance was assessed as the average percentage of correct classification in the validation set (doing cross validation) as well as in the test set. The best parameters for each method were obtained from the cross-validation process with the training–validation set.

These results show the attractive characteristic of the WA, used as a feature extraction method, combined with the well-recognized capacity of SVM, the powerful discrimination of RBFNN, and the classical LDA for classification. The highest classification rates were obtained using wavelet decomposition,

together with SVM with an RBF type of the kernel. Specifically, for a decomposition level 5 with $C = 2048$ and $\sigma = 0.1$, an average classification rate of 90% in the test set was obtained (see Table X).

LDA with wavelet extraction showed good performance in training–validation, giving classification rates of over 85% (see Table III), but in the test set, the best classification rate was 80% for level 5 (see Table IV). We believe that in general, the reduced number of samples gathered in the databases affected the performance. LDA discrimination is heavily based on the following two assumptions: 1) The data follow a normal multivariate distribution, which can be guaranteed only for large set of data; and 2) the covariance matrices of the classes are all equal, which is a condition that is also true for large set of data.

RBFNN showed good performance in the test set with classification rates of 88% when wavelet extraction of level 5 and selectivity 0.1 were used (see Table VIII). However, a lower classifier performance was obtained when PCA was used for feature extraction (see Table IX).

The developed wine classification methodology highlights what can be achieved by just using wine aroma information, and a result is accomplished in a couple of minutes, which makes it practically an online classification system. Furthermore, it should be emphasized that the methodology developed in this paper is general and could be used in other applications as well, such as classification of coffee, olive oil, explosives, etc. A point that needs further study is the analysis of the effects of using other mother wavelets that could improve the results obtained in this paper. In addition, the analysis of different kernel functions for SVM is a subject to be studied and compared with the RBF type of kernel used in this paper. Another important aspect to consider in the future is to prove these methodologies in a more populated database.

REFERENCES

[1] M. D. Cabezudo, M. Herraiz, and E. F Gorostiza, "On the main analytical characteristics for solving enological problems," *Process Biochem.*, no. 18, pp. 17–23, 1983.

[2] P. Etievant and P. Schilich, "Varietal and geographic classification of French red wines in terms of mayor acids," *Agric. Food Chem.*, no. 47, pp. 421–498, 1989.

[3] J. Aires-de-Sousa, "Verifying wine origin: A neural network approach," *Amer. J. Enol. Vitic.*, vol. 47, no. 4, pp. 410–414, 1996.

[4] 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," *J. Food Eng.*, vol. 67, no. 4, pp. 483–490, Apr. 2005.

[5] 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," *J. Food Eng.*, vol. 75, no. 1, pp. 1–10, Jul. 2006.

[6] M. A. Bustos, M. A. Duarte-Mermoud, N. H. Beltrán, S. A. Salah, A. I. Peña-Neira, E. A. Loyola, and J. W. Jalocha, "Clasificación de vinos Chilenos usando un enfoque Bayesiano," *Viticultura y Enología Profesional*, no. 90, pp. 63–70, 2004.

[7] S. Aeberhard, O. de Vel, and D. Coomans, "Comparative analysis of statistical pattern recognition methods in high dimensional settings," *Pattern Recognit.*, vol. 27, no. 8, pp. 1065–1077, Aug. 1994.

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining, Interference, and Prediction*. New York: Springer-Verlag, 2001.

[9] A. Webb, *Statistical Pattern Recognition*. Hoboken, NJ: Wiley, 2002.

[10] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sens. Actuators B, Chem.*, vol. 18, no. 1–3, pp. 210–211, Mar. 1994.

[11] K. Brudzewski, S. Osowski, and T. Markiewicz, "Classification of milk by means of an electronic nose and SVM neural networks," *Sens. Actuators B, Chem.*, vol. 98, no. 2/3, pp. 291–298, Mar. 2004.

[12] S. Ampuero and J. O. Bosset, "The electronic nose applied to dairy products: A review," *Sens. Actuators B, Chem.*, vol. 94, no. 1, pp. 1–12, Aug. 2003.

[13] Y. J. Lin, H. R. Guo, Y. H. Chang, M. T. Kao, H. H. Wang, and R. I. Hong, "Application of the electronic nose for uremia diagnosis," *Sens. Actuators B, Chem.*, vol. 76, no. 1–3, pp. 177–180, Jun. 2001.

[14] M. García, M. Aleixandre, J. Gutiérrez, and M. C. Horrillo, "Electronic nose for wine discrimination," *Sens. Actuators B, Chem.*, vol. 113, no. 2, pp. 911–916, Feb. 2006.

[15] J. A. Ragazzo-Sanchez, P. Chalier, and C. Ghommidh, "Coupling gas chromatography and electronic nose for dehydration and desalcoholization of alcoholized beverages: Application to off-flavour detection in wine," *Sens. Actuators B, Chem.*, vol. 106, no. 1, pp. 253–257, Apr. 2005.

[16] A. Yamazaki and T. B. Ludermir, "Classification of vintages of wine by an artificial nose with neural networks," in *Tercer Encuentro Nacional de Inteligencia Artificial*, Fortaleza, Brazil, 2001.

[17] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1994.

[18] M. Gaeta, M. Marsella, S. Miranda, and S. Salerno, "Using neural networks for wine identification," in *Proc. IEEE Int. Joint Symposia Intell. Syst.*, 1998, pp. 418–421.

[19] J. P. Santos, J. Lozano, H. Vásquez, J. A. Agapito, M. A. Martín, and J. González, "Clasificación e identificación de vinos mediante un sistema de estado sólido," in *Proc. XXI Jornadas de Automática*, Sevilla, Spain, 2000.

[20] *7100 Fast GC Analyzer: Operation Manual*, Electron. Sensor Technol., Newbury Park, CA, 1999.

[21] C. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 2002.

[22] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[23] J. Ghosh and A. Nag, *An Overview of Radial Basis Functions Networks*. Heidelberg, Germany: Physica-Verlag, 2000.

[24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[25] V. Vapnik, *Statistical Learning Theory*. Hoboken, NJ: Wiley, 1998.

[26] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[27] D. Michie, J. Spiegelhalter, and C. Taylor, *Machine Learning, Neural and Statistical Classification*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[28] C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*. Boston, MA: Kluwer, 2000.

[29] K. Fukunaga and R. Hayes, "Estimation of classifier performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 10, pp. 1087–1101, Oct. 1989.

[30] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Process. Mag.*, vol. 8, no. 4, pp. 14–38, Oct. 1991.

[31] D. Y. Pan, "Digital audio compression," *Digit. Tech. J.*, vol. 5, no. 2, pp. 28–40, 1993.

[32] C. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification*. Taipei, Taiwan, R.O.C.: Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., 2003.

[33] M. Kiefte, *Discriminant Analysis Toolbox*. Edmonton, AB, Canada: Univ. Alberta, 2000.

[34] C. Burges, B. Scholkopf, and A. Smola, *Advances in Kernel Methods Support Vector Learning*. Cambridge, MA: MIT Press, 1999.

[35] S. Keerthi and C. J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, 2003.

[36] C. C. Chang and C. J. Lin, *OSU Support Vector Machines (SVMs) Toolbox*. Taipei, Taiwan, R.O.C.: Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., 2002.

[37] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, 3rd ed. San Mateo, CA: Morgan Kaufmann, 2004.

[38] L. C. Jain and B. Lazzerini, Eds., *Knowledge-Based Intelligent Techniques in Character Recognition*, Boca Raton, FL: CRC, 1999.

[39] A. Dhawan, *Medical Image Analysis*. Hoboken, NJ: Wiley, 2003.

[40] W. Chou and B. H. Juang, Eds., *Pattern Recognition in Speech and Language Processing*, Boca Raton, FL: CRC2003.

[41] E. J. Staples and S. Viswanathan, "Ultra-high speed chromatography and virtual chemical sensors for detecting explosives and chemical warfare agents," *IEEE Sensors J.*, vol. 5, no. 4, pp. 622–631, Aug. 2005.

Nicolás H. Beltrán (M'86–SM'04) received the degree in electrical engineering from the University of Chile, Santiago, Chile, in 1974 and the Master's degree in electrical engineering and the Doctoral degree in applied sciences from Katholieke Universiteit Leuven (KUL), Leuven, Belgium, in 1981 and 1985, respectively.

From 1975 to 1979, he was with the Venezuelan Institute for Scientific Research (IVIC). Since 1985, he has been with the Electrical Engineering Department, University of Chile, where he has conducted research on ceramic sensors and intelligent instrumentation.

Manuel A. Duarte-Mermoud (M'01) received the degree of electrical engineer from the University of Chile, Santiago, Chile, in 1977 and the M.Sc., M.Phil., and Ph.D. degrees in electrical engineering from Yale University, New Haven, CT, in 1985, 1986, and 1988, respectively.

From 1977 to 1979, he was a Field Engineer with Santiago Subway. Since 1979, he has been with the Electrical Engineering Department, University of Chile, where he is currently a Professor. His main research interests are in robust adaptive control (linear and nonlinear systems), system identification, signal processing, and pattern recognition. He is focused on applications to the mining and wine industries, sensory systems, and electrical machines and drives.

Dr. Duarte is member of the IFAC. He is past Treasurer and past President of ACCA, the Chilean National Member Organization of IFAC, and past Vice-President of IEEE Chile.

Víctor A. Soto Vicencio received the degree of electrical engineer in 2004 and the Diploma in the regulation of electrical markets from the University of Chile, Santiago, Chile.

Since 2004, he has been a Project Engineer with the international engineering company AMEC-CADE, where he primarily works in the development and studies of power systems and industrial plants.

Sebastián A. Salah received the degree of electrical engineer and the Master of Science degree in electrical engineering from the University of Chile, Santiago, Chile, in 2004.

Since 2004, he has been with Inversiones Ultra, where he is developing solutions in the financial area as R&D Manager.

Matías A. Bustos was born in Santiago, Chile, in 1977. He received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Chile, Santiago, in 2004.

He is currently a Senior Research Engineer with the Services and Technology Division, Mining and Metallurgical Research Center, Chile, where he is involved in the development of computer vision and image processing algorithms used in soft sensor applications for mining processes.