Research Article

Relevance Analysis of Stochastic Biosignals for Identification of Pathologies

Lina María Sepúlveda-Cano, Carlos Daniel Acosta-Medina, and Germán Castellanos-Domínguez

Grupo de Procesamiento y Reconocimiento de Señales, Universidad Nacional de Colombia, Km. 9, Vía al Aeropuerto, Campus La Nubia, 17001000, Manizales, Colombia

Correspondence should be addressed to Lina María Sepúlveda-Cano, lmsepulvedac@bt.unal.edu.co

This paper presents a complementary study of the methodology for diagnosing of pathologies, based on relevance analysis of stochastic (time-variant) features that are extracted from *t-f* representations of biosignal recordings. Dimension reduction is carried out by adapting in time commonly used latent variable techniques for a given relevance function, as evaluation measure of time-variant transformation. Examples of both unsupervised and supervised training are deliberated for distinguishing the set of most relevant stochastic features. Besides, two different combining approaches for feature selection are studied. Firstly, when the considered input set comprises a single type of stochastic features, that is, having the same principle of generation. Secondly, when the whole input set of parameters is taken into consideration no matter of their physical meaning. For validation purposes, the methodology is tested for the concrete case of diagnosing of obstructive sleep apnea. Achieved results related to performed accuracy and dimension reduction are comparable with respect to other outcomes reported in the literature, and thus clearly showing that proposed methodology can be focused on finding alternative methods minimizing the parameters used for pathology diagnosing.

1. Introduction

Regarding to analysis and processing of a biological signal, termed biosignal, stochastic modeling has been under continuous formulation for extracting valuable information, when directly taking into account evolution of random biological variables along a given argument (time, space, etc.). Stochastic modeling assumes that each observation comprises an ordered sequence of measures (or trajectory) and may be intrinsically more difficult to assess if comparing to the standard scenario, where each measurement consists of a mere collection of unconnected values. Indeed, the use of stochastic modeling precedes the necessity of building a suitable method of processing. So, it is well known that the complexity of stochastic modeling increases because of the need to carry out the nonstationary estimation of parameters derived from biosignal recordings; one can refer to that issue as the most important difference between static and argument-variant statistical processing.

In this line of analysis, the time-frequency representation (TFR), planned to determine the energy distribution along the frequency axis at each time instant, has been proposed before to investigate the time-variant properties of the spectral parameters during either transient physiological or pathological episodes [1]. It has established the discriminating capability of frequency bands of biological activity between normal and pathological patterns, and for that reason, the set of TFR-based stochastic features to be considered should be suitable estimated by spectral subband methods. Nonetheless, if considering that classification is often more accurate when the pattern is simplified through representation by important features, then another specificity concerning the construction of stochastic modeling is the analysis of relevance of stochastic features as a justification for reducing

dimensionality. Also, there is a need to cope with the problem of either overtraining data or incrementing the computational overhead, which is particularly exasperating in those cases of multivariate TFR analysis. Extraction of relevant information from stochastic feature sets had been discussed in the past as a means to improve performance during and after training in learning processes. Namely, as pointed out in [2], to get an effective feature selection algorithm, in the context of an inference, two main issues are to be solved: the same measure associated to a given relevance function (i.e., a proper measure of distance among time series), and the multivariate transformation through the time axis, which is assumed to maximize the measure of relevance present in the features by their projection onto a new space [3].

In the present work, a methodology for diagnosing of pathologies is discussed that is based on relevance analysis of stochastic features extracted from *t-f* representations of biosignal recordings. Dimension reduction is carried out by adapting in time commonly used latent variable techniques, in such a way that the data information is maximally preserved, given a relevance function as evaluation measure of time-variant transformation, and therefore distinguishing relevant stochastic features. Specifically, since the maximum variance is assumed as a measure of relevance, time-adapted Principal Component Analysis (PCA) approach is developed. But PCA maximizes the variability on the data set, and for classification purposes, the attained components are not always related to the most discriminative information [4]. Thus, similar time-variant decomposition approach, but grounded on Partial Least Squares (PLS), is deliberated as a supervised multivariate transformation that yields components maximally related with labels. Besides, two different combining approaches for selecting the best set of contours based on relevance analysis are studied: firstly, when considering partly divided sets with a single type of stochastic features, having the same principle of generation. Secondly, when the best features are chosen among the whole set of considered time-variant parameters no matter of their physical meaning.

This paper is organized as follows: the time-varying spectral analysis is introduced, including a brief description of the extraction of stochastic features using either short-time fourier Transform or wavelet transform for calculation of TFR. Then, the methodology for stochastic-based relevance of extracted features is described in detail. Lastly, the effectiveness of a feature set representing the dynamics of the biological activity is illustrated for the concrete case of obstructive sleep apnea detection, followed by a discussion of the results and conclusions.

2. Background

2.1. Generation of TFR-Based Time-Variant Features.
Because of physical restrictions that are usually imposed on real random signals, rather than the linear time-frequency (*t-f*) representation, $\langle x, \phi \rangle$, estimation of time-dependant energy spectral density is carried out, which for a given

$x(t) \in \mathbb{R}(T)$ is directly calculated by the *spectrogram*, $\mathbf{S}_x(t, f)$

$$| \langle x, \phi \rangle |^2 = \left| \int_T x(\tau)\phi(\tau - t)e^{-j2\pi f\tau}d\tau \right|^2 \triangleq \mathbf{S}_x(t, f),$$
$$t, \tau \in T, \ \mathbf{S}_x(t, f) \in \mathbb{R}^+. \tag{1}$$

The commonly used TFR calculated by Short Time Fourier Transform (termed STFT) introduces a time localization concept by using a tapering window function, ϕ, of short duration going along the studied signal, x, but the window length remains constant for (1). Therefore, the extraction of information with fast changes in time (i.e., high frequency components) should be accomplished with short and well-timed localized intervals. In contrast, the low-frequency components have to involve large time intervals of analysis. As a result, given a nonstationary signal of relatively small time-bandwidth product, the STFT spectrogram might be not good enough for revealing properly the *t-f* dynamics. Certainly, this drawback is used to be partially solved by different windowing functions [5]. Instead, another worthy approach to analyze nonstationary signals is to expand them into basis functions, assembled from shifted and scaled versions of a given mother function, but keeping the energy concentrated in short intervals of the *t-f* plane (referred as *wavelets*). The Wavelet Transform (WT) gives an appropriate time resolution and poor frequency resolution at high frequencies, but providing a good frequency resolution at low frequencies. This approach makes sense, especially, when the signal has high-frequency components during a short period, while exhibiting low-frequency components for large intervals, which is true in most cases of biological signals.

Basically, some TFR-based short-time parameters have been widely accepted for characterizing the biosignals that are computed by filter-bank decomposition, when efficiently combining frequency and magnitude information from the short-term power spectrum input signals. Time-variant outputs of these filters that might be chosen so as to cover the most relevant part of the frequency range are regarded as the set of time-variant features $\mathbf{x}_n = \{x_n(l) : n \in \mathbb{N}, l \in T\}$. Therefore, sampled vector over discrete time, l, of each narrow-band feature, $\mathbf{x}_n$, is attained by filter bank modeling, for instance, using the set of *Linear Frequency Cepstral Coefficients* (LFCC) that is extracted by discrete cosine transform of triangular log-filter banks, $\{F_m(k) : m = 1, \ldots, n_M\}$, linearly spaced in the frequency domain

$$x_n(l) = \sum_{m=1}^{n_M} \log(s_m(l)) \cos\left(n\left(m - \frac{1}{2}\frac{\pi}{p}\right)\right), \tag{2}$$

where p is the number of desired LFCC trajectories to be considered, and $s_m(l)$ is the weighted sum of each frequency filter response set, $s_m(l) = \sum_{k=1}^{n_K} S_\xi(l, k)F_m(k)$, being m, l, and k the indexes for filter ordinal, time, and frequency axes, respectively; n_K stands for the number of samples in the frequency domain.

Other effective way of generating *t-f* based time-variant features can be achieved through computation of the histograms of the subband spectral centroids that are estimated for each filter in the frequency domain, $F_m'(k)$, by

$$x_n(l) = \frac{\sum_{k=1}^{n_K} kF_n'(k)S_\xi^\gamma(l,k)}{\sum_{k=1}^{n_K} F_n'(k)S_\xi^\gamma(l,k)}, \quad (3)$$

where γ is a parameter that represents the dynamic range of the spectrum, used in the computation of the centroid, and the filters $F_n'(k)$ are linearly distributed along the spectrum. In addition, the energy around each centroid can be also considered as time-variant feature that for a fixed bandwidth Δk is computed by means of

$$x_n(l) = \sum_{k=\hat{x}_n(l)-\Delta k}^{\hat{x}_n(l)+\Delta k} S_\xi(l,k), \quad (4)$$

where $\hat{x}_n(l)$ is the actual value of the time-variant centroid that is estimated by (3).

2.2. Definition of Feature Relevance. Relevance analysis distinguishes variables that are effectively representing the subjacent physiological phenomena according to some evaluation measure, named *relevant features*; it tries to reject variables, whose contribution to representation target is as near zero as possible (*irrelevant features*), as well as those that have repeated information (*redundant features*). Thus, when providing feature selection in the context of any inference, the foremost issue is defining the notion of relevancy [6].

Let $\mathbf{X}_\xi \in \mathbb{R}^{N \times p}$ be an object set comprising N observation samples that are described by a p-dimensional feature set $\xi = \{\xi_i : i = 1,\ldots,p\}$. Besides, each sample is associated to one, and only one, element from the class label set $\mathbf{c} = \{c_k \in \mathbb{N} : k = 1,\ldots,K,\}$, where K is the amount of classes under consideration. Then, given $\mathbf{X}_\xi$, and for each one of the features $\xi_i \in \xi$, the relevance function g is defined as follows:

$$g : \mathbb{R}^{N \times p} \times \xi \longrightarrow \mathbb{R},$$
$$(\mathbf{X}_\xi, \xi_i) \longmapsto g(\mathbf{X}_\xi, \xi_i) \in \mathbb{R}, \quad (5)$$

where over the introduced feature relevance function the following properties are determined.

  (i) *Nonnegativity,* that is, $g(\mathbf{X}_\xi, \xi_i) \geq 0,$ for all i.

  (ii) *Nullity,* the function $g(\mathbf{X}_\xi, \xi_i)$ is null if feature ξ_i has not relevance at all.

  (iii) *Nonredundancy,* if $\xi_i = \alpha\xi_j + \eta$, where the real-valued $\alpha \neq 0$ and η is some noise with mean zero and unit variance, then, $|g(\mathbf{X}_\xi, \xi_i) - g(\mathbf{X}_\xi, \xi_j)| \to 0$.

To ensure the lack of redundancy between any pair of considered features, (ξ_i, ξ_j), then a direct way to remove one of these redundant features is verifying that the zeroth lag of their normalized correlation function does not exceed

a given small positive-valued threshold $\mu \to 0$, and thus the relevance is recomputed as follows:

$$g(\mathbf{X}_{\hat{\xi}}, \xi_i) = \begin{cases} 0, & \left| \mathbf{E}\{\xi_i, \xi_j\} \right| \leq 1 - \mu, : \\ & i, j = 1,\ldots,p; i \neq j, \\ g(\mathbf{X}_\xi, \xi_i), & \text{otherwise,} \end{cases} \quad (6)$$

where $\mathbf{E}\{\cdot\}$ is the expectance operator.

The evaluation of $g(\mathbf{X}_\xi, \xi_i)$ is called *relevance weight*. The main assumption in the proposed approaches is that the largest weight is associated to the most relevant feature. So, we reorder the weights according to their relevance, forming the relevance array

$$\mathbf{g}(\mathbf{X}_{\hat{\xi}}, \xi) = \begin{bmatrix} g_1 \cdots |g_i| \cdots |g_p \end{bmatrix}^\top \in \mathbb{R}^p, \quad \text{with } g_i \geq g_{i+1}. \quad (7)$$

Finally, the proper number of selected parameters having essential information is achieved by truncating the set of stochastic features given in (7). This subset is proposed to be assessed as follows:

$$\mathbf{Z} = \left\{ \xi_i : \sum_{j=1}^q g_j^2 = \sum_{j=1}^p g_j^2 - \epsilon : i = 1,\ldots,q \leq p \right\}, \quad (8)$$

with $\epsilon \in \mathbb{R}$ being a predetermined small threshold positive value.

2.3. Feature Selection by Stochastic-Based Relevance Analysis. Any *stochastic feature* $\xi(t)$ refers to random numeric values that represent measures evolving over time; that is, there is a certain set of parameters, $\Xi = \{\mathbf{x}_i = \xi_i(t) : i = 1,\ldots,p\}$, that changing along the time axis, $t \in T$, is supposed to carry temporal information of the nonstationary signals. Because of high computational cost of stochastic feature-based training, dimension reduction of input spaces is to be carried out, being latent variable techniques widely used for this aim that finds a transformation reducing p-dimensional stochastic feature arrangement, $\Xi \in \mathbb{R}^{p \times T}$, into q-dimensional stochastic set, $\mathbf{Z} \in \mathbb{R}^{q \times T}$, $q \leq p$, in such a way that the data information is maximally preserved. Besides, as the relevance function g, the evaluation measure of transformation is given that distinguishes variables effectively representing the subjacent physiological phenomena, termed *relevant stochastic features*.

Given a set of stochastic features, $\{\mathbf{x}_i\}$, with observation assemble comprising N objects that are disposed in the input observation matrix $\mathbf{X}_\Xi = [X_1| \cdots |X_i| \cdots |X_N|]$, denoting every object X_i, $i = 1,\ldots,N$, that is described by the respective observation set of time-variant arrangements, $\{\mathbf{x}_{ji} \subset \Xi, j = 1,\ldots,p\}$, such that $X_i = [\mathbf{x}_{1i}| \cdots |\mathbf{x}_{ji}| \cdots |\mathbf{x}_{pi}]^T$, $X_i \in \mathbb{R}^{p \times T}$, where $\mathbf{x}_{ji} = [x_{ji}(1) \cdots x_{ji}(t) \cdots x_{ji}(T)]$ is each one of the measured or estimated short-term features from biosignal recordings, equally sampled evolving through the time, and being $x_{ij}(t)$, the jth stochstic feature for the ith object upon a concrete t instant of time.

For the sake of simplicity, the reduction dimension is developed when projecting by the simplest time-evolving latent variable approach. So, given the observation matrix, $\mathbf{X}_\Xi$, there will be a couple of orthonormal matrixes, $\mathbf{U}, \mathbf{V}$, plus diagonal matrix $\mathbf{\Sigma}_\mathbf{X}$, as well, so that the following linear decomposition takes place

$$\mathbf{X}_\Xi = \mathbf{U}\mathbf{\Sigma}_\mathbf{X}\mathbf{V}^\top, \quad \mathbf{U} \in \mathbb{R}^{N \times N}, \ \mathbf{\Sigma}_\mathbf{X} \in \mathbb{R}^{N \times pT}, \ \mathbf{V} \in \mathbb{R}^{pT \times pT}, \tag{9}$$

where $\mathbf{\Sigma}_\mathbf{X} \in \mathbb{R}^{pT \times pT}$ holds first ordered q as most relevant eigenvalues of matrix $\mathbf{X}_\Xi$, $\nu_1 \geqslant \nu_2, \ldots, \geqslant \nu_q \geqslant \nu_{q+1}, \ldots, \geqslant \nu_{pT} \geqslant 0$, that implies the relevance measure to be considered. It must be quoted that inherently to basic latent variable approach, the minimum mean squared-based error is assumed as the evaluation measure of transformation, $g(\mathbf{X}_\Xi, \mathbf{Z}) \sim \min \mathbf{E}\{\|\Xi - \mathbf{Z}\|_2\}$, (where $\| \cdot \|_2$ is the norm squared value); that is, maximum variance is preferred as relevance measure, when the following estimation of covariance matrix is carried out

$$\mathrm{cov}\{\mathbf{X}_\Xi\} = \mathbf{X}_\Xi^\top \mathbf{X}_\Xi = \mathbf{V}\mathbf{\Sigma}_\mathbf{X}^2\mathbf{V}^\top. \tag{10}$$

To make clear the contribution of each time-variant value $x_{ij}(t)$, (10) can be further extended in the form $\mathbf{X}_\Xi = \sum_{j=1}^q \nu_j U_j V_j^T$, where U_j is the jth column of matrix $\mathbf{U}$. So, each vector time row $\tau = 1, \ldots, pT$ of $\mathbf{X}_\Xi$ is written by terms of the sum of the respective eigenvalues, $\{\nu_j\}$; that is, $\chi(\tau) = \sum_{j=1}^q \lambda_{\tau j} V_j$, where $\lambda_{\tau j} = \nu_j U_{\tau j}$, and $U_{\tau j}$ being the τjth element of $\mathbf{U}$. Each time row τ of $\mathbf{X}_\Xi$ is a vector comprising $\lfloor \tau/(T+1) \rfloor + 1$-th stochastic variable, measured upon the whole object set for a given $\tau \bmod (T)$ moment of time, where $\lfloor z \rfloor$ denotes the largest integer less than or equal to z.

Consequently, the amount of variance captured at moment t by the decomposition eigenvalue set is assessed as time-variant relevance measure, which is associated to the variance of respective time row, τ, for each feature along the time as follows:

$$g(\mathbf{X}_\Xi, \mathbf{Z}; \tau) = \mathrm{var}(\chi(\tau)). \tag{11}$$

Nonetheless, based on minimum mean squared-based relevance measure, one might take advantage of the extra information given by the label class set $\mathbf{c} = \{c_k\}$, thus considered relevance function $g(\mathbf{X}_\Xi, \xi_j; \tau, c_k)$ becomes also label dependent. Linear multivariate technique based on PLS regression might be used as a supervised technique that somehow generalizes PCA orthogonal decomposition of (10), but with different way to chose the latent vectors where additional conditions are required, new basis vectors, $\{\gamma_i : i = 1, \ldots, q\}$, are resulting after simultaneous decomposition

$$\mathbf{X}_\Xi = \mathbf{\Gamma}\mathbf{P}^\top + \varepsilon_X, \quad \mathbf{c} = \mathbf{\Gamma}\mathbf{Q}^\top + \varepsilon_c, \tag{12}$$

with $\mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}$, where matrices ε_X and ε_c are the error terms, assumed independent and identically distributed. Matrixes $\mathbf{P}$ and $\mathbf{Q}$ contain the weights revealing the influence of individual $\mathbf{X}_\Xi$ and $\mathbf{c}$-variables, respectively.

3. Experimental Setup

The proposed methodology for automatic identification of pathologies, based on analysis of stochastic features that are extracted from TFR, appraises the next stages: (a) preprocessing, (b) biosignal TFR-based enhancement, (c) feature selection and extraction embracing dimension reduction of time-variant parameters derived from TFR, and (d) OSA detection. Figure 1 shows experimental outline of OSA detection and the methods subject to investigation. Testing of proposed methodology for relevance analysis is carried out on stochastic features that are calculated for Heart-Rate-Variability (HRV) time series, which are computed from electrocardiographic recordings (ECG).

3.1. Database and Preprocessing. This collection of ECG recordings was downloaded from PhysioNet [7]. Database consists of 30 ECGs (modified lead V2); each one includes a set of reference annotations obtained from the study of simultaneously recorded respiration signals, and added every minute of the recording that indicate either the presence or absence of apnoea during each segment of time. Labels are divided into two groups: class apnea, having 100 min with sleep apnea and class normal, with only 5 min or less with sleep apnea (no apnea). These recordings were obtained from men and women between the ages 27 through 63 years old, with weights ranging from 53 to 135 kg. Biosignal recording were digitized at 100 Hz with 16 bit resolution. It must be quoted that training conditions for this database are selected as recommended in [8]. Histograms of ECG recording length are depicted for both considered classes in Figure 2, and it can be seen that their values mostly range within (50–80) heart-per-minute interval, that is, having a no considerable variation from the length mean value.

Basically, automatic OSA diagnosis requires for the extraction of HRV time series from each ECG recording, which in turn can be estimated precisely if an accurate recognition of the *QRS* complex fiducial point is achieved. In this work, complex detection is carried out by method proposed in [9] that includes linear filtering, followed by a non linear transformation, and adaptive decision rules, as well. Further smoothing of anomaly valued peaks of assembled *RR* time series is achieved [8]. Then, the HRV time serie is normalized, termed $x'(l)$, as recommended in [10]

$$x'(l) = \frac{2(x(l) - \mathbf{E}\{\mathbf{x}\})}{\max_{\forall l}\{\mathbf{x}\} - \min_{\forall l}\{\mathbf{x}\}}, \quad l \in T. \tag{13}$$

3.2. TFR Enhancement and Feature Generation. Figure 3 shows examples of estimated *t-f* representations, performed for both typical normal and pathological HRV time series. Parameter tuning for a considered TFR is achieved by the procedure developed for biosignals, discussed in [5]. In the beginning, the STFT-based quadratic spectrogram, which has been assumed as the TFR baseline, is computed by sliding Hamming windows of 32.5 ms length and 50% of overlapping, based on the above-explained spectral HRV properties. Regarding to WT alternative, the respective TFR is performed using Morlet wavelet mother function that is
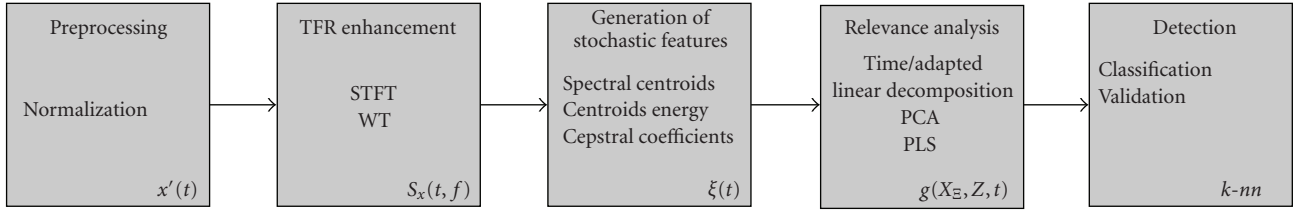
Preprocessing		TFR enhancement		Generation of stochastic features		Relevance analysis		Detection	
Normalization		STFT WT		Spectral centroids Centroids energy Cepstral coefficients		Time/adapted linear decomposition PCA PLS		Classification Validation	
	$x'(t)$		$S_x(t, f)$		$\xi(t)$		$g(X_\Xi, Z, t)$		k-nn

FIGURE 1: Schematic representation of an automated system for pathology diagnosing based on analysis of stochastic features.
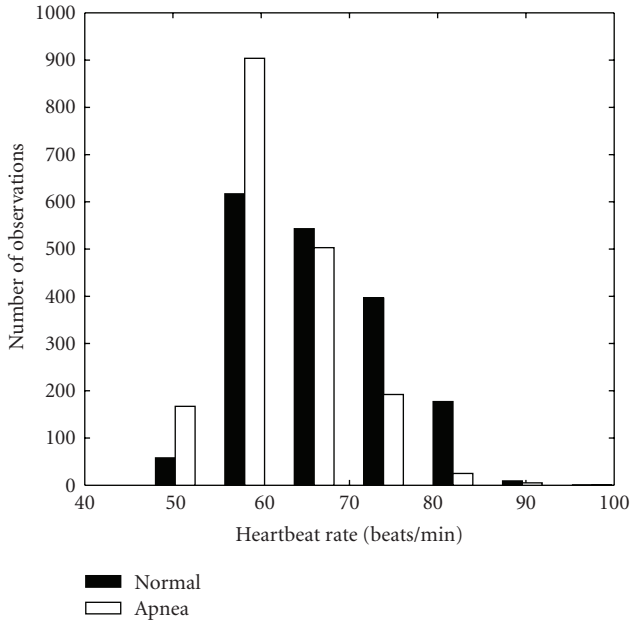

FIGURE 2: Histogram of ECG recording lengths for both considered class labels (normal and pathologic).

likely to be an efficient means for HRV studies [11]. Based on the dominant frequency components of the underlying HRV signal, the number of WT decomposition levels is chosen within the interval (2–512), where the value 2 is regarded to the highest assumed frequency value, while value of 512 reflects the lowest frequency.

As result, the input data space includes the following 52 time-variant features to be studied: the first 20 spectral centroids and their respective energy (see (4)) are estimated by using Hamming filters with 30% overlap and linear response distribution. In case of cepstral coefficients, their first 12 vectors are considered, being computed by 32 triangular response filters with 50% overlap.

3.3. Training Based on Stochastic Relevance Analysis

3.3.1. Estimation of Relevance Weights by Time-Adapted Multivariate Approach.
As said before, each time-dependent feature is assumed to have a relative associated weight of relevance; the largest weight, the most relevant the parameter. However, any estimate of relevance weight is conditioned by the given stochastic feature set taken for its

calculation. For the concrete case of OSA diagnosing, the best set of selected parameters can be achieved using, at least, two different combining approaches of comparison: firstly, when taking a partially divided set that comprises just a single type of performed stochastic features, having the same principle of generation (see §II-A). Secondly, when the best set is chosen among the whole set of features, no matter of their physical meaning. Even that the former approach is more commonly used because of the convenient physical interpretation of selected set, and tuning of training procedures is provided throughout this work based on the latter combining way. Both combining approaches of stochastic feature selection are studied in terms of accuracy performance and reduction dimension, as well.

Figure 4(a) illustrates the attained estimates of normalized relevance weights. Estimation is provided after removing redundant stochastic features in accordance to (6). Weights are ordered by ordinal contour number, which are calculated when taking a partially divided set. Next, Figure 4(b) displays in detail weights, ordered by decreasing relevance, computed when taking into account the whole set of stochastic features. All 45 referred stochastic features in Figure 4 are nonzero weighted, meaning that the other 7 left parameters are totally redundant. So, the number of selected features for case of STFT enhancement is 35 whereas for wavelet-based TFR, that number becomes 30. As seen, estimation of weights evidences no big variations on dependence on the achieved TFR enhancement.

As explained above, further improvement for stochastic features selection might be reached by considering the advantage of the extra information given by the label class set, that is, if using PLS, when relevance weights are calculated according to eigenvalue estimation, given in (12). As a result, the assessed number of selected features selected is around 42 for both considered cases of TFR enhancement, which is higher that the reached amount of features estimated for PCA.

3.3.2. Classification Performance Outcomes.
Tuning of the different schemes of considered parameterizations throughout this training procedure is carried out by using the average classification accuracy for the automatic OSA detection, which is estimated using a simple k-nearest neighbor k-nn classifier, followed by the well-known cross-validation methodology. Several reasons account for the widespread use of this classifier: it is straightforward to implement, it generally leads to a good recognition performance (thanks

(a) Normal—by STFT enhancement

(b) Apnoea—by STFT enhancement

(c) Normal—by WT enhancement
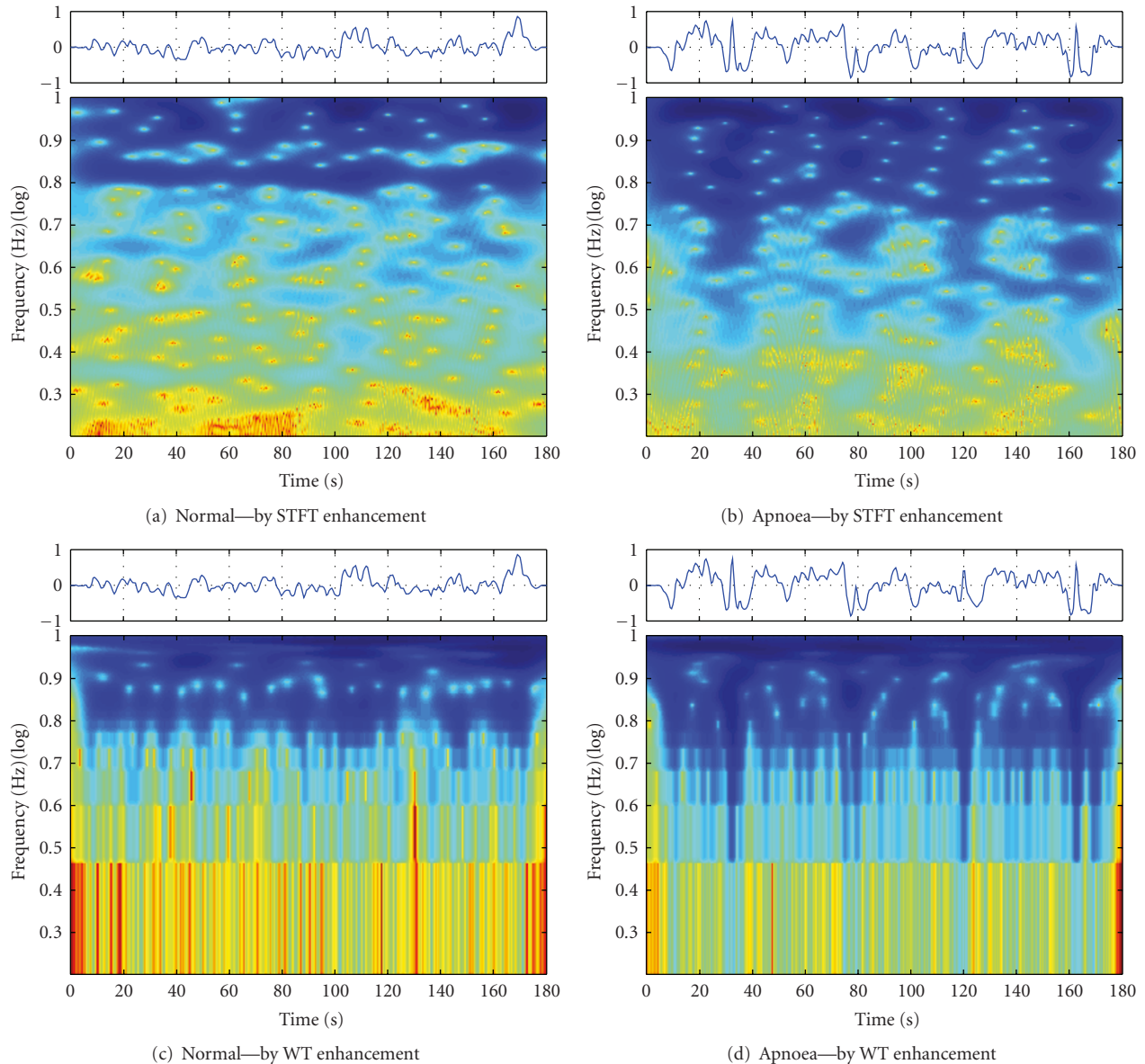
(d) Apnoea—by WT enhancement

FIGURE 3: Examples of TFR on dependence on considered enhancement (STFT and WT) that are estimated for HRV recordings with labels: normal and apnoea.

to the nonlinearity of its decision boundaries), and its complexity is assumed to be independent of the number of classes. So, the tuning of *k-nn* classifier is carried out by calculating the optimal number of neighbors in the sense of accuracy performance. In particular, Figure 5 illustrates how the estimated performance changes when ranging from 1 to 30 the number of considered principal components, which are mutually linear independent combination of original variables. Some fluctuations are also observed when varying the number of neighbors within the interval from 1 to 21. As seen for both compared cases of TFR enhancement, an adequate number on neighbors as much as 7 can be adjusted, which is estimated based on the amount of explained variance, when high-order components are not greatly contributing. But, certain differences are noted between both

used feature extraction approaches. Explicitly, the number of latent components needed in case of STFT-based TFR is just 22 (reaching an accuracy value of some close to 92%), while if using WT the component amount is 8, (for reached accuracy close to 91%). Therefore, even than the STFT-based TFR reaches slightly better accuracy, at the same time WT performs better dimension reduction.

Regarding to comparison among partially divided sets of stochastic features, Figure 6 depicts estimated accuracy values of time-dependent parameters, which are computed when adding one by one their weights ordered by decreasing relevance. From Figures 6(a) and 6(b), attained for both considered methods of TFR enhancement, STFT and WT, respectively, it can be seen that the subset of centroid coefficients achieves the better accuracy performance, regardless
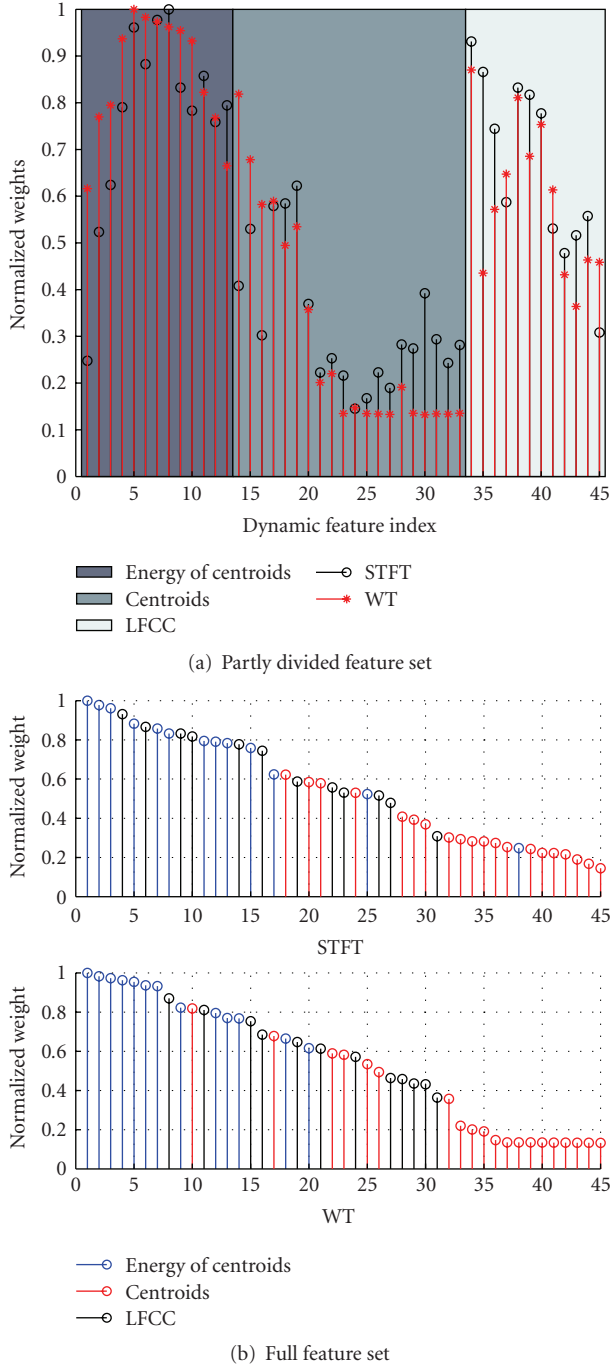
(a) Partly divided feature set

(b) Full feature set

FIGURE 4: On computing relevance weights for considered combining approaches for stochastic feature selection.
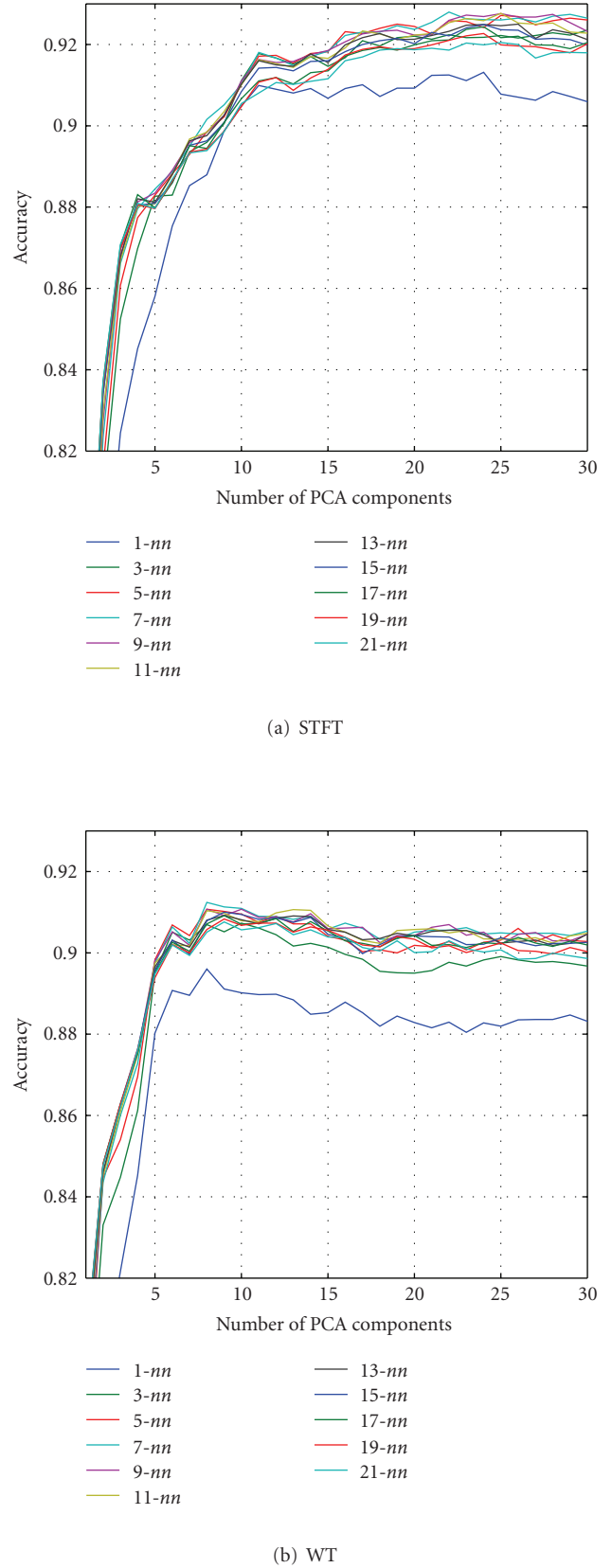

(a) STFT

(b) WT

FIGURE 5: Accuracy performance depending upon the number of *k-nn* neighbors when varying the number of used principal components, for both considered cases of feature extraction enhancement.

of their exhibited lower values of relevance. At the same time, coefficients of centroid energy turn to have the worse accuracy, no matter of their reached before higher relevance weights.

In addition, detailed observation of Figure 6 makes clear that after certain number of added weights the assessed accuracy performance tends to an asymptotical value. For this case of training when taking a partially divided set, the best performance value achieved is 88,7%, if using STFT, and
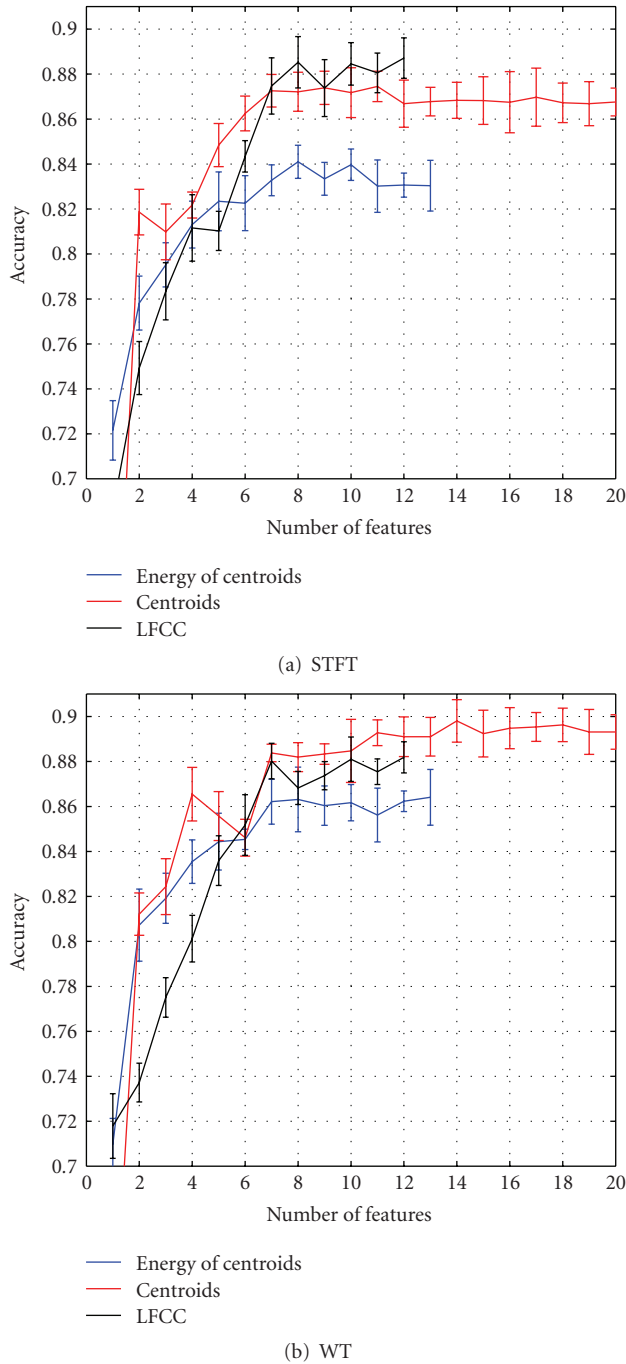
(a) STFT

(b) WT

Figure 6: Accuracy performance for partially divided sets of selected dynamic features.

89.8% if using WT; both values are lower that their respective estimates computed for a whole set of features. Yet, even that the accuracy is lower, the reduction on computational load is higher for last studied approach of selecting weights. In the end, achieved minute-by-minute classification accuracy for either case of HRV-based training are comparable to those outcomes referred in previous works (90% [12], 90.16% [13], and 93% [14]), being attained for similar training conditions.
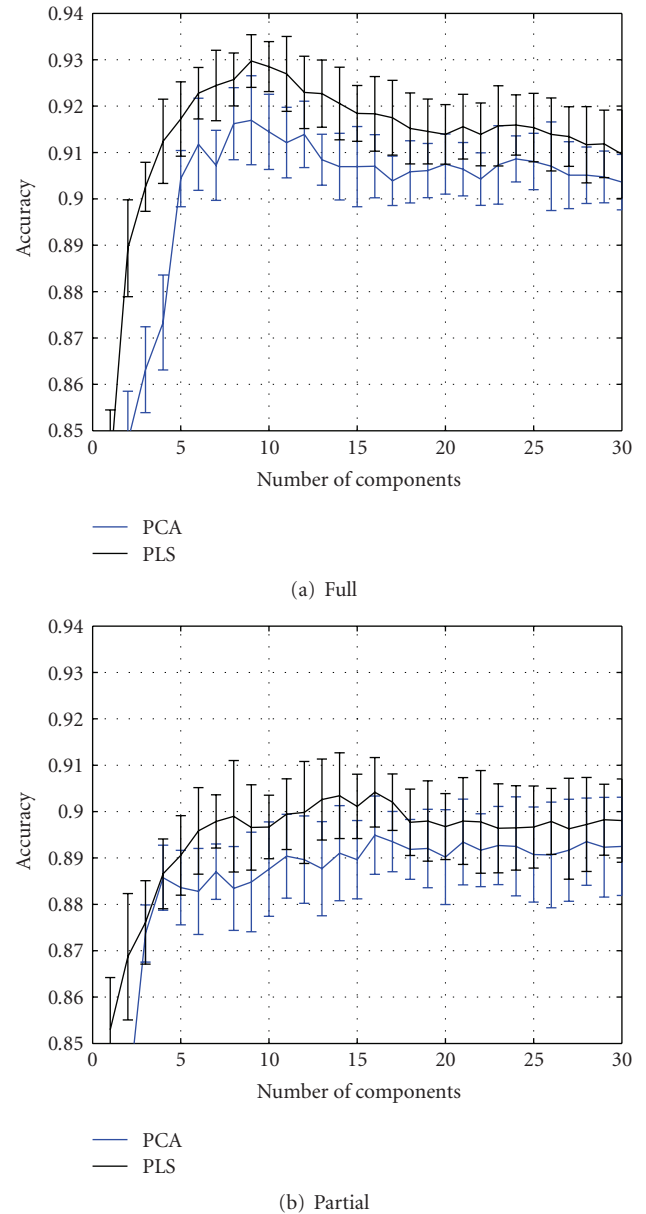

(a) Full

(b) Partial

Figure 7: Improvement of accuracy performed by using PLS technique.

Lastly, Figure 7 that resumes the accuracy performed by PLS shows that the estimation of accuracy tends to be more stable, thus, pointing out on the robustness of the suggested methodology using stochastic features. In fact, the achieved performance when using *k-nn* classifier is summarized in Table 1.

4. Discussion

It should be remarked that the main goal of the present paper is to use a complex of signal processing algorithms for pathology diagnosing, where features are computed from biosignal recordings, and therefore taking advantage from extracting

TABLE 1: Achieved sensitivity and specificity for training for improved methodology for performed accuracy of classification.

Feature set	Sensibility [%]	Specificity [%]	Precision [%]
PCA—total set	89.00	91.00	91.00
PCA—parcial set	88.60	94.20	91.83
PLS—total set	89.00	93.00	93.00
PLS—parcial set	81.60	92.83	92.66

additional relevant information about the ongoing correlated process. Thus, there are investigated clue aspects of proposed methodology that is based on analysis of relevance, given a set of stochastic features that are extracted from t-f representation. Although this approach had been used previously for detection of other pathologies from biosignal recordings [3, 5], the present study is framed on analysis of relevance of stochastic processes derived from filter-bank-based TFR features. Discussed training methodology lies on the hypothesis that each time-dependent parameter holds a relative associated relevance weight, and in this connection, the results also evidence the following aspects to take into consideration.

(i) The enhanced feature estimation that is carried out by introducing t-f representation should be regarded as a remarkable factor for adequate generation of any studied set of stochastic features. Here, feature enhancement is performed by means of nonparametric TFR that had been reported to be appropriate for the analysis of nonstationary biological signals. Nonetheless, results of Figure 5 make evident that even that the STFT-based TFR reaches slightly better accuracy, and still WT performs better dimension reduction. Yet, since the WT decomposition of signals requires an adequate, regular, and localized mother wavelet function [15], whose selection is out of the scope of the present work, an additional improvement of pathology detection performance should be expected in this regard.

(ii) Several studies have established the discriminating capability of frequency bands of biological activity between normal and pathological signals, thus the set of considered TFR-based dynamic measures are to be related to the time-variant features suitable estimated by spectral subband methods. Namely, linear frequency cepstral coefficients, spectral centroids, and their energy contours have been considered. Besides, since more efforts should be done to define the features carrying fundamental information for the classification of pathologies, as quoted in [12], then the set of considered features could fulfil with this requirement because of their easier interpretation.

(iii) With regard to feature selection, proposed methodology for analysis of stochastic relevance is based on time-adapted linear component approach. At this point, two main issues are to be considered: the same measure associated to a given relevance function and the multivariate transformation through the time axis, which is assumed to maximize the measure of relevance present in the contours by their projection onto a new space. As a measure of relevance, the maximum variance is assumed. Specifically, time-adapted versions of both linear transformations, PCA and PLS, are

used throughout this paper as unsupervised and supervised methods, respectively, to perform dimensionality reduction. Results of accuracy, showed in Figure 7, point out on clear advantage of the extra information given by the label class set, that is, relevance weights that are recomputed, according to PLS-based eigenvalue estimation by (12), become more evident, and thus explaining their better performed accuracy.

(iv) Two different combining approaches for selecting the best set of contours are studied: firstly, when taking a partially divided set that comprises a single type of stochastic features, having the same principle of generation. Secondly, when the best contours are chosen among the whole set of features, no matter on their physical meaning. Achieved performance, for both considered methods of feature extraction, leads to the conclusion that latter approach of selecting contours brings higher accuracy though the former approach might be also contemplated because of the convenient physical interpretation of selected feature set. In either case, attained results can be oriented in research focused on finding alternative methods minimizing the parameters used for pathology diagnosing, but strongly requiring biosignal interpretation.

5. Conclusions

The training methodology for detection of pathologies is explored, which is based on relevance analysis of stochastic features extracted from nonparametric t-f representation. Achieved results related to performed accuracy and dimension reduction are comparable with respect to another outcomes, reported in the literature and clearly showing that proposed methodology can be focused on finding alternative methods minimizing the parameters used for pathology diagnosing. Therefore, discussed training approach of feature selection that is capable of capturing the stochastic information, which lies on the hypothesis that each time-dependent characteristic holds a relative associated relevance weight, is proved to be valid for considered cases of pathology diagnosing from biosignal recordings. In this line of analysis, based on simple multivariate techniques, like PCA or PLS, the implementation of time-adapted linear component approach turns to be adequate to maximize the measure of relevance present in the stochastic features. Regarding to the same assumed measure of relevance—the maximum variance—some lacks of its estimation are fixed. Therefore, exploring studies of more robust measures (information entropy, mutual information, correntropy, etc.) are to be further carried out.

Another aspect worthy of mention is the enhanced feature estimation carried out by using the STFT-based quadratic spectrogram, as a baseline TFR, and by wavelet transform. Even that both TFR had been previously reported to be an efficient means for biosignal analysis, no substantial differences on performance are measured. However, additional improvement of pathology detection performance should be expected if properly tuning the WT decomposition.

As future work, further efforts on testing different measures of relevance, should be focused on extended studies to

corroborate the potential of another approaches in conjunction with biosignals analysis and pathology diagnosing.

Acknowledgments

References

[1] M. Sun, M. L. Scheuer, and R. J. Sclabassi, "Decomposition of biomedical signals for enhancement of their time-frequency distributions," *Journal of the Franklin Institute*, vol. 337, no. 4, pp. 453–467, 2000.

[2] Y. Zhao and S. Zhang, "Generalized dimension–reduction framework for recent–biased time series analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 2, pp. 231–244, 2006.

[3] G. Daza-Santacoloma, J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Rutz, and G. Castellanos-Domínguez, "Dynamic feature extraction: an application to voice pathology detection," *Intelligent Automation and Soft Computing*, vol. 15, no. 4, pp. 665–680, 2009.

[4] L. D. Avendaño-Valencia, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Feature extraction from parametric time frequency representations for heart murmur detection," *Annals of Biomedical Engineering*, vol. 38, no. 8, pp. 2716–2732, 2010.

[5] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals," *Annals of Biomedical Engineering*, vol. 38, no. 1, pp. 118–137, 2010.

[6] E. Delgado-Trejos, A. Perera-Lluna, M. Vallverdú-Ferrer, P. Caminal-Magrans, and G. Castellanos-Domínguez, "Dimensionality reduction oriented toward the feature visualization for ischemia detection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 590–598, 2009.

[7] T. Penzel, "Is heart rate variability the simple solution to diagnose sleep apnoea?" *European Respiratory Journal*, vol. 22, no. 6, pp. 870–871, 2003.

[8] M. J. Lado, X. A. Vila, L. Rodríguez-Liñares, A. J. Méndez, D. N. Olivieri, and P. Félix, "Detecting Sleep Apnea by Heart Rate Variability Analysis: Assessing the Validity of Databases and Algorithms," *Journal of Medical Systems*, pp. 1–9, 2009.

[9] L. Sörnmo and P. Laguna, *Biolectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier, Amsterdam, The Netherlands, 2005.

[10] V. Kudriavtsev, V. Polyshchuk, and D. L. Roy, "Heart energy signature spectrogram for cardiovascular diagnosis," *BioMedical Engineering Online*, vol. 6, no. 1, article 16, 2007.

[11] D. A. Newandee and S. S. Reisman, "Wavelet representation comparison for heart rate variability analysis," in *Proceedings of the IEEE 29th Annual Bioengineering Conference*, pp. 112–113, Newark, NJ, USA, 2003.

[12] M. O. Mendez, J. Corthout, S. Van Huffel et al., "Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis," *Physiological Measurement*, vol. 31, no. 3, pp. 273–289, 2010.

[13] M. Al-Abed, M. Manry, J. R. Burk, E. A. Lucas, and K. Behbehani, "A method to detect obstructive sleep apnea using neural network classification of time-frequency plots of the heart rate variability," in *Proceedings of the 29th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '07)*, pp. 6102–6105, 2007.

[14] A. Hossen, B. A. Ghunaimi, and M. O. Hassan, "Subband decomposition soft-decision algorithm for heart rate variability analysis in patients with obstructive sleep apnea and normal controls," *Signal Processing*, vol. 85, no. 1, pp. 95–106, 2005.

[15] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37–48, 2009.