Dimensionality Reduction Oriented Toward the Feature Visualization for Ischemia Detection

Edilson Delgado-Trejos, *Member, IEEE*, Alexandre Perera-Lluna, Montserrat Vallverdú-Ferrer,
Peré Caminal-Magrans, *Member, IEEE*, and Germán Castellanos-Domínguez, *Member, IEEE*

Abstract—An effective data representation methodology on high-dimension feature spaces is presented, which allows a better interpretation of subjacent physiological phenomena (namely, cardiac behavior related to cardiovascular diseases), and is based on search criteria over a feature set resulting in an increase in the detection capability of ischemic pathologies, but also connecting these features with the physiologic representation of the ECG. The proposed dimension reduction scheme consists of three levels: projection, interpretation, and visualization. First, a hybrid algorithm is described that projects the multidimensional data to a lower dimension space, gathering the features that contribute similarly in the meaning of the covariance reconstruction in order to find information of clinical relevance over the initial training space. Next, an algorithm of variable selection is provided that further reduces the dimension, taking into account only the variables that offer greater class separability, and finally, the selected feature set is projected to a 2-D space in order to verify the performance of the suggested dimension reduction algorithm in terms of the discrimination capability for ischemia detection. The ECG recordings used in this study are from the European ST–T database and from the Universidad Nacional de Colombia database. In both cases, over 99% feature reduction was obtained, and classification precision was over 99% using a five-nearest-neighbor classifier (5-NN).

Index Terms—Dimensionality reduction, feature extraction, feature selection, feature visualization, ischemia detection, multidimensional analysis, relevance, wavelet transform (WT).

I. INTRODUCTION

AUTOMATIC pathology detection systems are based on a feature space, constructed from informative attributes that allow class discrimination. Usually, feature space is composed of clinical measures that reflect a specific biological behavior of the heart, but some features are extracted as abstract mathematical representations without a particular physiological meaning. Information in ECG signals related to the discrimination of the functional states is therefore represented by a large dimensional space, which hinders a proper interpretation of the embedded symbolic physiology in the feature space. Issues from large feature space dimensionality are inevitable, given the huge amount of data generated in medical experimental systems. The known difficulties of large dimensionality spaces should be emphasized: high processing times, high signal acquisition time and complexity, and the well-known effect of the curse of dimensionality [1].

Mostly, dimensionality reduction is related to the following three applied areas: multivariate data classification, data analysis, and data visualization [2]. Several methods have been studied for data dimensionality reduction, such as principal component analysis (PCA) [3], independent component analysis (ICA) [4], and multidimensional scaling (MDS) [5], among others. These methods are well understood, being robust and easy to implement, but unfortunately, their limitation is the strong assumption of linearity, and the new variables may not always be easy to interpret. The main goal in dimensionality reduction is to preserve the maximum information from original data given an optimization criterion (relevance) [6]. A solution of a feature selection problem by means of the classification rate is proposed in [7], where the initial feature space is composed of a number of features greater than 50. Even floating search methods in their backward or forward derivatives [sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS)] are suitable for small and medium feature sets; genetic algorithms (GAs) are more desirable for datasets with large dimensions (high computational complexity). Linear techniques have difficulties in reducing dimensionality from real and nonlinear data; two algorithms could be highlighted to deal with nonlinearities: Isomap [8] and local linear embedding (LLE) [9]; both of them aim to keep local topology properties while trying to achieve nonlinear embeddings [10]. Both Isomap and LLE have been used in visualization and classification, however, these techniques loose accuracy with real data with noise contamination. Supervised Isomap (S-Isomap) is proposed in [11], which uses class information to guide the nonlinear dimensionality reduction procedure. For S-Isomap, the vicinity graph for input data is built according to certain types of dissimilarity between data, easing the integration of class information. Although, when the training dataset is sparse (multimodal), vicinities could disconnect. S-Isomap is not

E. Delgado-Trejos is with the Machine Intelligence and Pattern Recognition Group, Research Center, Instituto Tecnológico Metropolitano, Medellín 57, Colombia (e-mail: edilsondelgado@itm.edu.co).

A. Perera-Lluna, M. Vallverdú-Ferrer, and P. Caminal-Magrans are with the Centre for Biomedical Engineering Research (CREB), Department of Systems, Automatic and Industrial Informatics Engineering, Polytechnical University of Catalunya, 08034 Barcelona, Spain, and also with the Center for Biomedical Research Network in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), 50018 Saragossa, Spain (e-mail: alexandre.perera@upc.edu; montserrat.vallverdu@upc.edu; pere.caminal@upc.edu).

G. Castellanos is with the Control and Digital Signal Processing Group, Department of Electrical, Electronics and Computer Engineering, Universidad Nacional de Colombia, 64-60 Manizales, Colombia (e-mail: cgcastellanosd@unal.edu.co).

Digital Object Identifier 10.1109/TITB.2009.2016654

suitable for these cases, which unfortunately are frequent in real-life applications. In general, facilities for connecting multidimensional spaces of abstract mathematic representation with their physical source generator by dimensionality reduction are not clearly reported in the literature. Several techniques have been proposed for ischemic beat classification, which evaluate the ST segment changes and the T-wave alterations using different methodologies. Their evaluation process is based upon parametric modeling [12], [13], wavelet transform (WT) [14], set of rules [15], [16], and artificial neural networks (NNs) [17], [18]. Although these methods have achieved satisfactory results, and some improvements can be suggested, none of the aforementioned techniques combine high performance with the ability to interpret the classification results [19].

In this study, a dimensionality reduction scheme at three levels is proposed. The first level is based on the linear mapping of the data to a time–frequency representation space with smaller dimension. The second level uses a statistical measure (second moment) to group the variables that contribute in a similar way in the covariance reconstruction. This unsupervised procedure endeavors to find information that facilitates the connection between the representation variables and the input data. The third level optimizes a cost function based on the measurement of data separability; in order to find the sets of variables that improve classification precision. Cluster analysis is followed by the optimization of a projection index in order to obtain 2-D visualization, and thus, verify the performance of the dimension reduction algorithm in terms of the discriminant capability for detecting ischemia using ECG signals.

II. STATE OF THE ART

Recent studies on cardiac disease detection show that accurate ischemic episode detection in the recorded ECG is based on the correct classification of the ischemic cardiac beats [19]. For the problem of maximizing the performance in the detection of ischemia episodes, the state-space consists of regions that lie near class separation boundaries, and require the construction of discriminant complexes, while for the remaining regions, the classification task is significantly simpler [20]. Computer-based techniques can be grouped depending on the computational paradigm on which they are based (e.g., rule-based expert systems, artificial NNs, pattern recognition, etc.) [15]. A Karhunen Loève (KL) transform-based algorithm was developed in [21] for robust automated detection of transient ST segment episodes. The algorithm incorporates a single-scan trajectory-recognition technique in feature space using the Mahalanobis distance function between the feature vectors. A pattern vector may be represented as a linear combination of uncorrelated KL basis functions. On the other hand, NNs have been widely used as pattern and statistical classifiers in many application areas, including medicine [22]. A supervised NN-based algorithm was used in [23] for the automated detection of ischemic episodes resulting from ST segment elevation or depression. The NN training was carried out by an adaptive backpropagation (BP) algorithm, which drastically reduced training time. In another approach

cited in [18], the network was trained using a Bayesian regularization method, and in order to reduce the dimensionality of the input pattern, a PCA was employed. However, in terms of connection with the subjacent physiological phenomenon, NNs have an inherently high degree of complexity, resulting in a model that is difficult to interpret. A knowledge-based technique for the automated detection of ischemic episodes can provide some level of interpretation. Rule-based methods exhibit certain advantages such as direct transformation of medical knowledge to rules, low computational load, and explanation of the diagnostic decisions. However, their diagnostic value depends on the appropriate selection and combination of the rules and on the method of feature value extraction used in the rules [15]. Supervising network self-organizing map (sNet-SOM) model allows computationally effective solutions to be designed for the particular ischemic detection problem [20]. Specifically, the sNet-SOM uses unsupervised learning for the simple regions and supervised for the difficult ones in a two-stage learning process. This approach proceeds until the total number of training patterns that are mapped to neurons with high entropy (ambiguous classification) are reduced to a manageable size with a capable supervised model. The second learning phase (supervised training) has the objective of constructing better decision boundaries at the ambiguous regions [20]. The multicriteria classification issue, known as a sorting problematic [24], deals with assigning patterns to one or several categories. The assignment is achieved in [19] through the examination of the intrinsic value of the pattern by referring to preestablished rules. A category can be defined either by one reference pattern prototype (being a monoprofile category) or by a set of patterns (a multiprofile one). The latter is harder to treat, but it offers greater flexibility as far as category modeling is concerned. Specifically, the procedure deals with assigning cardiac beats to predefined multiprofile categories (ischemic or not), which are not ordered, and this implies that there are no boundaries defining each category (this problem is called nominal sorting) [25]. Adaptive systems are of great interest in many applications, due to their flexibility and the capability to adjust them to the user, thus achieving better performances. The adaptation process consists of adapting a standard system, initially conceived for any user, to a particular user, considering his/her characteristics. The only vestige of adaptation that some systems present is threshold adjustment, which can be performed manually or automatically [26]. Other systems use classifiers based on NNs [23], for which the adaptation process is time-consuming and hard to implement. Hidden Markov models (HMMs) are also employed, which have been shown to be suitable for adaptation, due to the low complexity and the suitability of their structure in implementing online adaptation. However, an adaptation process can be implemented to adapt the HMMs to the ECG signal of a particular individual [27].

III. FEATURE VISUALIZATION: PROPOSED APPROACH

Effective pattern representation (visualization) involved in physiological behavior from feature space analysis, is generally obtained through several approaches based on dimensionality

reduction techniques. In this sense, it is important to define an evaluation function for directing the reduction context (relevance) [6].

A. Relevance Analysis

Relevance analysis distinguishes variables that are effectively representing the subjacent physiological phenomena according to some evaluation measure, named *relevant features*; moreover, it tries to reject variables that have repeated information (*redundant features*), and the ones whose contribution to the representation target is considered as null (*irrelevant features*).

Let $\boldsymbol{\Omega}$ be an object set where each one is composed by a set of random variables $\boldsymbol{\xi} = \{\xi_1, \xi_2, \ldots, \xi_p\}$ and is associated to one and only one class label from the label class set $\boldsymbol{c} = \{c_k | k \in K\}$. Thus, $\xi_i : \boldsymbol{\Omega} \to \mathbb{R}$, and it is defined as $c : \boldsymbol{\Omega} \to \{c_k | k \in K\}$, therefore, if $\boldsymbol{a} \in \boldsymbol{\Omega}$ then $\mathbf{x} = [\xi_1(\boldsymbol{a}), \xi_2(\boldsymbol{a}), \ldots, \xi_p(\boldsymbol{a})]$ and $c(\boldsymbol{a}) \in \boldsymbol{c}$.

Definition 3.1 (Generalized relevance): Let $\boldsymbol{\xi}$ be a variable set, $\varphi(\boldsymbol{\xi})$ parts of $\boldsymbol{\xi}$ and $\mu : \varphi(\boldsymbol{\xi}) \to \mathbb{R}$ a function such that $\forall \check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$ and $\mu(\check{\boldsymbol{\zeta}}) \leq \mu(\boldsymbol{\xi})$. Let $[\boldsymbol{\xi}] = \left\{ \check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi}) : \mu(\check{\boldsymbol{\zeta}}) = \mu(\boldsymbol{\xi}) \right\}$ be an equivalence class, since $\boldsymbol{\xi}$ is a finite set, then $[\boldsymbol{\xi}]$ is finite too. Therefore, $[\boldsymbol{\xi}] = \left\{ \hat{\boldsymbol{\zeta}}_i \right\}_{i=1}^{L}$ and every $\hat{\boldsymbol{\zeta}}_i \in [\boldsymbol{\xi}]$ is named **relevant variable subset** according to a function μ.

Remark 3.1 (Relevance stratification): If $L = 1$, all variables of $\boldsymbol{\xi}$ are relevant. If $L > 2$, and taking into account that $\boldsymbol{\xi}$ is a finite set, the subsets of $\hat{\boldsymbol{\zeta}}_i$ can be ordered by inclusion, obtaining $\{\boldsymbol{\zeta}_i\}_{i=1}^{d}$, which is the set of minimal subsets. The procedure used for discriminating relevant variables from the minimal subsets, which according to a function μ, captures the whole information (or nearly) of $\boldsymbol{\mu}$ without admitting repeated information, is called **strong relevance**. In fact, the *strongly relevant features* are within the intersection of all minimal subsets. On the other hand, the procedure for discriminating the variables that are contained in minimal subsets $\boldsymbol{\zeta}_i$, and their information can be inferred by others, is named **weak relevance**. In fact, the union of all minimal subsets subtracting the intersection conforms the subset of *weakly relevant features* of $\boldsymbol{\xi}$ according to μ.

B. Evaluation Function

The evaluation function involved in pattern recognition tasks and effective data representation is directly associated to the relevance function μ. According to this, the following notation will be used: given an observation set $\mathbf{X} = \{\mathbf{x}_j\}_{j=1}^{n}$ where each observation has one and only one class label assigned from the class label set $\boldsymbol{c} = \{c_k | k \in K\}$ and a relevance measure μ is imposed on a variable subset $\check{\boldsymbol{\zeta}} \in \varphi(\boldsymbol{\xi})$; if this relevance measure takes into account the class label information of each observation, this measure will be referred to as *evaluation function,* and can be written as $f_{\check{\boldsymbol{\zeta}}} \equiv f(\boldsymbol{c}, \check{\boldsymbol{\zeta}})$. Therefore, every $f_{\check{\boldsymbol{\zeta}}}$ has associated the relevance measure $\mu(\check{\boldsymbol{\zeta}})$ in the sense of Definition 3.1.
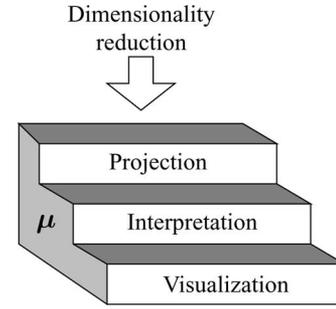
Fig. 1.   Operative levels of dimensionality reduction.

C. Operative Levels of Dimensionality Reduction

The goal of dimensionality reduction is to find a representation of that manifold (a coordinate system) that will allow the data vectors to be projected on it, and obtain a low-dimensional and compact representation of the data. The general purpose for reducing the dimension is the desire to preserve most of the relevant information of the original data according to some optimality criterion (evaluation function) using a relevance measure μ to direct the representation context.

Definition 3.2 (Dimensionality reduction): Given an observation set $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^p : j : 1, \ldots, n\}$ where each observation $\mathbf{x}_j$ is associated to the p-dimensional feature set $\boldsymbol{\xi}$. The term **reduced set of features** is given to the q-dimensional set $\boldsymbol{\zeta}$, $q < p$ is as small as possible, which is associated to the representation space $\mathbf{Z}$ under the following conditions.

1) There is a dimensionality reduction mapping G

$$G : \mathbf{X} \to \mathbf{Z}$$

$$\boldsymbol{\xi} \mapsto \boldsymbol{\zeta} = \mathrm{G}\{\boldsymbol{\xi}\}.$$

2) There is a smooth and nonsingular reconstruction mapping g

$$g : \mathbf{Z} \to \mathbf{Y} \subset \mathbf{X}$$

$$\boldsymbol{\zeta} \mapsto \tilde{\boldsymbol{\xi}} = g\{\boldsymbol{\zeta}\}.$$

Moreover, the following constraints must be considered.

1) The manifold $\mathbf{Y} \overset{\triangle}{=} g\{\mathbf{Z}\}$ approximately contains all the observation points: $\{\mathbf{x}_j\}_{j=1}^{n} \overset{\subseteq}{\sim} \mathbf{Y}$.

2) Given a metric $m : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^+ \cup \{0\}$ and taking into account that $\tilde{\mathbf{x}}_j = g\{G\{\mathbf{x}_j\}\}$, a distance $\boldsymbol{d} = \{d_j : d_j = m(\tilde{\mathbf{x}}_j, \mathbf{x}_j)\}$ can be obtained. There is a function $\varepsilon : \mathbb{R}^n \to \mathbb{R}$, which computes the *reconstruction error* $\varepsilon(\boldsymbol{d})$ between $\mathbf{X}$ and $\mathbf{Y}$, where it is required that $\varepsilon(\boldsymbol{d}) \to 0$.

According to Definitions 3.1 and 3.2, the dimensionality reduction task can be seen as an optimization routine

$$\mathbf{Z} = f(\boldsymbol{c}, \mathbf{X}) \to \underset{\mu(\varphi(\boldsymbol{\xi})), [\boldsymbol{\xi}]}{\text{Optimize}} \ f_{\check{\boldsymbol{\zeta}}}. \tag{1}$$

Interpretability of multidimensional spaces using dimensionality reduction can be taken in three layers defined according to their operative capability (see Fig. 1): projection, interpretation, and visualization.

1) Level 1 (Projection): High-dimension data can be projected in a lower dimension space by a transformation function that optimizes a specific criterion in the resulting space [6]. Literature usually refers to the finding of this image space as feature extraction [28]. The aim of this reduction level is to represent data in a lower dimension space in order to decrease computational complexity and increase classification accuracy. The following techniques are commonly used for mapping data on a new space where their properties can facilitate pattern recognition: PCA, ICA, Hermite expansions (HEs), and WT, among others. In this study, PCA and WT [29] were used as transformation principles of the data in order to carry out the feature extraction stage.

2) Level 2 (Interpretation): This level constitutes a deeper layer in dimension reduction, as the purpose is not only to search for a representation space of lower dimension, but also to exert some type of distinction between the input variables or to establish the connection between the output and input space, thus determining which are the variables of the initial feature space that contribute effectively to the representation. This subset taken from the initial variable set is known as *relevant features*, and this level is frequently known as *feature selection*. In general, given a set of observations where each observation is associated to one and only one class label from the class label set c, the problem of feature selection is to choose the subset $\zeta \subset \xi$ of size q from a set of p features where optimal and suboptimal strategies can be considered. For the optimal case, if the cardinal of ζ is q, and all the q-cardinal subsets are in $\breve{\xi} = \left\{ \breve{\zeta}_i \subset \xi : i = 1, \ldots, M \right\}$, the subset ζ is that which optimizes an evaluation function $f_{\breve{\zeta}_i}$, such as $f(c, \zeta) = \max_{\breve{\xi} \subset \xi} f(c, \breve{\zeta}_i)$ [30]. When $f_{\breve{\zeta}_i}$ uses information of the classification in order to minimize the classifier error, this procedure is known as *wrapper-type selection*, whereas it is called *filter-type selection* when the procedure consists of data preprocessing by the optimization of $f_{\breve{\zeta}_i}$ with regards to a metric (independent of the classification results) for rejecting irrelevant and redundant variables [31]. Popular suboptimal algorithms such as sequential forward search (SFS) and sequential backward search (SBS) start from the empty/complete set where step by step new subsets are generated by adding/discarding a feature selected according to some evaluation function. In floating search methods, SFFS and SBFS, the number of added/removed features can also change at each step, and these wrapper routines carry out the search in a considerably smaller number of subsets [1]. Algorithms that carry out data clustering also belong to this level, since they allow the representation of a determined set of data in a reduced set of clusters based on a metric. *Iterative self-organizing data (Isodata) analysis techniques* are considered amongst the most well known, where the cost function $f_{\breve{\zeta}_i}$ includes a spatial metric to group sufficiently near elements.

3) Level 3 (Visualization): Data projection/interpretation, which carries out dimensionality reduction of the multivariate data to a small dimension by a single or a combination of the different methods mentioned before, is followed by visualization of the resulting feature space. Visual representation helps

to acquire knowledge of the multidimensional space structure. In this case, dimension reduction is more demanding, since the representation space does not surpass the three dimensions, besides capturing the corresponding structure in the most accurate way possible. In recent literature different methods like scatter plot [32], S-Isomap [11], and weight watcher [2] are reviewed to obtain an interpretive visual representation of multidimensional data. In general, several techniques can be used in visualization applications although under strongly constrained conditions. The problem consists of adequately defining the relevance target. In this study, the feature visualization concept is taken from a more demanding point of view since the corresponding structures of multidimensional spaces and the connection between visualization spaces, and subjacent physiological phenomena are sought after. Taking into consideration the constraint 2) of Definition 3.2, this dimensionality reduction oriented towards feature visualization is subject to the following condition.

Condition 3.1 (Feature visualization): Visualization of information embedded in high-dimensional spaces and connected with subjacent physiological phenomenon can be obtained by dimensionality reduction procedures as long as there is a smooth and nonsingular reconstruction mapping g where $\varepsilon \to 0$.

In fact, from Condition 3.1, it is possible to build visualization procedures pursuing the following mathematical scheme given by the expression

$$\mathbf{Z} = f_r \circ f_{r-1} \circ \cdots \circ f_2 \circ f_1 (c, \mathbf{X})$$
$$\text{such that } \exists f_1^{-1}, f_2^{-1}, \ldots, f_{r-1}^{-1}, f_r^{-1} \quad (2)$$

where each f is an evaluation function with a relevance measure μ associated.

IV. EXPERIMENTAL OUTLINE

A. Databases

1) BD-UNAL-GCPDS-ECG: The ECG recordings of this database belong to the Universidad Nacional de Colombia. This database is made up of 189 ECG recordings of 12 leads during 10 min, taking into account the recommendations of the European Society of Cardiology. They were taken from 91 men between 29 and 75 years of age, and from 98 women aged between 45 and 75 years. The annotations were carried out by three cardiologists in order to obtain a representative selection of ischemic episodes following the same criteria used in the E-STT database, such as displacements of the base line and the ST$-$T complex due to hypertension, ventricular dyskinesia, and medication effects. The device used was *cardio card PC-based ECG/resting PC ECG/PC EKG* system made by Nasiff Associates. This acquisition card was used under the following operational parameters: Sampling frequency $f_s = 500$ Hz, 13 bits of resolution, rank of 10 mV, three channels, and 12 leads. From the complete set of recordings from the database, 27 recordings of 12 leads corresponding to normal subjects (men and women of similar ages) were chosen, as well as 27 recordings of 12 leads of patients diagnosed with episodes of ischemic cardiopathy (men and women of similar ages). From each recording,
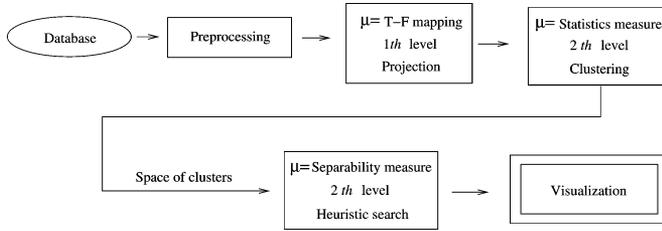
Fig. 2.   Block diagram of the proposed procedure.

TABLE I
SPECTRAL REPRESENTATION USING WT

WT	N-order
Daubechies	2-10
Symlet	2-10
Biorthogonal	13, 15, 24, 26, 28, 55, 68
Reverse Biorthogonal	22, 24, 26, 28, 44, 55, 68

Algorithm 1 Feature visualization by clustered loads

Require: $\mathbf{X}_{n \times p}$.
1) Estimate the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$.
2) Find the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ from $\boldsymbol{\Sigma}_{\mathbf{X}}$.
3) Estimate the q value $\star$\number of eigenvalues calculated for an accumulated variance of 99%\$\star$
4) $\mathbf{A}_q \leftarrow q$ eigenvectors associated to the q greatest eigenvalues by columns.
5) The rows of $\mathbf{A}_q$ are taken as $\mathbf{v}_i$, with $i = 1, \ldots, p$.
6) $\boldsymbol{v}_j = \text{Isodata}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ with $j = 1, \ldots, m$. $\star$\feature clustering into m clusters\$\star$
7) $\hat{\boldsymbol{\xi}} = [\mathrm{E}\{\boldsymbol{v}_1\}, \mathrm{E}\{\boldsymbol{v}_2\}, \ldots, \mathrm{E}\{\boldsymbol{v}_m\}]$ $\star$\each column of this matrix is obtained by estimation of the expected value corresponding to each cluster\$\star$
8) $\varsigma = \text{SFFS}\left\{\varphi\left(\hat{\boldsymbol{\xi}}\right)\right\}$, using a k-NN classifier as evaluation function. $\star$\$\varphi\left(\hat{\boldsymbol{\xi}}\right)$ means all possible subsets built from $\hat{\boldsymbol{\xi}}$\$\star$

$\varsigma = \{$ effective feature set $\}$
Algorithm validation: k-NN classifier (50% for training and 50% for validation).

---

several averaged beats were obtained (averages of 20 beats approximately); therefore, yielding a set of 900 normal beats and 900 beats with evidence of ischemia.

2) BD-ESTT: The European ST−T database is intended to be used for evaluating algorithms by analyzing ST and T-wave changes. This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. The subjects were 70 men aged 30–84, and eight women aged 55–71. Myocardial ischemia was diagnosed or suspected in each subject. Each record is two hours in duration and contains two signals, each sampled at 250 samples/s with 12-bit resolution over a nominal 20 mV input range. The sample values were rescaled after digitization with reference to calibration signals in the original analog recordings, in order to obtain a uniform scale of 200 ADC units/mV for all signals [33]. For this study, 1800 representative beats were chosen: 900 considered normal and 900 beats with evidence of ischemia.

B. Proposed Procedure: Clustered Loads

Fig. 2 shows the proposed procedure, where combinations of the techniques previously mentioned, corresponding to different operative levels of dimensionality reduction, were used in order to obtain feature visualization, taking into account Condition 3.1 and expression (2).

1) Preprocessing: Before extracting the information contained in the electrocardiographic recordings, procedures were applied to reduce the effects of electrical interference, patient movement, and breathing, using the WT of the *Daubechies* family, in the following way: electrical interference (*Daubechies* 3, approximation level 3), fluctuation of base line caused by breathing (*Daubechies* 4, approximation level 8), and electromyographical noise (*Daubechies* 3, approximation level 4). The method used to determine the location of QRS was the intersection of WT through zero. This is because the maximum module and the intersections through zero of the WT correspond to fine edges of the signal, and the QRS complex produces two maximum modules with opposed signs intersecting each other through zero [34]. Therefore, these are determined by applying decision rules (thresholds) to the WT of the ECG signal. After detecting the QRS complexes, every beat from each of the recordings is segmented [35].

2) Feature Space: Representation space was generated using wavelet analysis. Since there is an abundant variety of wavelet functions, conformation of feature space was established by the collection of wavelet coefficients corresponding

to different families (as shown in Table I) in two ways: 1) coefficients of the third approximation level and 2) decomposition up to level 6, and in each approximation scale, the four coefficients with maximum absolute value, in order to verify the performance of the proposed algorithm in the identification of redundant variables, since variation in terms of the representation offered by the different wavelets should not be noticeable. The total extracted features were 840. The observations were separated in two groups of equal size for calibration and validation (each matrix has dimensions 900×840), using a method of random election.

3) Feature Visualization: This procedure is detailed in Algorithm 1, where the clustered features are analyzed with the purpose of establishing a pattern of similarity that will reveal interpretation parameters that will help clinical knowledge or the adjustment of parameters associated with the processing techniques used. Specifically, the representation of n ECG recordings is obtained by estimating p features that belong to $\boldsymbol{\xi}$ and are located in the matrix $\mathbf{X}_{n \times p}$, where the columns contain the features and the rows the training recordings. The Isodata procedure groups the most similar vectors $\mathbf{v}_i$, that compose the matrix $\mathbf{A}_q$, with the aim of discovering the loads that equally contribute in the covariance reconstruction. Thus, indirectly it is possible to group the redundant features and to reduce the dimensionality taking only one component per group. On the other hand, the SFFS algorithm is taken as a second-level routine of *wrapper*-type, to reduce dimensions by selecting the components that offer higher separability between the classes. The variable $\hat{\boldsymbol{\xi}}$ is the set of components that represents the initial features by clusters, and $\varphi\left(\hat{\boldsymbol{\xi}}\right)$ indicates all possible subsets built from $\hat{\boldsymbol{\xi}}$,
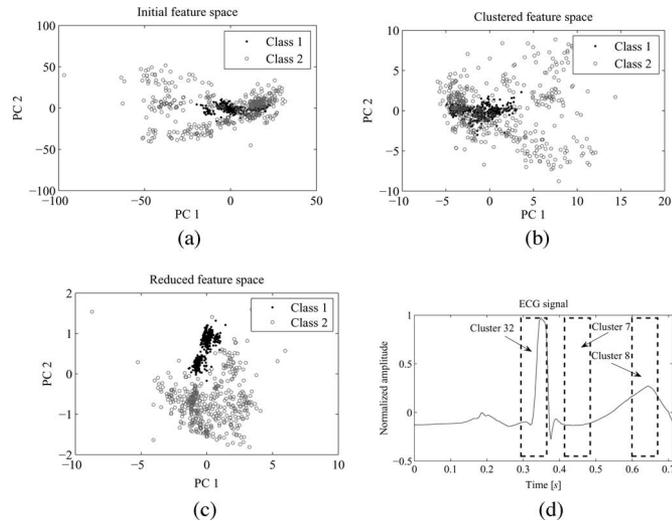
Fig. 3. Results in the different stages. (a) Principal components of $\mathbf{X}$. (b) Principal components of $\hat{\boldsymbol{\xi}}$. (c) Principal components of ς. (d) Connection with ECG trace.

TABLE II
CONFUSION MATRIX

Beat class	BD-UNAL-GCPDS-ECG		BD-ESTT	
	Normal	Ischemic	Normal	Ischemic
Normal	449	3	448	5
Ischemic	1	447	2	445

TABLE III
S, *P*, AND Sp RESULTS

	BD-UNAL-GCPDS-ECG	BD-ESTT
S (%)	99.8	99.6
P (%)	99.3	98.9
Sp (%)	99.3	98.9

where an evaluation function f_ς based on the k-nearest neighbor rule is used in the search for the most representative subset with regards to separability. Thus, there is inverse mapping for each transformation stage associated to μ, in order to assure result interpretation by feature visualization. It is important to mention that ischemic episodes in terms of ischemic beats are defined according to the same set of criteria as those used in [20]. Correctly detected episodes are termed true positive (TP) episodes. Missed episodes are termed false negatives (FNs). Also, when a nonischemic episode is correctly detected it is called true negative (TN) and if it is detected as ischemic, a false positive (FP) situation has occurred. There are important indexes for evaluating the detection capability of ischemic episodes as sensitivity (S), predictivity (P) and specificity (Sp) that are defined as $S = (\text{TP})/(\text{TP} + \text{FN}) \times 100\%$, $P = (\text{TP})/(\text{TP} + \text{FP}) \times 100\%$, and $\text{Sp} = (\text{TN})/(\text{TN} + \text{FP}) \times 100\%$, respectively.

V. RESULTS AND DISCUSSION

Using the BD-UNAL-GCPDS-ECG database, a projection matrix $\mathbf{A}_q$ with $q = 15$ principal components was obtained for 99% of the data variability (840 features represented by 15 dimensions). In Fig. 3(a), the representation of the initial training space is shown by the projection in two principal components. It can be noted that the observations for each one of the classes have a geometric disposition that makes pattern separability difficult. The Isodata algorithm grouped the set of features into 34 clusters $\mathbf{v}_i$, with which a new $\hat{\boldsymbol{\xi}}$ matrix was made, which has the same number of observations as the training set but with 34 columns, each one of them being the mean of the clustered features. The representation by principal components of the new $\hat{\boldsymbol{\xi}}$ space is shown in Fig. 3(b). After applying the SFFS algorithm on $\hat{\boldsymbol{\xi}}$, it was found that a set of three variables is sufficient for class separation (the classifier used as cost function was a 5-NN). The reduced set of features obtained was $\{\varsigma = \mathbf{v}_7, \mathbf{v}_8, \mathbf{v}_{32}\}$, where these indexes correspond to the columns of the $\hat{\boldsymbol{\xi}}$ ma-

trix. The reduced feature space ς is shown by projecting the data in the plane made up by the two principal components in Fig. 3(c). When comparing Fig. 3(b) and (c), the effect of feature selection can be observed when a reduced space is obtained where separability of the observations is visible. After achieving the reduced space, for example, taking the BD-UNAL-GCPDS-ECG database, the connection of the WT features belonging to clusters 7, 8, and 32, was made with the time representation of the ECG signal. In Fig. 3(d), the regions of the ECG signal can be seen in which most of the coefficients are concentrated, corresponding to clusters that comprise the reduced space ς.

For the training and validation groups, matrices ς_t and ς_v of 900×3 dimensions were created by random selection, with which the confusion matrix for the 5-NN classifier showed in Table II was obtained. It is important to mention that by using of the BD-ESTT database similar results were found. Thus, the classification accuracy obtained was 99.5% for the BD-UNAL-GCPDS-ECG database and 99.2% for the BD-ESTT database, under simple validation criterion (no validation observation was used in the training procedure). Additionally, the sensitivity, predictivity, and specificity results are presented in Table III. In general, the approach proposed in this paper allows a reduced representation space to be built in order to distinguish different ischemic beats from a nonlabeled beat set. Fig. 4 shows the averaged normal shape of an ECG signal and some examples of ischemic episodes taken into account in this study. This procedure can be improved by changing the representation target of the loads for coefficient vectors, thus obtaining a final space composed of statistically independent variables, since this analysis is less constrained and has the possibility of carrying out blind analysis without requiring a lot of prior information.

A. Interpretation and Visualization of Results

From analysis of clustered loads, clusters associated with the resulting components of ς, revealed that not only the time–frequency representation of the ST−T complex contains representative information of abnormalities related to ischemic events, but in the bands corresponding to the QRS complex, there is also discriminant information for the automatic detection of ischemia [see Fig. 3(d)]. Although it was not very
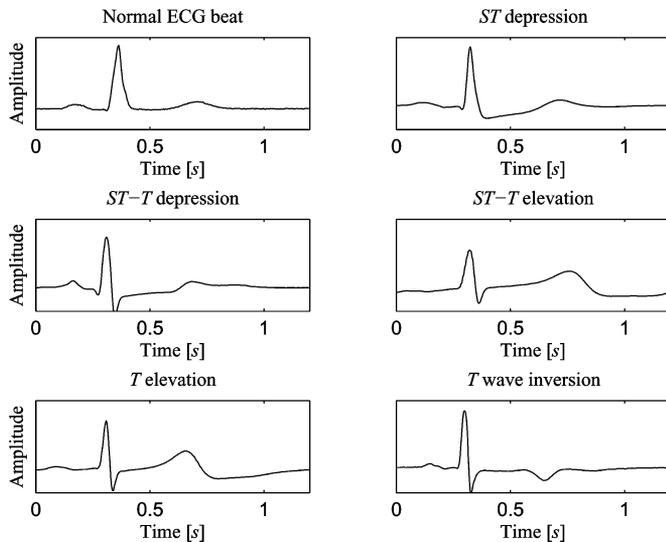
Fig. 4. Example of normal and ischemic averaged beats.

Approach	S (%)	Sp (%)
GA & MCDA [19]	91	91
ANN & PCA [18]	90	90
ANN & Parametric modeling [13]	81	84
ANN Backpropagation [23]	89	78
Incremental HMM [27]	86	85
Karhunen–Loève & Mahalanobis [21]	87	88
Rule-based [15]	70	63
Clustered loads (*Current work*)	99	99

noticeable in the signals chosen for training, this could be due to the changes displayed by the Q-wave as one of the manifestations of an infarct, which demonstrates that wavelets have sufficient representation power to detect changes in the ECG signal caused by ischemic problems. Moreover, the results confirm that the analysis can be made using a single type of wavelet, as long as the wavelet offers suitable representation corresponding to the complexes ST−T and QRS, since the information of these time-frequency bands is contained in ς. This means that construction of the representation space using several types of wavelets entails in itself correlated variables, and in terms of separability, different wavelets have similar contributions.

B. Comparison With Other Models

Using the same feature set on the databases described in Section IV-A, this algorithm was compared with previously developed models, and is reported in [6] and [35]. The first model of comparison includes the induction of a decision tree (weak relevance) for rejecting irrelevant features. Tuning of the selection model was carried out by GAs simultaneously seeking to minimize the number of features, as well as the classification error. Moreover, computational complexity is considerably higher in comparison with the inherent cost of the procedure proposed in this study (relation 30:1), as well as a less effective dimension reduction (13 variables). This approach was also compared with a model that in the first stage develops separability analysis by means of hypothesis tests, followed by an analysis of linear independence by means of correlation analysis, and finally, feature selection using a heuristic search algorithm based on distances, known as *Relief*. Dimension reduction to 16 variables, greater computational complexity (relation 10:1) and lower classification precision was obtained. Finally, it was compared with a multivariate model based on a feature selection algorithm (SFFS) which uses MANOVA algorithm as an evaluation function, achieving a reduction to seven variables, greater computational complexity (relation 10 : 1), and comparable clas-

sification precision. These last two algorithms strongly depend on the stage of data preprocessing (removal of anomalous values, verification of normality, and transformation of distributions), whereas the algorithm proposed in this paper displays better performance without this type of data preprocessing. Additionally, the approach presented in this study offers the capacity of interpreting the results, when the analysis of clusters is allowed revealing the nature of the information contained by the features. On the other hand, the BD-ESTT database was probed in order to compare to other system performances on the same database reported in the literature (see Table IV), where the proposed approach was compared to other approaches in terms of beat classification, such as a combination between multicriteria decision analysis and GAs [19], PCA and feed-forward NNs [18], parametric modeling combined with feed-forward NNs [13], adaptive BP NNs [23], incremental HMM [27], KL transform with Mahalanobis distance [21], and rule-based detection [15].

VI. CONCLUSION

Dimension reduction taken according to the clustered loads scheme shown in Fig. 2, facilitates interpretation and visualization of spaces with high dimensionality. Additionally, feature visualization reveals clinical patterns (connection with subjacent physiological phenomenon) that help to elect the smallest possible set of features that best represent all the cardiac information contained in the ECG recording.

The proposed procedure for the automatic detection of ischemic events, through the information contained in electrocardiographic signals, consisted of the following steps.

1) Mapping of the time–frequency space by means of the WT.
2) Using a statistical metric, a projection to the principal components space was obtained and, using an algebraic metric, a clustering of features was carried out from the loads matrix.
3) Each cluster was represented by the average (statistical metric), then a wrapper-type heuristic search algorithm SFFS was used to find the variables that offered more separability (geometrical metric), and were projected to a 2-D space, maximizing a projection index based on the variability of the data (PCA), therefore obtaining the visualization.

This study revealed that not only the temporal and low-frequency structure of the ST−T complex offers discriminant

information in ischemic events, but also the spectral variations that are reflected in the QRS complex. Thus, the interpretation restricted the feature extraction to the frequency bands in which the changes that contribute to class discrimination occur.

The proposed dimensionality reduction scheme, shown in Fig. 1, permits the analysis of multidimensional data through relevance measures associated to an evaluation function (taking into consideration the inherent constraints with regards to strong or weak relevance). It is important to note that the success of the analysis strongly depends on the adequate election of these relevance measures. In this sense, the scheme facilitates the creation of procedures for finding subsets of representative variables with regards to a relevance measure, which is involved in improving the representation capability of the reduced variable set. Finally, this scheme connects the solution subspace and the initial space with the purpose of finding patterns that can be related to the nature of the characterized signals, in two ways: cardiac functionality patterns and patterns that reveal hidden information, which is discriminant and consistent for pattern recognition.

REFERENCES

[1] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2004.

[2] A. König, "Dimensionality reduction techniques for multivariate data classification, interactive visualization, and analysis-systematic feature selection vs. extraction," in *Proc. 4th Int. Conf. Knowl.-Based Intell. Eng. Syst. Allied Technol.*, Brighton, U.K., Sep. 2000, pp. 44–55.

[3] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. ed. New York: Spring-Verlag, 2002.

[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis* (Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications and Control). New York: Wiley, 2001.

[5] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd ed. London, FL, U.K./Boca Raton: Chapman & Hall/CRC Press, 2000.

[6] L. F. Giraldo, E. Delgado, J. C. Riaño, and G. Castellanos, "Feature selection using hybrid evaluation approaches based on genetic algorithms," in *Proc. IEEE Comput. Soc. Press Proc., Electron., Robot. Automotive Mech. Conf. (CERMA 2006)*, vol. 2, pp. 242–247.

[7] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognit.*, vol. 33, pp. 25–41, 2000.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by local linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[10] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas, "Non-linear dimensionality reduction techniques for classification and visualization," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2002, pp. 645–651.

[11] X. Geng, D.-C. Zhan, and Z.-H. Zhou, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1098–1107, Dec. 2005.

[12] I. Pitas, M. G. Strintzis, S. Grippas, and C. Xerostylides, "Machine classification of ischemic electrocardiograms," presented at the IEEE Mediterranean Electrotech. Conf., Athens, Greece, 1983.

[13] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis, "An expert system for ischemia detection based on parametric modeling and artificial neural networks," in *Proc. Eur. Med. Biol. Eng. Conf.*, 2002, pp. 742–743.

[14] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 2, pp. 167–173, Mar./Apr. 1995.

[15] C. Papaloukas, D. I. Fotiadis, A. Likas, A. P. Liavas, and L. K. Michalis, "A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms," *Med. Biol. Eng. Comput.*, vol. 39, pp. 105–112, 2001.

[16] C. Papaloukas, D. I. Fotiadis, A. Likas, C. S. Stroumbis, and L. K. Michalis, "Use of a novel rule-based expert system in the detection of changes in the ST segment and the T wave in long duration ECGs," *J. Electrocardiol.*, vol. 35, no. 1, pp. 27–34, 2002.

[17] T. Stamkopoulos, K. Diamantaras, N. Maglaveras, and M. Strintzis, "ECG analysis using nonlinear PCA neural networks for ischemia beat detection," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3058–3067, Nov. 1998.

[18] C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis, "An ischemia detection method based on artificial neural networks," *Artif. Intell. Med.*, vol. 24, no. 2, pp. 167–178, 2002.

[19] Y. Goletsis, C. Papaloukas, D. I. Fotiadis, A. Likas, and L. K. Michalis, "Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 10, pp. 1717–1725, Oct. 2004.

[20] S. Papadimitriou, S. Mavroudi, L. Vladutu, and A. Bezerianos, "Ischemia detection with a self-organizing map supplemented by supervised learning," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 503–515, May 2001.

[21] F. Jager, G. B. Moody, and R. G. Mark, "Detection of transient ST segment episodes during ambulatory ECG monitoring," *Comput. Biomed. Res.*, vol. 31, no. 5, pp. 305–322, 1998.

[22] A. S. Miller, B. H. Blott, and T. K. Hames, "Review of neural network applications in medical imaging and signal processing," *Med. Biol. Eng. Comput.*, vol. 30, no. 5, pp. 449–464, 1992.

[23] N. Maglaveras, T. Stamkopoulos, C. Pappas, and M. G. Strintzis, "An adaptive backpropagation neural network for real-time ischemia episodes detection: Development and performance analysis using the european *ST–T* database," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 7, pp. 805–813, Jul. 1998.

[24] B. Roy, *Multicriteria Methodology for Decision Aiding*. Amsterdam, The Netherlands: Kluwer, 1996.

[25] J. Léger and J. Martel, "A multicriteria assignement procedure for a nominal sorting problematic," *Eur. J. Oper. Res.*, vol. 138, no. 2, pp. 349–364, 2002.

[26] J. P. Martínez, R. Almeida, S. Olmos, A. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[27] R. V. Andreao, S. M. T. Muller, J. Boudy, B. Dorizzi, T. F. Bastos Filho, and M. Sarcinelli Filho, "Incremental HMM training applied to ECG signal analysis," *Comput. Biol. Med.*, vol. 38, pp. 659–667, 2008.

[28] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2006.

[29] S. Pittner and S. V. Kamathi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 83–88, Jan. 1999.

[30] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. New York: Wiley, 2002.

[31] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learning Res.*, vol. 5, pp. 1205–1224, 2004.

[32] P. Craig and J. Kennedy, "Coordinated graph and scatter-plot views for the visual exploration of microarray time-series data," in *Proc. IEEE Symp. Inf. Vis.*, 2003, pp. 173–180.

[33] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European *ST-T* database: Standard for evaluating systems for the analysis of *ST-T* changes in ambulatory electrocardiography," *Eur. Heart J.*, vol. 13, pp. 1164–1172, 1992.

[34] J. S. Sahambi, S. Tandon, and R. Bhatt, "Using wavelet transform for ECG characterization," *IEEE Eng. Med. Biol.*, vol. 16, no. 1, pp. 77–83, Jan./Feb. 1997.

[35] E. Delgado, J. L. Rodríguez, F. Jiménez, D. Cuesta, and G. Castellanos, "Recognition of cardiac arrhythmias by means of beat clustering on ecg-holter records," *Comput. Cardiol.*, vol. 34, pp. 161–164, 2007.

Edilson Delgado-Trejos (M'06) received the undergraduate degree in electronic engineering, the M.Sc. degree in industrial automation, and the Ph.D. degree in engineering sciences from the Universidad Nacional de Colombia, Manizales, Colombia, in 2000, 2003, and 2008, respectively.

Since August 2008, he has been a Professor and a Researcher in the Research Center of the Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia, where he is currently the Director of the Machine Intelligence and Pattern Recognition Group (MIRP). His current research interests include pattern recognition, machine learning, nonlinear analysis, digital signal processing, and industrial automation.

Peré Caminal-Magrans (M'90) was born in Barcelona, Spain, in 1952. He received the M.S. and Ph.D. degrees in mechanical engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1974 and 1980, respectively.

He is currently a Full Professor of automatic control in the Department of Control Engineering, UPC, where he is also the Director of the Biomedical Engineering Research Centre (CREB) and the Director of the master program on biomedical engineering. He is also with the Center for Biomedical Research Network in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Saragossa, Spain. His current research interests include biomedical signal processing in the areas of cardiovascular system, respiratory system, and genomics.

Alexandre Perera-Lluna received two B.S. degrees in physics and in electrical engineering, and the Ph.D. degree in physics from the Universitat de Barcelona, Barcelona, Spain, in 1996, 2001, and 2003, respectively.

From 2003 to 2005, he was with Texas A&M University as a Postdoctoral Research Associate. He is currently a Ramon y Cajal Researcher at the Center for Biomedical Research, Universitat Politècnica de Catalunya (UPC), Barcelona. He is also with the Center for Biomedical Research Network in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Saragossa, Spain. His current research interests include applied pattern recognition focusing on chemical sensors, machine olfaction, dimensionality reduction and bioinformatics.

Germán Castellanos-Domínguez (M'96) received the undergraduate degree in radiotechnical systems and the Ph.D. degree in processing devices and systems from the Nauchno Issledovatelsky Institut, Perm, Russia, in 1985 and 1990, respectively.

He is currently a Professor in the Department of Electrical, Electronic and Computer Engineering, Universidad Nacional de Colombia, Manizales, Colombia, where he is also the Director of the Control and Digital Signal Processing Group. His current research interests include information and signal theory, digital signal processing, and bioengineering.

Montserrat Vallverdú-Ferrer received the M.S. degree in electrical engineering and the Ph.D. degree in biomedical engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.

She is currently an Assistant Professor in the Department of Automatic Control, Biomedical Engineering Research Center, UPC. She is also with the Center for Biomedical Research Network in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Saragossa, Spain. Her current research interests include complexity analysis of the hidden information in biological systems.