Dynamic Relevance Analysis in Biosignals

Lina María Sepúlveda Cano

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electrical, Electronic and Computer Engineering
Manizales
2013
Dynamic Relevance Analysis in Biosignals

Lina María Sepúlveda Cano

Thesis for the degree of
Dr. Eng. — Automatics

Advisors
Prof. César Germán Castellanos Domínguez
Prof. Carlos Daniel Acosta Medina

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electrical, Electronic and Computer Engineering
Manizales
2013
Análisis Dinámico de Relevancia en Bioseñales

Lina María Sepúlveda Cano

Tesis para optar al título de
Dr. Ing. — Automática

Directores
Prof. César Germán Castellanos Domínguez
Prof. Carlos Daniel Acosta Medina

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales
2013
CONTENTS

7.3 Discussion ................................................. 49

8 Time–Frequency Decomposition Enhancement for Apnea Patient Detection 57

8.1 Preprocessing .............................................. 58
8.2 Continuous Decompositions .................................. 60
  8.2.1 Time–Frequency decomposition enhancement ............. 60
  8.2.2 Results of Time–Frequency Decomposition Enhancement ... 63
8.3 Discussion .................................................. 64

9 Features Selection for Epilepsy Detection 68

9.1 Preprocessing .............................................. 69
9.2 Time–Varying Decompositions ................................ 69
  9.2.1 Stochastic Features Subset Selection ....................... 70
  9.2.2 Relevance Analysis ....................................... 70
  9.2.3 Results of Time–Varying Decompositions .................. 70
9.3 Time–Frequency Decompositions ................................ 71
  9.3.1 Continuous Decomposition Enhancement by Spectral Splitting ... 73
  9.3.2 Results of Time–Frequency Decompositions ............... 73
  9.3.3 EEG–Rhythms Analysis .................................... 74
9.4 Discussion .................................................. 76

10 Respiratory Frequency Extraction in Treadmill Exercise Test 82

10.1 Preprocessing .............................................. 83
10.2 Signal Decomposition ...................................... 83
  10.2.1 Time–Frequency Decomposition ......................... 83
  10.2.2 SSA ....................................................... 84
  10.2.3 Empirical Mode Decomposition (EMD) .................... 89
10.3 Results ..................................................... 89
10.4 Discussion .................................................. 90

IV Conclusions and Future Work 92

11 Conclusions .................................................. 93

V Appendix .................................................... 95

A List of Publications .......................................... 96
List of Figures

6.1 Proposed methodology - General scheme . . . . . . . . . . . . 34
6.2 Static relevance analysis for DLM decomposition technique . . 36
6.3 Selection of frequency bands using relevance analysis . . . . . 38

7.1 Example of HRV time–frequency and time–scale representations. 45
7.2 Static relevance weights based on non–supervised measurement 47
7.3 On computing dynamic relevance weights based on non–supervised measurement 48
7.4 Selection of frequency bands using relevance analysis for OSA diagnosis. 53
7.5 Scalogram enhancement for OSA diagnosis . . . . . . . . . . . 54

8.1 Histogram of heart beat rate per minute for a given set of labeled PPG fragments 60
8.2 Example of time–frequency representations . . . . . . . . . . . 61
8.3 Tuning of the number of TFR–based dynamic features . . . . . 62
8.4 On computing relevance weights based on static relevance . . . 62
8.5 On computing dynamic relevance weights based on non–supervised measurement 63
8.6 Performed ROC curves for both considered fragment lengths . 64

9.1 Static relevance analysis for stochastic features selection . . . 71
9.2 Dynamic relevance analysis for stochastic features selection . . 71
9.3 Example of rhythms spectrum for different data-sets . . . . . . 75

10.1 Examples of maximum frequency at the time–frequency representation - HRV 84
10.2 Examples of original and reconstructed signals - HRV . . . . . 85
10.3 Examples of correlation matrix of the eigen triples . . . . . . 86
10.4 Examples of reconstructed and filtered signals - HRV . . . . . 87
10.5 Examples of eigen–triples instantaneous frequency . . . . . . 87
10.6 Comparison between the processed and the original signals - HRV 88
10.7 Results for EMD and instantaneous frequency . . . . . . . . . 89
Acknowledgments

This work is supported under project “Servicio de monitoreo remoto de actividad cardiaca para el tamizaje clinico en la red de telemedicina del departamento de caldas”, financed by UNAL-UCALDAS (Colombia) and 'Becas para estudiantes sobresalientes de posgrado de la Universidad Nacional de Colombia'. 
Abstract

In this work, a methodology for biosignal analysis (e.g. pathology diagnosis) is discussed, which is based on dynamic relevance analysis of stochastic features extracted from different decomposition techniques 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 for a given relevance function. Specifically, since the maximum variance is assumed as a measure of relevance, time-adapted supervised approaches are developed. Additionally, in the case of high dimensionality data with significant correlation among the whole set, a dimensionality reduction technique is proposed, based on time-frequency relevance maps. The proposed approaches are experimentally assessed on real-world data sets, allowing to confirm whether the proposed feature selection algorithm is adequate for classification purposes.

The conjunction of these advances conforms a methodology for training pattern recognition systems, which is a fully automatized dimensionality reduction method that allows the use of functional representations. The main advantage of the proposed methodology, is that preserves the maximum information among the high dimensional input data. In this terms of classification performance, the proposed methodology is efficient and competitive, outperforming other similar methods.
Part I

Preliminary
Chapter 1

Motivation

Pattern recognition is the study of how an automated system can observe the environment, learn to distinguish patterns of interest, and take reasonable decisions based on these patterns. It is usual in pattern recognition tasks, the representation or characterization of the objects is performed by means of static features, i.e. numerical values representing some attributes of the signal assumed to be constant through some dimensions, such time or space. However, there is another type of characteristics known as dynamic features, and are numerical values representing some attribute of the signal, which change with respect to any associated dimension. the dynamic features are commonly referenced in the literature as stochastic variables, time series, curves or functional data [1].

Particularly, in the biosignal analysis, stochastic modeling has been under continuous formulation for extracting valuable information due to the high non–stationarity inherent of this kind of time series. In this signals, is assumed that changes in the dynamics of a biological system are usually related to status changes observed [2]. The representation in terms of the signal stochastic behavior of the underlying biological information, could identify changes episodes and pathologies on the hidden dynamics of the variables, when other types of techniques, conventionally employed, fail [2].

For all these reasons, in the present work, a methodology for pathology diagnosis is discussed, which is based on relevance analysis of stochastic feature selection, extracted from different decomposition techniques 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 for a given relevance function. Specifically,
since the maximum variance is assumed as a measure of relevance, time–adapted supervised approaches are developed. Additionally, in the case of high–dimensional data with significant correlation among the whole set, a reduction dimension technique is proposed, based on time–frequency relevance maps. The proposed approaches are experimentally assessed on real-world data sets, allowing to confirm whether the proposed feature selection algorithm is adequate for classification purposes; those data sets are: i) synthetic control chart time series, ii) Apnea data sets (ECG or EKG and PPG), iii) Epilepsy data set (EEG), and iv) Treadmill exercise data set (HRV). In the cases when the classification accuracy is discussed, the 10–folds technique is used, with a training subset of 70% and a testing subset of 30%.
Chapter 2

Objectives

2.1 General Objective

To develop a methodology for dynamic features selection based on a time-varying relevance function, aimed to improve the classification accuracy and the pattern recognition on biosignals.

2.2 Specific Objectives

– To suggest a time-varying relevance function, its properties and constraints, for dynamic features analysis in biological signals.

– To suggest a dynamic feature selection criterion, based on the dynamic relevance function proposed.

– To develop a dynamic feature selection algorithm, based on the relevance temporal analysis, aiming to improve the classification accuracy and the pattern recognition on biosignals.

– To analyze the behavior of the proposed algorithm when the inputs are strong-varying biosignals, specially for treadmill exercise signals.
Part II

Theoretical Framework
A time series $\mathbf{x} = \{x_t : t = 1, \ldots, T\}$, with $\mathbf{x} \in \mathbb{R}^T$, is a collection of observations made chronologically, and its nature includes: large in data size, high dimensionality and update continuously [3]. When there is one time series feature at each time instance, it is called an univariate time series (UTS), and when there are more than one time series feature, it is called a multivariate time series (MTS) [4].

Time series data mining techniques analyze time series data in search of interesting patterns that were previously unknown to information users [5]. Most classic data mining algorithms do not perform or scale well on time series data, due to its intrinsic structural characteristics such as the high dimensionality and feature correlation, combined with the measurement-induced noises that beset real-world time series data, pose challenges that render classic data mining algorithms ineffective and inefficient for time series [6]. The following typical tasks for time series data mining are discussed in [7]:

- Preprocessing: digital filtering, feature selection, feature extraction.
- Similarity: numeric and symbolic time series distances.
- Visualization.
CHAPTER 3. MULTIVARIATE TIME SERIES DATA MINING

- Prediction.
- Clustering.
- Segmentation.
- Classification.

In this thesis, the analyzed data mining tasks for multivariate time series (specifically, biological signals) are the time series representation and preprocessing.

3.1 Non–Parametric Time–Varying Decompositions

Usually, the exact values of each time series data point is not interesting in data mining; rather, the focus is centered in the trends, shapes and patterns contained within the data [8]. In the analysis and processing of biosignals, stochastic modeling has been under continuous formulation for extracting valuable information due to the high non–stationarity inherent to this kind of time series, called stochastic features.

Any stochastic feature $\xi \in \mathbb{R}^T$ refers to random numeric values that represent measurements evolving along the time $t \in T$, i.e., there is a certain set of parameters, $\Xi = \{\xi_i : i = 1, \ldots, p\}$, that change along the time axis, and it is supposed to carry temporal information of the non–stationary signals. In this chapter, different approaches are presented as candidates for time series decomposition in the stochastic feature sense.

Based on generalized spectral representation, time series approximations can be carried out by decomposing the underlying biosignal into a linear combination of harmonic, exponential, or even polynomial series [9], that is, the problem is addressed as modeling multiple time-evolving data streams governed by some linear recurrent formula. As a result, different time series representations have to be prove, aiming to find the most suitable stochastic features set.

3.1.1 Empirical Mode Decomposition

A first approximation to time series representation is the EMD procedure, that decomposes the signal $\mathbf{x}$ into a finite set of frequency and amplitude
modulated signals, being the representation basis called Intrinsic Mode Functions IMF. Each IMF is computed by an iterative process \cite{10}: i) The extrema of the signal is detected (maximum and minimum), ii) The upper and lower envelopes are computed, connecting the maximum and minimum independently by cubic spline interpolation, iii) The local mean of the envelopes is calculated, iv) If the local mean is zero, it is subtracted of the original signal; the result is the first IMF, v) If the following conditions are not met, a sifting process is carried out. If the conditions are met, this IMF is considered, then, the remainder is the new signal and the process begins again.

**Conditions for IMF** Aiming to compute the IMF, two conditions should be considered \cite{11}: The first condition is that in the whole data the IMF must have an equal number of extrema and zero crossings or they can differ by at most one. The second condition is that the mean value calculated from the envelopes must be have to be zero. According to \cite{10}, the Eq. (3.1) for signal reconstruction can be expressed as:

\[
x = \sum_{i=1}^{K} s_i + r_K,
\]

(3.1)

where \( \{s_i\} \) is the set of IMF, \( r_K \) is the remainder and \( K \) is the number of the IMF extracted of the original data. The first IMF is related to the highest frequency and the last is related to the lowest frequency.

A signal can be decomposed into the linear sum of a series of component signals by Hankel matrix-based SVD, and essentially what the component signals reflect are projections of original signal on the orthonormal bases \cite{12}. The following techniques are hankel–based decompositions with different weights for each orthonormal base.

### 3.1.2 Finite Rank Series

Assume a real-valued time series \( x = \{x_t : t = 1, \ldots, T\} \), with \( x \in \mathbb{R}^T \), which is mapped into the multidimensional time series set \( X = \{x^\nabla_k : k = 1, \ldots, K\} \), \( X \in \mathbb{R}^{K \times L} \), termed Hankel Matrix, where \( \nabla \) is the truncation operator, comprising \( K \)-lagged vectors \( x^\nabla_k = [x_{k-1}, \ldots, x_{k+L-2}]^T \); \( x^\nabla_k \in \mathbb{R}^L \), and \( x^\nabla_k \subset x \) where \( K = T - L \), being \( 3 \leq L \leq T \) \cite{13}. Furthermore, if the dimension dim \( \mathfrak{L}^{(L)}\{x\} = \text{span}\{x^\nabla_k : \forall k\} = d \), for \( 0 \leq d \leq L \), then, the
series is regarded as $L$-rank $d$, noted as $\text{rank}_L \{ \mathbf{x} \}$. In turn, when the equality $\text{rank}_L \{ \mathbf{x} \} = d$ holds for all suitable $L$, the series $\mathbf{x}$ has rank $d$, that is, $\text{rank} \{ \mathbf{x} \} = d$. If such $d$ exists, the time series is called a finite rank series \[9\].

Generally, a non–zero series $\mathbf{x}$ is governed by the following linear sequence of dimension $d \geq 1$ if

$$x_{t+d} = \sum_{i=1}^{d} w_i x_{t+d-i},$$

(3.2)

for a certain set $\{w_i \in \mathbb{R} : i = 1, \ldots, d\}$ with some $w_i \neq 0$, and $0 \leq t \leq T - d$.

Since $\text{rank}_L \{ \mathbf{x} \} = \text{rank}_L \{ \mathbf{X} \} = \text{rank}_L \{ \mathbf{XX}^\top \}$, the orthonormal system of eigenvectors $\{ \mathbf{u}_i : i = 1, \ldots, d \}$ (corresponding to $d$ positive eigenvalues of the matrix $\text{rank}_L \{ \mathbf{XX}^\top \}$) constitutes the left singular vectors of the singular value decomposition of the matrix $\mathbf{X}$. This concrete model of Finite Rank Time Series is known as Singular Spectral Analysis (SSA) \[14\] and the eigen–decomposition of the trajectory or Hankel matrix is called eigen–triple.

Suppose a sufficiently large $T$ of the series $\mathbf{x}$, under the assumption that $d < \min(L, K)$, so, taking into account Eq. (3.1), the $K$-lagged vectors $\{ \mathbf{x}_k^\nabla \}$ satisfy the following vector recurrent equation:

$$\mathbf{x}_{k+d}^\nabla = \sum_{i=1}^{d} w_i \mathbf{x}_{k+d-i}^\nabla,$$

(3.3)

The coefficient set in Eq. (3.1) can be found as follows:

$$w = \frac{1}{1 - \mu^2} \sum_{i=1}^{d} u_{di} \mathbf{u}_d^\nabla,$$

(3.4)

where $w = [w_d, \ldots, w_1]^\top$ is the coefficient vector, $\mathbf{u}_d^\nabla = [u_{1i}, \ldots, u_{di}]$ is the $i$–th eigenvector, and $\mu^2 = \sum_{i=1}^{d} u_{di}^2$ is the verticality coefficient.

### 3.1.3 Exponentially Damped Sinusoidal

Moreover, the model (3.1) can be represented as a linear combination of Exponentially Damped Sinusoids (EDS), when the signal is modeled by a finite sum of discrete-time exponentially damped complex sinusoids, as follows \[15\]:

$$x_t = \sum_{i=1}^{d} a_i \exp(j\phi_i) \exp((-h_i + j2\pi f_i)T\Delta t)$$

(3.5)
The amplitudes \( a_i \), damping factors \( h_i \), phases \( \phi_i \), frequencies \( f_i \), and the model order \( d \) are free parameters of the non-linear model. Through Prony analysis presented in [16], the non-linear problem stated in Eq. (3.4) can be decoupled by solving, in a least square sense, a set of linear equations. Thus, model (3.4) is rewritten by using Eq. (3.1), as follows:

\[
x_t = - \sum_{i=1}^{d} w_i x_{t-i},
\]

(3.6)

where the parameter vector \( w = [w_1, \ldots, w_d] \) is given as [17]:

\[
w = \sum_{i=1}^{d} \sigma_i u_i v_i^\top,
\]

(3.7)

being \( \sigma = [\sigma_1, \ldots \sigma_d] \), \( \{v_i : i = 1, \ldots, d\} \) and \( \{u_i : i = 1, \ldots, d\} \) the singular values, and the right and left singular vectors of \( X \).

### 3.2 Parametric Time–Varying Decomposition

Model in Eq. (3.1) can be rewritten in terms of an autoregressive model, as follows:

\[
x_t = wx_{t-1},
\]

(3.8)

where \( x_{t-1} = [x_{t-1} \ldots x_{t-d}]^\top \in \mathbb{R}^d \) is the autoregressive vector and \( w \in \mathbb{R}^{d \times 1} \) is the corresponding parameter vector, given by: \( w = [w_1, \ldots, w_d] \); that is assumed to be constant in time, such a model (3.1) can be also extended to represent a time–varying autoregressive model of order \( d \), termed TVAR:

\[
x_t = \sum_{i=1}^{d} w_{i,t} x_{t-i},
\]

(3.9)

where \( w_t = [w_{1,t}, \ldots, w_{d,t}] \) is the autoregressive parameter at \( t \)-th time sample.

An alternative structure of (3.8) is the Dynamic Linear Model (DLM), given by (3.9) [18]:

\[
x_t = 1^\top H_t W_t x_{t-1} + \delta_t,
\]

(3.10)
where \( \mathbf{1} = [1, \ldots , 1] \in \mathbb{R}^{d \times 1} \), \( \mathbf{H}_t = \text{diag} (\mathbf{U}_t^\top \mathbf{f}) \mathbf{U}_t^{-1} \), \( \mathbf{H}_t \in \mathbb{R}^{d \times d} \), \( \mathbf{f} = [1, 0 \ldots 0] \top \in \mathbb{R}^{d \times 1} \) is the vector of regressors, \( \mathbf{\delta} = [\delta_1, \ldots , \delta_t] \top \) is a zero–mean normal distributed innovation vector, and \( \mathbf{W}_t \in \mathbb{R}^{d \times d} \) is the parameter matrix expressed as:

\[
\mathbf{W}_t = \mathbf{U}_t \mathbf{A}_t \mathbf{U}_t^{-1} = \begin{bmatrix}
   w_{1,t} & w_{2,t} & \cdots & w_{d-1,t} & w_{dt,t} \\
   1 & 0 & \cdots & 0 & 0 \\
   \vdots & \vdots & \ddots & \vdots & \vdots \\
   0 & 0 & \cdots & 1 & 0 
\end{bmatrix}, \quad t = 1 + d, \ldots , T - d
\] (3.11)

being \( \mathbf{U}_t \in \mathbb{R}^{d \times d} \) the eigenvector matrix and \( \mathbf{A}_t \in \mathbb{R}^{d \times d} \) a diagonal matrix containing in the main diagonal the eigenvalues of \( \mathbf{W}_t \), at time instant \( t \).

In this sense, it is important to be able to embed a TVAR model in the DLM form in order to distinguish, and ultimately infer, the process structure from contaminating noise [19]. Particularly, a DLM describing a TVAR process, at moment time \( t \), has the form of Eq. (3.7).

Therefore, the TVAR model in (3.8) can be rewritten by means of vector recurrent expression (3.1):

\[
x_t = \sum_{i=1}^{d} h_{i,t} a_{i,t} u_{i,t} x_{t-i} + \mathbf{\delta}_t,
\] (3.12)

where \( \mathbf{h}_t = [h_{1,t}, \ldots , h_{d,t}] \) is \( \text{diag} (\mathbf{U}_t^\top \mathbf{f}) \) and \( \mathbf{a}_t = [a_{1,t}, \ldots , a_{d,t}] \) is the eigenvalue vector of \( \mathbf{W}_t \).

### 3.3 Non–Parametric Time–Frequency Decompositions

The commonly used time–frequency representation, termed TFR, calculated by the Short Time Fourier Transform, introduces a time localization concept by using a tapering window function \( w(t - \tau) \) (being \( \tau \) the shifting factor) of short duration going along the studied signal, \( \mathbf{x} \in \mathbb{R}^T \), with the window length remaining constant, namely:

\[
\mathbf{x}.\mathbf{w} = \frac{1}{2\pi} \int S(\omega)\phi(\omega)d\omega,
\] (3.13)
where $\phi(\omega) = e^{j\omega t}$ is the basis and $S = \{S(\omega, t)\} \in \mathbb{R}^{Q \times T}$ is the decomposition coefficients matrix.

On the other hand, the signal decomposition through continuous wavelet, termed CWT, can be expressed in two dimensions, as follows [20]:

$$x = \int \int S(a, b) \phi(a, b) \frac{dadb}{a^2},$$  \hspace{1cm} (3.14)

where $\{S(a, b)\}$ are the coefficients of the continuous wavelet, $a$ is scale factor, $b$ is the position factor, and $\phi(a, b) \in \mathbb{R}^T$ is the mother wavelet. The CWT becomes a convenient time–scale representation when the mother wavelet is chosen by taken into account its similitude with the analyzed time series.

The coefficient matrix $S$ of enhanced representations describing the time series dynamics are the spectrogram, in case of TFR, and scalogram for CWT; both of them, respectively, defined as:

$$S_{STFT} = \left| \int_T x(\tau)w(\tau-t)e^{-j\omega t}d\tau \right|^2$$  \hspace{1cm} (3.15)

$$S_{CWT} = \int_T \frac{1}{\sqrt{a}}x(t)\phi^*(t, a, b)dt,$$  \hspace{1cm} (3.16)

where $\phi^*(t, a, b)$ is the conjugated mother wavelet.

Aiming to synthesize the information contained in the decomposition described above, some TFR/CWT–based short–time parameters have been widely accepted for characterizing the biosignals. These are computed by filter–banked decomposition, because the efficiently combine frequency/scale (termed $\lambda$) 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 $\lambda$ range, are regarded as the set of time–variant features $\Xi = \{\xi_i : i \in \mathbb{N}\}$. Therefore, each narrow–band feature, $\xi_i$, is attained by filter banked modeling, for instance, using the set of Linear Frequency Cepstral Coefficients (LFCC) being extracted by Discrete Cosine Transform of triangular log–filter banks, $\{F_m(\lambda) : m = 1, \ldots, n_M\}$, linearly spaced in $\lambda$ domain:

$$\xi_i = \sum_{m=1}^{n_M} \log(s_m) \cos \left(i \left(m - \frac{\pi}{2p}\right) \right) \hspace{1cm} (3.17)$$

where $p$ is the number of desired LFCC features to be considered, and $s_m \in \mathbb{R}^T$ is the weighted sum of each frequency filter response set, $s_m =$
\[ \sum_{\lambda=1}^{\Lambda} S(\lambda) F_m(\lambda) \]

being \( m \) and \( \lambda \) the indexes for filter ordinal frequency or scale axes, respectively; \( \Lambda \) stands for the number of bins in the frequency/scale domain.

Other effective way of generating time–\( \lambda \) based time–variant features can be achieved through histograms computation of the sub-band spectral centroids, that are estimated for each filter \( F_m' (\lambda) \), in the \( \lambda \) domain by:

\[
\xi_{i} = \frac{\sum_{\lambda=1}^{\Lambda} \lambda F_i' (\lambda) S^\gamma(t, \lambda)}{\sum_{\lambda=1}^{\Lambda} F_i' (\lambda) S^\gamma(t, \lambda)} \tag{3.18}
\]

where \( \gamma \) is a parameter representing the dynamic range of the spectrum, used in the computation of the centroid, and the filters \( F_i' (\lambda) \) 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 \( \Delta \lambda \) it is computed by means of:

\[
\xi_{i} = \sum_{\lambda=\hat{\xi}_i(t) - \Delta \lambda}^{\hat{\xi}_i(t) + \Delta \lambda} S(t, \lambda) \tag{3.19}
\]

where \( \hat{\xi}_i \) is the actual value of the time–variant centroid that is estimated by Eq. (3.17).

### 3.4 Summary

A single input signal can be decomposed into an unknown number of recurrent sequences, which may vary from observation to observation. A strong constraint on this decomposition approach is related to the modeling time series holding themselves different stochastic structures. Each time–variant component of the signal could be used as a dynamic feature. Table 3.1 summarizes the time–series decomposition techniques analyzed in this work.
## Table 3.1: Decomposition Techniques

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Equation</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non–Parametric Time–Varying Decompositions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMD</td>
<td>$x = \sum_{i=1}^{K} s_i + r_K$</td>
<td>Number of IMFs</td>
<td>IMFs</td>
</tr>
<tr>
<td>SSA</td>
<td>$x_{t+d} = \sum_{i=1}^{d} w_i x_{t+i-d-i}$</td>
<td>rank_L(x) = d</td>
<td>Based on Hankel</td>
</tr>
<tr>
<td></td>
<td>$w = \frac{1}{1-p^2} \sum_{i=1}^{d} u_i u_i^\top$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDS</td>
<td>$x_t = - \sum_{i=1}^{d} w_i x_{t-i}$</td>
<td>rank_L(x) = d</td>
<td>Based on Hankel</td>
</tr>
<tr>
<td></td>
<td>$w = \sum_{i=1}^{d} \sigma_i u_i v_i^\top$</td>
<td></td>
<td>Non-linear</td>
</tr>
<tr>
<td><strong>Parametric Time-Varying Decompositions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLM</td>
<td>$x_t = w x_{t-1}$</td>
<td>Model Order</td>
<td>AR Model</td>
</tr>
<tr>
<td></td>
<td>$W_t = U_t A_t U_t^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non–Parametric Time–Frequency Decompositions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STFT</td>
<td>$S = \left</td>
<td>\int_{t} x(\tau) w(\tau - t) e^{-j\omega t} , d\tau \right</td>
<td>^2$</td>
</tr>
<tr>
<td>Wavelet</td>
<td>$S = \int_{t} \frac{1}{\sqrt{n}} x(t) \phi^*(t, a, b) , dt$</td>
<td>Wavelet, Scale</td>
<td>Time–Scale</td>
</tr>
<tr>
<td>FCC</td>
<td>$\xi_i = \sum_{m=1}^{n_M} \log(s_m) \cos \left( i \left( m - \frac{1}{2} \right) \right)$</td>
<td>-Coefficients Grid</td>
<td>Stochastic Features</td>
</tr>
</tbody>
</table>
Chapter 4

Stochastic Feature Subset Selection

Feature subset selection (FSS) is one of the techniques to pre-process the data before to perform data mining tasks as classification and clustering [21]. FSS is to identify a subset of original features from a given data set while removing irrelevant and/or redundant features [4]. An MTS item or observation is naturally represented as an $\mathcal{E} \in \mathbb{R}^{p \times T}$ matrix, where $p$ is the number of variables or features, and $T$ is the number of time instants [22]. However, the state of the art feature subset selection techniques, such as Recursive Feature elimination (RFE), require each item to be represented in one row [4]. In the case of stochastic characterization, and given a set of stochastic features, $\{\xi_i \in \mathbb{R}^T\}$ with observation assemble comprising $N$ objects, could be disposed in the input observation matrix $X_{\mathcal{E}} = [X_1|\cdots|X_i|\cdots|X_N]$. Every object $X_i$, $i = 1, \ldots, N$, is described by the respective observation set of time–variant arrangements, $\{\xi_{ji} \subset \mathcal{E}, j = 1, \ldots, p\}$, such that, $X_i = [\xi_{i1}|\cdots|\xi_{ji}|\cdots|\xi_{ip}]^T$, $X_i \in \mathbb{R}^{p \times T}$, where $\xi_{ji} = [\xi_{ji}(1) \ldots \xi_{ji}(t) \ldots \xi_{ji}(T)]$ is each one of the measured or estimated short–term features from time–series recordings. Each element $\xi_{ij}(t)$, is the $j$–th stochastic feature for the $i$–th object upon a concrete $t$ instant of time. As can be seen, the MTS vectorized form contains a large set of data to be analyzed. Feature subset selection techniques and similarity measures from machine learning area are suitable for static features while MTS data is characterized by dynamic features; then, feature selection and classification techniques for dynamic features are necessary for processing MTS data in variety of problems [23].
4.1 Feature Subset Selection Techniques

A summary of the current techniques for feature subset selection and classification for MTS data sets is found in [23]. The following techniques are tested in this thesis aiming to compare the results of the proposed methodology.

4.1.1 Recursive Feature Elimination (RFE)

The goal of the RFE is to find the features subset that maximizes the performance of the predictor, using a Support Vector Machine (SVM) classifier [24]. The importance of a dimension is determined by the influence it has on the margin of a trained SVM [25]. The iterative algorithm proposed by [26] is shown in Algorithm 1. The RFE is considered as a wrapper technique, i.e. a classifier–dependent technique.

4.1.2 Minimum Redundancy and Maximum Relevance (mRMR)

Max-Relevance is to search features satisfying Eq. (4.1), which approximates the maximum dependency on the target class, with the mean value of all mutual information values between individual feature $\xi_i(t), i = 1, \ldots, p_T$ and class $c_k, k = 1, \ldots, K$ [27].

$$\max D\{\Xi, c\}, D = \frac{1}{|\Xi|} \sum_{\xi_i(t) \in \Xi} I(\xi_i(t); c),$$  \hspace{1cm} (4.1)

where $I\{\ldots\}$ the mutual information operator and $c$ is the labels vector.

As an additional criteria for features removal, the authors in [27] proposed the minimum redundancy (Eq. (4.2)), aiming to select mutually exclusive features.

$$\min R\{\Xi\}, R = \frac{1}{|\Xi|^2} \sum_{\xi_i(t), \xi_j(t) \in \Xi} I(\xi_i(t); \xi_j(t)),$$  \hspace{1cm} (4.2)

The criterion combining the above two constraints is called minimal-redundancy-maximal-relevance and is defined by Eq. (4.3).

$$\max \Phi\{D, R\}, \Phi = D - R,$$  \hspace{1cm} (4.3)
**CHAPTER 4. STOCHASTIC FEATURE SUBSET SELECTION**

**Input:** Training examples $X_Ξ \in \mathbb{R}^{N \times pT}$

**Output:** Feature ranked list $Ψ \in \mathbb{R}^r$, $r \leq pT$

Initialize

Subset of surviving features $Ξ = Ξ$

Feature ranked list $Ψ = [\ldots]$

while $Ξ \neq [\ldots]$ do

1. Restrict training examples to good feature index, $X_Ξ$

2. Train the classifier $α = SVM - train\{X_Ξ, c\}$, where $c = \{c_k \in \mathbb{N} : k = 1, \ldots, K\}$ is the class label set and $K$ is the number of classes under consideration.

3. Compute the weight vector of dimension $\mathbb{R}^{pT-r}$, $w = \sum_j α_j c_j X_{Ξj}$.

4. Compute the ranking criteria $\rho_i = (w_i)^2$, $\forall i$.

5. Find the feature with smallest ranking criterion $f = \arg\min\{\rho\}$.

6. Update feature ranked list $Ψ$.

7. Eliminate the feature with smallest ranking criterion.

end

**Algorithm 1:** Recursive Feature Elimination

The mRMR is considered as a *filter* technique, i.e. the selection is independent of the classifier accuracy.

### 4.1.3 Common Principal Component Loading-based Variable (CLeVer)

The CLeVer algorithm proposed by [28], involves three phases: *i*) principal components computation per MTS item (observation), *ii*) descriptive common principal components computation across all principal components (DCPC), and *iii*) variable subset selection using DCPC loadings of variables. The full algorithm described in [29] is listed in Algorithm 2.
Principal component analysis is a generalization of PCA for multivariate data items where all the data items have the same number of dimensions [23], that is a constrain for CLeVer implementation. The CLeVer is considered as an embedded-filter technique.

Input: Training examples $\mathbf{X} \in \mathbb{R}^{N \times pT}$
Output: Selected Variables $\Psi \in \mathbb{R}^r, r \leq pT$

foreach MTS item $n$ do

1. Find the correlation matrix, $\text{corr}\{X_n\}$, of each set of time series data.

2. Calculate singular value decomposition (SVD) of each correlation matrix, $\text{corr}\{X_n\} = \mathbf{U}\Sigma\mathbf{V}^\top$.

3. The variance will be contained in the diagonal entries of $\Sigma$.

4. Find $q$, the number of diagonal entries of $\Sigma$ where the variance is less than a prescribed percent of the total variance for each matrix.

end

5. Use the maximum $q$ for all of the time series matrices and use that number of rows as a cut off point of the loadings. Create the matrix $\mathbf{L}$ defined as $\mathbf{L} = \sum_n \mathbf{M}_n^\top \mathbf{M}_n$, where $\mathbf{M}_n$ is the matrix of the first $q$ principal components in the $n$–th time series matrix.

6. Compute SVD of $\mathbf{L}$.

7. Use the first $q$ rows of $\mathbf{V}^\top$. These are the DCPCs.

8. Do k-means clustering on the DCPCs.

9. The variables closest to the cluster centers will be the selected variables.

Algorithm 2: CLeVer
4.2 Proposed Stochastic Relevance Analysis

Relevance analysis aims to distinguish variables that are effectively representing the subjacent physiological phenomena (in the case of biosignals), named relevant features, according to some evaluation measure; and tries to reject variables with less contribution to the representation target (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 relevance \cite{30}. In this chapter the proposed methodology for stochastic feature selection, based on relevance analysis, is discussed.

4.2.1 Static Relevance

Given $X_\xi$, and for each one of the features $\xi_i \in \xi$, the relevance function $g$ is defined as follows:

$$ g : \mathbb{R}^{N \times p} \times \xi \to \mathbb{R}^+$$

$$(X_\xi, \xi_i) \mapsto g(X_\xi, \xi_i) \in \mathbb{R}^+$$  \hspace{1cm} (4.4)

Relevance Properties

For the introduced feature relevance function the following properties are established:

- **Non-negativity**, i.e. $g(X_\xi, \xi_i) \geq 0$, $\forall i \in [1, p]$.

- **Nullity**, the function $g(X_\xi, \xi_i)$ is zero if the feature $\xi_i$ has no relevance at all.

- **Non-redundancy**, if $\xi_i = \alpha \xi_j + \eta$, where the real-valued $\alpha \neq 0$ and $\eta$ is some standard Gaussian noise with mean zero and unit variance, then, $|g(X_\xi, \xi_i) - g(X_\xi, \xi_j)| = \epsilon$, where $\epsilon \to 0$.

A direct way to ensure a low redundancy between any pair of considered features, $\{\xi_i, \xi_j\}$, is verifying that the zero lag of their normalized correlation function ($\text{corr}\{,\}$) does not exceed a given small positive-valued
CHAPTER 4. STOCHASTIC FEATURE SUBSET SELECTION

threshold $\epsilon \rightarrow 0$, and thus the relevance is recomputed as follows:

$$g(X_{\xi}, \xi_i) = \begin{cases} 0, & |\text{corr}\{\xi_i, \xi_j\}| \leq 1 - \epsilon, \quad : i, j = 1, \ldots, p; i \neq j \\ g(X_{\xi}, \xi_i), & \text{otherwise} \end{cases}$$

(4.5)

**Feature Selection Criterion**

The value of $g(X_{\xi}, \xi_i)$ is called *relevance weight*. The main assumption in the proposed approach, is that the largest weight is associated to the most relevant feature. Hence, the weights are sorted according to their relevance weight, forming the relevance array:

$$g(X_{\xi}, \xi) = [g_1 \cdots g_i \cdots g_p]^\top \in \mathbb{R}^p, \quad g_i \geq g_{i+1}$$

(4.6)

Finally, the proper number of selected parameters having essential information is achieved by truncating the set of stochastic features given in (4.6). This subset is proposed to be assessed as follows:

$$\psi = \{\xi_i : \sum_{j=1}^q g_j^2 - \sum_{j=1}^p g_j^2 \leq \epsilon : i = 1, \ldots, q \leq p\}$$

(4.7)

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

### 4.2.2 Dynamic Relevance

Because of high computational cost of stochastic–feature–based training, dimensionality reduction of input spaces is carried out. In this sense, latent variable techniques are widely used, aiming to reduce the $p$–dimensional stochastic feature arrangement, $\Xi \in \mathbb{R}^{p \times T}$, into $q$–dimensional stochastic set, $\Psi \in \mathbb{R}^{q \times T}$, $q \leq p$; 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, $\{\xi_i \in \mathbb{R}^T\}$, and the input matrix $X_{\Xi}$, the relevance function $g(\Xi, \xi)$ is defined as follows:

$$g : \mathbb{R}^{N \times pT} \times \xi \rightarrow \mathbb{R}^T$$

$$(X_{\Xi}, \xi_i) \mapsto g(X_{\Xi}, \xi_i)$$

(4.8)
CHAPTER 4. STOCHASTIC FEATURE SUBSET SELECTION

Dynamic Relevance Properties

For the introduced feature relevance function the following properties are established:

- **Non-negativity**, i.e. \( g(X_\Xi, \xi_i) \geq 0, \forall i \in [1, p], \) and \( \forall t \in \mathbb{R}^T \)

- **Nullity**, the function \( g(X_\Xi, \xi_i) \) is zero if feature \( \xi_i \) has not relevance at all \( \forall t \in \mathbb{R}^T \).

- **Non-redundancy**, if \( \xi_i = \alpha \xi_j + \eta \), where the real-valued \( \alpha \neq 0 \) and \( \eta \) is some noise vector with mean zero and unit variance, then, both stochastic features could be grouped by an operator.

Dynamic Selection Criterion

The main assumption about dynamic relevance is that the less varying the relevance weight, the more relevant the feature. The proper number of selected parameters having essential information is achieved by inspecting the set of stochastic features, according with the following criterion:

\[
\Psi = \{ \xi_i : |\text{var}\{g(X_\Xi, \xi_i)\} - \min\{\text{var}\{g(X_\Xi, \xi_j)\}\}| \leq \epsilon, \forall j \neq i \} \tag{4.9}
\]

being var\{\} the variance operator and \( \epsilon \to 0 \).

Special Case: Highly Correlated Features

In the case of high dimensional data with considerable correlation among the whole set, like the continuous decompositions (\( time-\lambda \)) and its enhancement through filter–banked–stochastic features (e.g. the LFCC), could be improvement applying the dynamic relevance analysis. Usually, the mentioned features are computed by an heuristic approach as in the Algorithm 3. Nevertheless, this approach does not take advantage of the information about the irregular energy spatial distribution of the \( time-\lambda \) decomposition. In fact, it would be desirable to accomplish sub–band partitions enclosing spectral components with alike time–evolving behavior, i.e. holding similar dynamic relevance. For this purpose, the sub–band partition set can be determined by introducing the concept of relevance–time distribution measuring as the amount of useful information within every spectral component.

The Algorithm 4 describes the process for splitting the \( \lambda \) axis by means of the relevance measures.
CHAPTER 4. STOCHASTIC FEATURE SUBSET SELECTION

Input: $x(t)$
Output: Frequency bands

\begin{algorithm}
\begin{algorithmic}
\Procedure{Class $k$}{do}
\Procedure{Observation $i$}{do}
1. Calculate $t$-\(\lambda\) decomposition $(k,i) \in \mathbb{R}^{A \times T}$ of $x(t)$;
\end{algorithmic}
\end{algorithm}

\begin{algorithm}
\Procedure{Filter bank $j = 1 : n_{\Lambda_{\text{max}}}$}{do}
1. Compute $j$ stochastic features corresponding to $j$ triangular filters linearly spaced over the frequency domain;

\begin{algorithm}
\Procedure{Features subset $p = 1 : j$}{do}
1. Create a features subset corresponding to the first $p$ LFCC;
2. Dimension reduction of the input data;
3. Obtain the accuracy for this feature subset with a classifier
\end{algorithm}

\end{algorithm}
\end{algorithm}

Select the frequency bands ($n_{\Lambda}$ and $p$) when the accuracy rate is maximized;

Algorithm 3: Frequency bands selection by heuristic approach
Input: $x(t)$
Output: Frequency bands

foreach Class $k$ do
  foreach Observation $i$ do
    1. Calculate $\Xi$ for $t-\lambda$ decomposition;
  end
end

1. Calculate $X_\Xi$;
2. Calculate $g(X_\Xi)$;
3. Reshape the relevance vector into a relevance matrix
   
   $$G = [g(X_\Xi, \xi_1)| \ldots | g(X_\Xi, \xi_\lambda)| \ldots | g(X_\Xi, \xi_\Lambda)]^T \in \mathbb{R}^{A \times T};$$
4. Calculate the relevance matrix marginal along the time axis $\in \mathbb{R}^{A \times 1};$
5. Select the frequency bands where the relevance presents significant changes on its behavior i.e. the local minimum of the marginal with the lowest variance;

Algorithm 4: Frequency bands selection by relevance analysis
4.2.3 Relevance Weight

Unsupervised Measurement

For the sake of simplicity, the dimensionality reduction is developed using the simplest time–evolving latent variable approach. Given the observation matrix, \( X_\xi \), there will be a couple of orthonormal matrices, \( U, V \), as well as a diagonal matrix \( \Sigma_X \), so that the following linear decomposition takes place:

\[
X_\xi = U \Sigma_X V^T; U \in \mathbb{R}^{N \times N}, \Sigma_X \in \mathbb{R}^{N \times p_T}, V \in \mathbb{R}^{p_T \times p_T}, N > p_T
\]  

where \( \Sigma_X \) holds the first \( q_T \) relevant eigenvalues of matrix \( X_\xi \), \( \nu_1 \geq \nu_2, \ldots, \geq \nu_q \geq \nu_{q+1}, \ldots, \geq \nu_{p_T} \geq 0 \), with \( q_T < p_T \). It must be quoted that inherently to basic latent variable approach, the minimum mean squared error is assumed as the evaluation measure of transformation of subspaces, \( g(X_\xi) \sim \min E\{\|X_\xi - X_\psi\|_2\} \), (where \( \| \cdot \|_2 \) is the norm squared value), being \( X_\psi \) the reconstruction of \( X_\xi \), that is \( g(X_\xi) \sim \min E\{\|X_\xi - VX_\xi\|_2\} \), where \( V \) is the eigenvector set of diagonal covariance matrix holding the sorted singular values.

To make clear the contribution of each time–variant value \( \xi_{ij}(t) \), expression (4.10) can be further extended in the form:

\[
X_\xi = \sum_{j=1}^{q_T} \nu_j U_j V_j^T,
\]

where \( U_j \) is the \( j \)-th column of matrix \( U \). The magnitudes of the eigenvector entries that span the representation basis is then chosen as the relevance measure, i.e. \( g_j(\tau) = \sum_{k=1}^{q_T} \|\nu_k V_j(\tau)\|_2 \), where \( V_j \) is the \( j \)-th column of matrix \( V \) and \( \tau = 1, \ldots, p_T \). Therefore, the contribution of each \( j \)-th stochastic feature is given by \( g_j(t) = [g(j-1)T + 1 \ldots g(j-t)T + 1 \ldots g(jT)] \).

Supervised Measurement

The following supervised relevance measures for continuous representations are assessed as [31]:

a. **Linear Label-conditioned Correlation** that is given by

\[
g(S(t, \lambda)|e) = \left| \frac{E\{(S'(t, \lambda) - \overline{S}(t, \lambda))(c^i - \overline{c})\}}{\sqrt{E\{(S'(t, \lambda) - \overline{S}(t, \lambda))^2\}E\{(c^i - \overline{c})^2\}}} \right|, \quad (4.11)
\]

where \( \overline{S}(t, \lambda) = E\{S'(t, \lambda) : \forall i\} \), the measured value of \( S(t, \lambda) \) for the \( i \)-th object, \( i = 1, \ldots, M \), and \( \overline{c} = E\{c^i : \forall i\} \). Likewise, \( c^i \) is the label of
the $i$-th object given to the $S^i(t, \lambda)$. The notation $E\{\cdot : \forall \mu\}$ stands for the expectation operator over variable $\mu$.

b. Symmetrical Label-conditioned Uncertainty given by:

$$g(S(t, \lambda)|c) = \frac{H\{S^i(t, \lambda) : \forall i\} - H\{S^i(t, \lambda)|c^i\}}{H\{S^i(t, \lambda) : \forall i\} - H\{c^i : \forall i\}}$$

being $H\{\cdot : \forall \mu\}$ the entropy operator over variable $\mu$.

### 4.2.4 Summary

Table 4.1 summarizes the relevance measurements and the features selection algorithms tested in this work.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter</td>
<td>mRMR</td>
<td>Data, Labels</td>
</tr>
<tr>
<td></td>
<td>CLeVer</td>
<td>Data, Labels</td>
</tr>
<tr>
<td>Wrapper</td>
<td>RFE</td>
<td>Data, labels, SVM parameters</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relevance Measurement</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>Linear Correlation</td>
<td>Data, Labels, TFR</td>
</tr>
<tr>
<td></td>
<td>Symmetrical Uncertainty</td>
<td>Data, Labels, TFR</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>Maximum Variance</td>
<td>Data</td>
</tr>
</tbody>
</table>
Part III
Experiments
Chapter 5

Experimental Setup

In this chapter, the databases and validation methodologies are explained. The databases correspond to artificial data sets and real-world biosignal recordings. The proposed methodology is conformed by three main topics, according to the specific objectives. The first and the second topics are: to suggest a time-varying relevance function, and to fix a criterion for stochastic features subset selection. In this sense, the function and the criterion were described in the theoretical part of this thesis; nevertheless, aiming to prove the hypothesis proposed, an algorithm is developed for stochastic features subset selection and it is tested with synthetic and real-world data sets. Each on of the data sets represents a different application with specific constrains. Due to the differences between applications, the data sets are presented in individual chapters. The following sessions give a general idea of the problem statement for each application and a summary of the decompositions and results.

5.1 Artificial Data Set

5.1.1 Synthetic Control Chart Time Series

This study uses the Synthetic Control Chart Time Series from the UCI KDD archive [32]. This data set contains 600 examples of control charts synthetically generated by the process in Alcock and Manolopoulos [33]. There are six different classes of control charts: Normal, Cyclic, Increasing trend, Decreasing trend, Upward shift, Downward shift and every chart has 60 time
values [31]. The control chart patterns are artificially generated by six equations; each equation represents a different type of pattern [35]. The purpose of the experiment is to show a general idea of the proposed methodology. The signal representations and the best results for this data set are shown in Table 5.1.

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Parameters</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-Varying</td>
<td>rank$_L = 2$</td>
<td>94.50 ± 1.21%</td>
</tr>
<tr>
<td></td>
<td>rank$_L = 2$</td>
<td>99.05 ± 0.64%</td>
</tr>
<tr>
<td>Parametric Time-Varying</td>
<td>BIC = 6</td>
<td>65.83 ± 2.88%</td>
</tr>
<tr>
<td>Time–Frequency</td>
<td>512 Frequency Bins</td>
<td>84.83 ± 2.06</td>
</tr>
<tr>
<td></td>
<td>[2 – 128] Decomposition Levels</td>
<td>97.22 ± 1.22</td>
</tr>
</tbody>
</table>

5.2 Biosignal Recordings

5.2.1 Apnea - ECG Database

The analyzed data come from the public Physionet database. The whole-night electrocardiographic (ECG) recordings contain all the events occurring during a night including apneas, arousals, movements, and also some wakefulness episodes. Each one of the apnea events was labeled either as obstructive or mixed. One-minute segments containing hypopnoea were also scored as apnea events. Apnea scoring was carried out on the basis of standard criteria by an expert sleep clinician. The subjects in the Physionet database were classified into three classes: A, B, and C. A recording are related to class A (Apnea) if it contained at least one hour with an AHCI (index of sleep apnea severity) of 10 or more, and at least 100 apnoea episodes. The subjects on the class A are fifteen men and one woman, with a mean age of 50 years (29-63). A recording is related to class B (termed borderline) if it contained, at least, one hour with an apnoea index of 5 or more, and between 5 and 99 minutes with apnoea; class B are four men and one woman, with a mean age of 46 years (39-53). Recordings containing less than 5 minutes of disordered breathing were put in the normal (control, or class C) group. The C group consisted of six male and five female subjects with a mean of 33 years (27-42) [36].
Although the whole Physionet database contains a total of 70 ECG signals, the learning set for the present study holds only 50 recordings that are selected according to [37]. Particularly, ECG signals with a large number of ectopic beats (more than 10% of the beats within the recording partition) are not included in the present research. Then, 25 recordings are chosen to be used as the classification training set, whereas a second group with the other 25 recordings is used as a validation set. As a result, the training set consists of 4950 apneic one-minute segments and 7127 non-apnoeic one-minute segments, while the testing set holds 4428 apnoeic and 7927 non–apnoeic one-minute segments.

ECG recordings are digitized at 100 Hz with 16 bit resolution. Basically, automatic obstructive sleep apnea (OSA) diagnosis requires the extraction of the heart rate variability (HRV) time series from each ECG recording, which in this case can be estimated more precisely if an accurate recognition of the QRS complex fiducial points is achieved. In this work, QRS complex detection is carried out by the method proposed in [38], including the following preprocessing: i) linear filtering, ii) non linear transformation, and iii) adaptive decision rules. Further smoothing of anomaly valued peaks of assembled beat-to-beat interval time series is performed [39]. Then, the HRV time series is normalized, termed $x'(t)$, as recommended in [40]:

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

The task for the apnea data set has two main sub-tasks: the segment classification by minute and the subject classification. Table 5.2 summarizes the best results for both sub-tasks.

<table>
<thead>
<tr>
<th>Segment Classification</th>
<th>Parameters</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decomposition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-Varying SSA</td>
<td>rank$_L$ = 2</td>
<td>65.78 ± 1.99%</td>
</tr>
<tr>
<td>Time-Varying EDS</td>
<td>rank$_L$ = 2</td>
<td>76.17 ± 0.37%</td>
</tr>
<tr>
<td>Parametric Time-Varying DLM</td>
<td>BIC = 10</td>
<td>74.25 ± 0.42%</td>
</tr>
<tr>
<td>Time–Frequency STFT</td>
<td>512 Frequency Bins</td>
<td>75.64 ± 1.14%</td>
</tr>
<tr>
<td></td>
<td>CWT</td>
<td>$[2 - 512]$ Decomposition Levels</td>
</tr>
<tr>
<td><strong>Subjects Classification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STFT</td>
<td>30 minutes</td>
<td>100%</td>
</tr>
</tbody>
</table>
5.2.2 Apnea - Photoplethysmography (PPG) Database

This study uses a collection of PPG recordings of 21 children that were acquired over all-night-long sessions, as detailed described in [41]. The children aging within 4.5 ± 2 years were referred to the Miguel Servet Children’s Hospital in Zaragoza for suspected sleep–disordered breathing. Electroencephalographic electrode positions C3, C4, O1, and O2, chin electromyogram, electrocardiographic leads I and II, eye movements, airflow, as well as chest and abdominal respiratory efforts were recorded by a digital polygraph (BITMED EGP800), according to the standard procedure of the American Thoracic Society [42]. PPG and arterial oxygen saturation (SaO₂) were measured continuously using a pulse oximeter (COSMO ETCO2/SpO2 Monitor Novametrix, Medical Systems). All signals were recorded with a sample rate of 100 Hz, but the electrocardiographic biosignals were sampled at 500 Hz. OSA evaluation from polisomnographic (PSG) data were scored by clinical experts using the standard procedures and criteria given in [43]. Children often desaturated with short apneas, as they have a lower functional residual capacity and a faster respiratory rate than adults. Therefore, obstructive apneas of any length are scored when interpreting pediatric sleep studies, as compared with the 10-s duration in adults. Children may develop clinical sequelae with what appears to be relatively mild OSA. Thus, an apnea index of 10 is considered to be relatively mild by most pediatric pulmonologists, whereas it is considered only mildly abnormal in adults. One reason why a low apnea index can be associated with severe clinical disease is that the apnea index, the parameter used most often to characterize disordered breathing in adults, does not give an accurate picture of the nature of the breathing disturbance in children [44]. Thus, ten children were diagnosed with OSA whereas the remained eleven were diagnosed as normal. In this application, the STFT and its respective stochastic features were used as training set. The best result was obtained with the cepstral coefficients (Acc = 95.40%). For the subject classification, all the subsets of stochastic features achieve the same accuracy with 83.33%

5.2.3 Epilepsy - EEG Database

Database One (DB1)

The electroencephalographic (EEG) signals correspond to 29 patients with medically intractable focal epilepsy. Signals where recorded at the Depart-
ment of Epileptology of the University of Bonn, by means of intracranially implanted electrodes [45] according to the International 10−20 System of Electrode Placement Standard. Recordings were acquired by an acquisition system of 128 channels EEG unit, using average common reference. Data was digitized at 173.61 Hz and 12 bits of resolution. The database comprises five sets (denoted as Z,O,N,F,S) composed of 100 EEG segments per channel of 23.6 s (4096 samples), which were selected and extracted after visual inspection from continuous multichannel EEG to avoid artifacts (e.g. muscular activity or eye movements). Z and O data sets consist of segments taken from scalp EEG recordings of five healthy subjects using the standard 10−20 electrode placement. Volunteers were awake and relaxed, with their eyes open (Z) and closed (O). N, F and S data sets were selected from epilepsy diagnosed EEG recordings. The signals were selected from five patients who achieved a complete control of the epileptic episodes after the dissection of one of the hippocampal formations, which was correctly diagnosed as the epileptogenic zone. Segments of D and C data sets were recorded in the epileptogenic and hippocampal zone of the brain, respectively. While N and F data sets only contain activity measured on inter–ictal intervals, S data set only contains recordings with ictal activity. Since discontinuities between the end and the beginning of a time series are known to cause spurious spectral frequency components, segments of 4396 samples were at first cut out of the recordings. In this work, the two–class and the five–class problems are addressed through the proposed methodology.

Database Two (DB2)

This collection is obtained from the Instituto de Epilepsia y Parkinson del Eje Cafetero database. Each data set (Z and S) contains 160 recorded scalp EEG signals from 20 channels corresponding to the electrodes placed on the head according to the International 10−20 System of Electrode Placement Standard. Z data set holds 80 recordings labeled as normal (seizure–free), whereas S data set holds 80 recordings labeled as epileptic; in this case, a neurologist reviewed the EEG to mark all epileptic events. Recordings, which were done under video control in order to have an accurate determination of the different seizure stages, had been sampled at a frequency of 256 Hz and 12 bits of resolution and a 2 min duration. All patients were examined by a neurologist. The data had been acquired in a non–regulated condition, and the noised data holds, besides awake background EEG activity, muscle
artifacts as well as 60 Hz power line interference.

Table 5.3 summarizes the best results for both data sets, each one with its respective decomposition techniques.

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Parameters</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-Varying</td>
<td>SSA</td>
<td>rank $L = 15$</td>
</tr>
<tr>
<td></td>
<td>EDS</td>
<td>rank $L = 5$</td>
</tr>
<tr>
<td>Parametric Time-Varying</td>
<td>DLM</td>
<td>$BIC = 20$</td>
</tr>
<tr>
<td>Time–Frequency</td>
<td>STFT</td>
<td>512 Frequency Bins</td>
</tr>
<tr>
<td></td>
<td>CWT</td>
<td>[1 − 256] Decomposition Levels</td>
</tr>
<tr>
<td>DB1 Two–Class Problem</td>
<td></td>
<td>acc.</td>
</tr>
<tr>
<td>Time-Varying</td>
<td>SSA</td>
<td>rank $L = 15$</td>
</tr>
<tr>
<td></td>
<td>EDS</td>
<td>rank $L = 5$</td>
</tr>
<tr>
<td>Parametric Time-Varying</td>
<td>DLM</td>
<td>$BIC = 20$</td>
</tr>
<tr>
<td>Time–Frequency</td>
<td>STFT</td>
<td>512 Frequency Bins</td>
</tr>
<tr>
<td></td>
<td>CWT</td>
<td>[1 − 256] Decomposition Levels</td>
</tr>
<tr>
<td>DB2 Two–Class Problem</td>
<td></td>
<td>acc.</td>
</tr>
<tr>
<td>Time-Varying</td>
<td>SSA</td>
<td>rank $L = 15$</td>
</tr>
<tr>
<td></td>
<td>EDS</td>
<td>rank $L = 5$</td>
</tr>
<tr>
<td>Parametric Time-Varying</td>
<td>DLM</td>
<td>$BIC = 20$</td>
</tr>
<tr>
<td>Time–Frequency</td>
<td>STFT</td>
<td>512 Frequency Bins</td>
</tr>
<tr>
<td></td>
<td>CWT</td>
<td>[1 − 256] Decomposition Levels</td>
</tr>
</tbody>
</table>

5.2.4 Treadmill Exercise Test - HRV Database

Twenty-three males volunteered to participate in the study, between the 23 available recordings, 10 were randomly chosen for the experiments. Some population study characteristics are reported in Table 5.4. All of them were physically active with at least 3 days per week of regular aerobic training. Subjects were all familiar with exercise testing, not taking prescribed medications and presented with normal blood pressure levels and electrocardiographic patterns. None of the subjects were under performing complain about any type of fatigue. The subjects were fully informed about the purpose and possible risks and benefits of their participation in the study before
giving them the written informed consent. The study was approved by the University Human Ethics Committee according to the declaration of Helsinki.

<table>
<thead>
<tr>
<th>Age [yr]</th>
<th>Height [cm]</th>
<th>Mass [kg]</th>
<th>Body mass index [kg.m$^2$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.8 ± 5.0</td>
<td>178.4 ± 5.7</td>
<td>74.8 ± 7.8</td>
<td>23.5 ± 2.5</td>
</tr>
</tbody>
</table>

Table 5.4: Study Population Characteristics - Mean ± Standard Deviation

R–R intervals were recorded beat-to-beat using a heart rate (HR) monitor (RS800, Polar Electro Oy, Kempele, Finland), which uses a sampling frequency of 1 kHz for the ECG signal, providing an accuracy of 1 ms for each R–R period. Moreover, the HR monitor recorded HR and running stride frequency, using a stride sensor (S3, Polar Electro Oy, Kempele, Finland). The synchronization between the open-circuit sampling system and the HR monitor measurements was assessed using the HR recorded by both devices. The main idea with this data set, is to find the respiratory frequency in the HRV, and after that, compare the results with the respiratory frequency given by the jeager. As a result, the mean of the error between both frequencies is 3.842 ± 3.5187

5.3 Assessment Criteria

For classification task, the goal is to map the data into a feature space in which the members from different classes are clearly separated. When the input data dimension is relatively high, most classification methods suffer from curse of dimensionality and get highly biased estimations. However, high dimensional data often represent phenomena that are intrinsically low dimensional. Thus, the problem of high dimensional data classification can be solved by first mapping the original data into a lower dimensional space and then applying some classification technique.

The classification accuracy have to be tested with different sets for training and validation. These sets are chosen randomly from the full data set. In this work, 10-folds are used, splitting the 70% of the objects for training, and the remaining 30% for testing. The classification accuracy (Acc) is computed as the ratio of the number of well classification samples to the whole number of samples. Finally, the expected error rate is calculated as the average of the 10 folds accuracy results computed.
Chapter 6

Features Selection for Control Chart Time Series Classification

Aiming to illustrate the proposed methodology in the most general way, the synthetic control chart data set is used. Control chart patterns are time series that show the level of a machine parameter plotted against time [35]. This data set is often used for clustering and classification algorithms test [35, 34, 48], due to the fact that some pair of classes are usually confused [34]: normal and cyclic, decreasing trend and downward shift, and increasing trend and upward shift.

Figure 6.1: Proposed methodology - General scheme

Figure 6.1 shows a scheme of the proposed methodology. The first step consists in to find the more suitable signal representation. Although the decomposition techniques analyzed in this work lie on the hypothesis that any signal can be decomposed into a linear combination of harmonic, exponential, or even polynomial series; not all the representations are suitable for a determined signal. The proper decomposition technique depends of the intrinsic
CHAPTER 6. FEATURES SELECTION FOR CONTROL CHART TIME SERIES CLASSIFICATION

dynamic of the studied phenomena. Different decomposition techniques have to be proved due to the stochastic behavior of the signals of interest. With this idea in mind, seven approaches are considered: SSA, DLM, EDS, TFR, CWT and the continuous decompositions enhancement.

6.1 Time–Varying Decompositions

Tuning of the different schemes of considered time–varying decompositions throughout training procedure is carried out by using the average classification accuracy for the six-class problem, which is estimated using a \( k\)-nn (6 neighbors aprox.) classifier, followed by the 10-folds methodology.

Regarding to the SSA and EDS decomposition techniques, order 2 is chosen. In DLM the the Bayesian Information Criterion (BIC) is used with an order of 6.

Feature Subset Selection

Although the number of stochastic features in discrete decompositions is low, the purpose of this chapter is give a general idea about the proposed methodology.

RFE and mRMR

Both techniques are tested with a fixed number of stochastic–features time–instants (300). The new features are selected from the vectorized version of the input matrix.

CLeVer

Aiming to compare the proposed methodology with other outcomes, the CLeVer methodology is tested for the DLM decomposition. This approach was tuning with three centroids for the \( k\)-means algorithm. As a results, three stochastic features are chosen (features 1, 3, 6).

6.1.1 Relevance Analysis

In this line, an unsupervised relevance analysis is made for the DLM representation, aiming to find the most relevant dynamic features as follows:

Static Relevance

Figure 6.2 shows the static relevance analysis for DLM. It can be seen, that for SSA the relevance weight in the static analysis is not very selective, indicating that all the subset is needed for a good classification.
CHAPTER 6. FEATURES SELECTION FOR CONTROL CHART TIME SERIES CLASSIFICATION

Dynamic Relevance

Figure 6.2: Static relevance analysis for DLM decomposition technique

Figure 6.3: Dynamic relevance analysis for DLM decomposition technique

Figure ?? shows the dynamic relevance analysis for DLM. The stochastic features chosen are 1, 3, 5. In this case, the algorithm is selective enough and indicates that for DLM three dynamic features have enough information for the classification task.

6.1.2 Results of Time–Varying Decompositions

Tuning of the different schemes of considered signal decomposition throughout this training procedure is carried out by using the average classification
accuracy. The different approaches are tested and compared, using a \( k-nn \) (6 neighbors aprox.) classifier.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Acc [%]</th>
<th>Stochastic Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSA</td>
<td>94.50 ± 1.21%</td>
<td>2</td>
</tr>
<tr>
<td>DLM</td>
<td>66.38 ± 2.38%</td>
<td>6</td>
</tr>
<tr>
<td>DLM - CLeVer</td>
<td>64.83 ± 2.22</td>
<td>3</td>
</tr>
<tr>
<td>DLM - RFE</td>
<td>65.72 ± 3.34%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - mRMR</td>
<td>65.83 ± 2.88%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - Dynamic Relevance</td>
<td>65.05 ± 3.41%</td>
<td>3</td>
</tr>
<tr>
<td>EDS</td>
<td>99.05 ± 0.64%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6.1: Performance outcomes for time series analysis using a single tuned \( k-nn \) classifier.

Table 6.1 summarizes the performed classification accuracy for the different approaches and its respective set of stochastic features. As seen, the results could give an idea about the signal behavior. The higher accuracy is reached by the decompositions of lower orders, i.e. it is possible that the signal only has two significant dynamics as much. Nevertheless, this hypothesis has to be confirmed proving more complex decomposition techniques. Regarding to the stochastic features subset selection, the tested techniques give almost the same accuracy than the full set of DLM dynamic features. Nevertheless, an incorrect selection of the features could significantly decrease the accuracy reached; for example, if the selected features are 2, 4, 6 for the relevance analysis, the classification accuracy is 55.5 ± 4.30%.

### 6.2 Time–Frequency Decompositions

The methodology for this section has two components: \( i \) Continuous decompositions as stochastic features, and \( ii \) Continuous decomposition enhancement with spectral splitting. The continuous decompositions parameters are fixed as follows: The STFT–based quadratic spectrogram is computed by sliding a 13 bins Hamming windows, 50% overlapping, and 512 frequency bins. Regarding the CWT alternative, the respective time-scale (\( t-s \)) representation is performed using Daubechies2 as mother function, as recommended in [49]. The number of CWT decomposition levels is chosen within
the interval \([2 − 128]\), where the value 2 is regarded to the highest assumed frequency value, while value of 128 reflects the lowest frequency.

### 6.2.1 Continuous Decomposition Enhancement with Spectral Splitting

The frequency splitting for stochastic feature extraction can be achieved using pixel-wise relevance analysis. Three different measures are considered: relevance based on linear correlation, relevance based on symmetrical uncertainty and relevance based on maximum variance (unsupervised technique). The Algorithm \([4]\) explained above, describes the process for splitting the frequency axis by means of the relevance measures. Figure \([6.2]\) shows the behavior of the relevance for each point in the TFR maps and Table \([6.2]\) shows the frequency bands that are chosen. Each band includes 1 time series of vector cepstral coefficients that is computed by 1 triangular response filter.

<table>
<thead>
<tr>
<th>Relevance</th>
<th>Frequency Bands [Hz]</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Correlation</td>
<td>(0 − 0.12, 0.12 − 0.20, 0.20 − 0.33)</td>
<td>3</td>
</tr>
<tr>
<td>Symmetrical Uncertainty</td>
<td>(0 − 0.05, 0.05 − 0.33)</td>
<td>2</td>
</tr>
<tr>
<td>Unsupervised Relevance</td>
<td>(0 − 0.058, 0.058 − 0.13, 0.13 − 0.33)</td>
<td>3</td>
</tr>
</tbody>
</table>

The analysis of highly correlated features could be carried over the CWT decomposition, due to the fact that this is not necessarily linear. The splitting of the scale axis is achieved in a similar way, using the relevance obtained for each set of scales. Figure \([6.3]\) shows the relevance map of the CWT distribution, and Table \([6.3]\) shows the scale bands that are chosen. Each band includes 1 time series of vector cepstral coefficients that is computed by 1 triangular response filter.

<table>
<thead>
<tr>
<th>Relevance</th>
<th>Scale Bands</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Correlation</td>
<td>(0 − 14, 14 − 128)</td>
<td>2</td>
</tr>
<tr>
<td>Symmetrical Uncertainty</td>
<td>(0 − 22, 22 − 128)</td>
<td>2</td>
</tr>
<tr>
<td>Unsupervised Relevance</td>
<td>(0 − 14, 14 − 128)</td>
<td>2</td>
</tr>
</tbody>
</table>
CHAPTER 6. FEATURES SELECTION FOR CONTROL CHART TIME SERIES CLASSIFICATION

Figure 6.4: Selection of frequency bands using relevance analysis.
CHAPTER 6. FEATURES SELECTION FOR CONTROL CHART TIME SERIES CLASSIFICATION

Figure 6.5: Selection of frequency bands using relevance analysis.


6.2.2 Results of Time–Frequency Decompositions

Tuning of the different schemes of considered signal decomposition throughout the training procedure, is carried out by using the average classification accuracy. Because of high computational cost of stochastic-feature-based training, dimension reduction of the input space is carried out by means of a time-evolving latent variable technique, proposed in \[50\]. Then, considered approaches are tested and compared using a \(k\)-nn (6 neighbors aprox.) classifier.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Relevance</th>
<th># Features</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFR</td>
<td>Baseline</td>
<td>512</td>
<td>82.66 ± 1.67</td>
</tr>
<tr>
<td></td>
<td>Linear correlation</td>
<td>3</td>
<td>73.05 ± 3.81</td>
</tr>
<tr>
<td></td>
<td>Symmetrical Uncertainty</td>
<td>2</td>
<td>89.72 ± 2.14</td>
</tr>
<tr>
<td></td>
<td>Maximum Variance</td>
<td>3</td>
<td>84.83 ± 2.06</td>
</tr>
<tr>
<td>CWT</td>
<td>Baseline</td>
<td>64</td>
<td>99.00 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>Linear correlation</td>
<td>2</td>
<td>94.55 ± 2.07</td>
</tr>
<tr>
<td></td>
<td>Symmetrical Uncertainty</td>
<td>2</td>
<td>97.22 ± 1.22</td>
</tr>
<tr>
<td></td>
<td>Maximum Variance</td>
<td>3</td>
<td>94.55 ± 2.07</td>
</tr>
</tbody>
</table>

Table 6.4 summarizes the performed classification accuracy for the proposed approaches and their respective set of stochastic features. As a result, the stochastic features behavior and the association of different dynamics could represent an improvement or a decline of the classification accuracy. Regarding to the relevance technique for spectral splitting, the symmetrical uncertainty presents the best performance with a significant accuracy improve, compare to the baseline for both decompositions.

6.3 Discussion

- Proposed training approach for control chart classification has been tested. The best performance is provided by the EDS decomposition in the discrete case, and by the CWT in the continuous decomposition case.

- Regarding the order selection for SSA and EDS, it can be seen that adding complexity at the linear model SSA, the trend becomes more
CHAPTER 6. FEATURES SELECTION FOR CONTROL CHART TIME SERIES CLASSIFICATION

approximate to a simple non–linear model, as EDS. Both models represent the signal with a lowest order, i.e. the decomposition converges to the original signal through the highest principal eigenvalue associated to the Hankel matrix. Besides, if the order increases, the accuracy decreases; this can be explained because the other eigenvalues are less informative for the process, adding noise.

- In the case of DLM, despite being a TVAR model, its accuracy is one of the lowest tested decompositions, although the classification task for this data set is not complex. It can be concluded, that the dynamic imposed by the model requires a high order, near to the signal length, aiming to follow the signal trend as the other models. However, even with the above problem, the decomposition is able to find differences between the different kind of control charts. The techniques for features subset selection do not improve the performance; nevertheless the input space is lower than the original DLM full set.

Regarding to the type of relevance weight measure, it can be noted that the unsupervised measure and the linear correlation show a similar behaviors; nevertheless, the classification results are better in the case of symmetrical uncertainty. This measure is convenient for the classification task when the data is correctly labeled.
Chapter 7

Features Selection for Apnea Detection

The obstructive sleep apnea syndrome (OSA) is a common sleep disorder, characterized by the obstruction in the airflow. To perform an OSA diagnosis, detection of repetitive episodes of apnea and hypopnea during sleep is carried out, mostly, by attended overnight polysomnography in a sleep laboratory. However, regarding to standard polysomnography test, the main disadvantage is the high amount of information required to be analyzed [51, 52].

Heart rate variability (HRV) analysis is one of the promising directions for a simple, less costly, noninvasive, reliable and ambulatory screening method for OSA detection [53]. Nonetheless, HRV-based features analysis must deal with non-stationary signals (typical of apnea episodes), making clear the importance of using time-variant or time-frequency representation (TFR) [37]. Most of the reported TFR-based features are analyzed by static approaches, causing loss of valuable information in the time-evolving process. Conversely, extracted stochastic data might be analyzed, and thus, there is a need for a feature extraction approach being able to capture the dynamic information. In this line of analysis, the proposed methodology is tested for feature extraction based on a time-evolving version of the well-known principal component analysis (PCA) and for dynamic feature selection based on dynamic relevance [50].
CHAPTER 7. FEATURES SELECTION FOR APNEA DETECTION

7.1 Time–Varying Decompositions

The HRV signals are highly non–stationary, thus, a better signal representation is required. In this work, three approaches are considered for signal decomposition: SSA, EDS and DLM. Tuning of the different schemes of considered signal decomposition throughout training procedure is carried out by using the average classification accuracy for the automatic OSA detection, which is estimated using a \( k-nn \) (29 neighbors aprox.) classifier, followed by the well–known cross–validation methodology. Since the stochastic features for HRV have a low dimension, a relevance analysis has not place in this.

Singular Spectrum Analysis: for SSA tuning, the model order is chosen according to the classification accuracy. The lower the model order, then the better accuracy then, the order 2 is selected. Exponentially Damped Sinusoidal: the model order is chosen according to the classification accuracy. Since the lowest order presents a better result, then, order 2 is chosen for this model. Dynamic Linear Model: the Bayesian Information Criterion (BIC) is used for model selection, aiming to find the recurrent coefficient of the embedded time-varying autorregresive (TVAR) model, in terms of DLM form. The estimated average of the order given by BIC, is 10.

7.1.1 Results of Time–Varying Decompositions

Tuning of the different schemes of considered signal decomposition throughout this training procedure is carried out by using the average classification accuracy for automatic OSA detection. The different approaches are tested and compared, using a \( k–nn \) (29 neighbors aprox.) classifier.

<table>
<thead>
<tr>
<th>Time–varying decompositions (parametric and non–parametric)</th>
<th>Acc [%]</th>
<th>Stochastic Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSA</td>
<td>65.78 ± 1.99%</td>
<td>2</td>
</tr>
<tr>
<td>DLM</td>
<td>74.25 ± 0.42%</td>
<td>10</td>
</tr>
<tr>
<td>EDS</td>
<td>76.17 ± 0.37%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7.1: Performance outcomes for time series analysis using a single tuned \( k-nn \) classifier.

Table 7.1 summarizes the performed minute–by–minute classification accuracy for the different approaches and its respective set of stochastic features. As seen, the classification accuracies are statistically similar.
CHAPTER 7. FEATURES SELECTION FOR APNEA DETECTION

7.2 Time–Frequency Decompositions

Due to the high dimension of both representations, the experiments are divided into three steps: i) Time–frequency decompositions as stochastic features, ii) Time–frequency decompositions enhancement with linear tuning, iii) Time-frequency decomposition enhancement with spectral splitting.

7.2.1 Time–Frequency Decompositions as Stochastic Features

Based on the spectral properties of the HRV signal, the STFT–based quadratic spectrogram is computed by sliding Hamming windows for the following set of estimation parameters: 32.5 ms processing window length, 50% overlapping, and 512 frequency bins.

Regarding the CWT alternative, the respective time-scale ($t$–$s$) representation is performed using Complex Morlet wavelet as mother function, as recommended in [54]. Based on the dominant frequency components of the underlying HRV signal, the number of CWT 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. Figure 7.1 shows an example of both representations for HRV signals.

![Figure 7.1: Example of HRV time–frequency and time–scale representations.](image)
7.2.2 Time–Frequency Decomposition Enhancement

Parameter tuning for a considered TFR is achieved by the procedure developed for biosignals, discussed in [55]. As a 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 Eq. (3.18)) 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.

For the concrete case of OSA diagnosis, two approaches are tested: firstly, the analysis of static relevance; secondly, the dynamic relevance and its respective selection criterion.

Stochastic Features Subset Selection

Aiming to reduce the number of stochastic features (52), different algorithms are tested, i.e. RFE, mRMR, CLeVer, static and dynamic relevance. The main idea is to chose those features that are significant for the phenomena identification.

RFE and mRMR

Both techniques are tested with a fixed number of stochastic–features time–instants (300 non–stochastic features). The new features are selected from the vectorized version of the input matrix.

CLeVer

Due to the CLeVer algorithm is \textit{k–means}–based, when the number of stochastic features is high enough, the centroids of each cluster are not always the same. In this line of analysis, and for the apnea data set, the experiment is repeated 10 times with the training set, to try to find the main features. Table 7.2 shows the results for each iteration and the selected features for 7 clusters.

As can be seen, the stochastic features are mixed no matter its nature, i.e. the features 4, 6, 11 are energies, the feature 35 is a centroid and the features 45, 50 are cepstral coefficients.

Static Relevance Analysis

Figure 7.2(a) illustrates the attained estimates of normalized relevance weights for the training set. Estimation is performed after removing redundant stochastic features in accordance to Eq. (4.5). Weights are ordered by ordinal contour number, which are calculated
Table 7.2: CLeVer subsets for 10 iterations

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Stochastic Features Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{6, 11, 13, 35, 43, 45, 50}</td>
</tr>
<tr>
<td>2</td>
<td>{4, 12, 35, 39, 41, 45, 50}</td>
</tr>
<tr>
<td>3</td>
<td>{4, 6, 11, 13, 35, 45, 50}</td>
</tr>
<tr>
<td>4</td>
<td>{4, 6, 12, 35, 39, 45, 50}</td>
</tr>
<tr>
<td>5</td>
<td>{2, 10, 12, 35, 41, 46, 50}</td>
</tr>
<tr>
<td>6</td>
<td>{4, 6, 11, 13, 35, 45, 50}</td>
</tr>
<tr>
<td>7</td>
<td>{4, 6, 11, 18, 35, 45, 50}</td>
</tr>
<tr>
<td>8</td>
<td>{6, 11, 13, 35, 43, 45, 50}</td>
</tr>
<tr>
<td>9</td>
<td>{4, 6, 9, 16, 35, 45, 50}</td>
</tr>
<tr>
<td>10</td>
<td>{4, 6, 12, 18, 35, 45, 50}</td>
</tr>
</tbody>
</table>

Selected subset: {4, 6, 11, 35, 45, 50}

Figure 7.2: Static relevance weights based on non-supervised measurement when taking a partially divided set. Next, Figure 7.2(b) shows 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 7.2 are nonzero weighted, meaning that the other 7 parameters left are totally redundant. Then, the number of selected features for case of STFT enhancement is 45. According to relevance analysis, the most relevant group is the energy features set.

Dynamic Relevance Analysis  Figure 7.3(a) illustrates the attained estimates of the normalized relevance weights along time axis for the training set. Estimation is provided after removing redundant stochastic features in accordance to Eq. (4.5). Next, Figure 7.3(b) displays in detail weights, or-
CHAPTER 7. FEATURES SELECTION FOR APNEA DETECTION

Figure 7.3: On computing dynamic relevance weights based on non-supervised measurement

ordered by decreasing relevance (increasing variability), computed when taking into account the whole set of stochastic features. All 45 referred stochastic features in Figure 7.3 are nonzero weighted, meaning that the other 7 left stochastic features are redundant. As in the static analysis, the number of selected features for case of STFT enhancement is 45. According to relevance analysis, the least relevant group is the energy feature set.

7.2.3 Time–Frequency Decompositions Enhancement with Spectral Splitting

It must be noted that because of an easier medical interpretation, the spectral splitting over TFR maps is carried out separately for each one of the two bands of interest (LF and HF).

• Relevance–Based Approach

The frequency splitting for stochastic feature extraction can be achieved using pixel-wise relevance analysis. Three different measures are considered: relevance based on linear correlation, relevance based on symmetrical uncertainty and relevance based on maximum variance (unsupervised technique). The Algorithm 4 explained above, describes the process for splitting the frequency axis by means of the relevance measures. Figure 7.4 shows the behavior of the relevance for each point in the TFR maps and Table 7.3 shows the frequency bands that are
chosen. Each band includes 1 time series of vector cepstral coefficients that is computed by 1 triangular response filter.

Table 7.3: Frequency bands selection for spectral–splitting

<table>
<thead>
<tr>
<th>Relevance</th>
<th>Frequency Bands [Hz]</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Correlation</td>
<td>LF 0.04 – 0.05, 0.05 – 0.11 0.11 – 0.15</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>HF 0.15 – 0.24, 0.24 – 0.31 0.31 – 0.39, 0.39 – 0.50</td>
<td>4</td>
</tr>
<tr>
<td>Sym Uncertainty</td>
<td>LF 0.04 – 0.06, 0.06 – 0.08 0.08 – 0.12, 0.12 – 0.15</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>HF 0.15 – 0.19, 0.19 – 0.30 0.30 – 0.50</td>
<td>4</td>
</tr>
<tr>
<td>Unsupervised Relevance</td>
<td>LF 0.04 – 0.07, 0.07 – 0.10 0.10 – 0.15</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>HF 0.15 – 0.23, 0.23 – 0.34 0.34 – 0.46, 0.46 – 0.50</td>
<td>4</td>
</tr>
</tbody>
</table>

The analysis of highly correlated features could be carried over the CWT decomposition, due to the fact that this is not necessarily linear. The splitting of the scale axis is achieved in a similar way, using the relevance obtained for each set of scales. Figure 7.5 shows the relevance map of the CWT distribution. In this case only three bans are chosen, at: 2 – 413; 413 – 485; 485 – 512 scales.

According to the sub–band selection by spectral splitting, both representations of the HRV are decomposed using cepstral coefficients. In the process, for each time a triangular response filter is applied, obtaining 7 stochastic features for spectrogram and 3 stochastic features for scalogram.

7.2.4 Results of Time–Frequency Decompositions

Tuning of the different schemes of considered signal decomposition throughout the training procedure, is carried out by using the average classification accuracy for the automatic OSA detection. Because of high computational cost of stochastic-feature-based training, dimension reduction of the input space is carried out by means of a time-evolving latent variable technique, proposed in [50]. Then, considered approaches are tested and compared using a $k$–nn (29 neighbors aprox.) classifier.
### Table 7.4: Classification accuracy

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Approach</th>
<th># Features</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFR Baseline</td>
<td></td>
<td>512</td>
<td>76.04 ± 0.43</td>
</tr>
<tr>
<td>Linear grid, full set</td>
<td></td>
<td>45</td>
<td>75.75 ± 1.57</td>
</tr>
<tr>
<td>Linear grid, energy of centroids</td>
<td></td>
<td>13</td>
<td>72.13 ± 0.96</td>
</tr>
<tr>
<td>Linear grid, centroids</td>
<td></td>
<td>20</td>
<td>74.90 ± 1.91</td>
</tr>
<tr>
<td>Linear grid, cepstral coefficients</td>
<td></td>
<td>12</td>
<td>73.54 ± 1.12</td>
</tr>
<tr>
<td>Linear grid, RFE</td>
<td>300 non–stochastic</td>
<td></td>
<td>68.44 ± 1.22</td>
</tr>
<tr>
<td>Linear grid, mRMR</td>
<td>300 non–stochastic</td>
<td></td>
<td>71.82 ± 1.01</td>
</tr>
<tr>
<td>Linear grid, CLeVer</td>
<td></td>
<td>6</td>
<td>67.18 ± 1.36</td>
</tr>
<tr>
<td>Linear correlation</td>
<td></td>
<td>7</td>
<td>75.64 ± 1.14</td>
</tr>
<tr>
<td>Symmetrical Uncertainty</td>
<td></td>
<td>7</td>
<td>74.80 ± 0.68</td>
</tr>
<tr>
<td>Maximum Variance</td>
<td></td>
<td>7</td>
<td>75.22 ± 0.58</td>
</tr>
<tr>
<td>CWT Baseline</td>
<td></td>
<td>512</td>
<td>74.19 ± 0.47</td>
</tr>
<tr>
<td>Maximum Variance</td>
<td></td>
<td>3</td>
<td>77.10 ± 0.58</td>
</tr>
</tbody>
</table>

Table 7.4 summarizes the performed minute–by–minute classification accuracy for the proposed approaches and their respective set of stochastic features. As a result, the stochastic features behavior and the association of different dynamics could represent an improvement or a decline of the classification accuracy.

### 7.2.5 Classification Improvement

For the classification improvement, the spectral splitting over TFR is chosen for medical interpretation simplicity. The principal hypothesis of such a modest training contribution (see Table 7.4) when gathering both bands of interest (LF,HF) may be the observed difference between them in terms of stochastic behavior (See Fig. 7.2). For this reason, the parallel combining \( k-nn \) classifier with median selection rule is used, aiming to improve the performance of the proposed approaches, thus, each dynamic feature subset is used separately. For both multi–band splitting approaches, the best performance (\( \text{Accuracy} \sim 80.6\% \), \( \text{Sensibility} \sim 76.2 \), \( \text{Specificity} \sim 82.2 \)) of the combining classifier that is achieved over testing data–set is shown in Table 7.5. Nonetheless, some degradation of performed sensitivity on validation set is fixed, this is due to the difficulty of properly labeling the apneic one–minute episodes by clinic professionals.
### Table 7.5: Best performance outcomes assessed for spectral splitting.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Acc [%]</th>
<th>Se [%]</th>
<th>Sp [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>Maximum variance</td>
<td>81.17 ± 0.88</td>
<td>83.11 ± 1.59</td>
</tr>
<tr>
<td>Testing set</td>
<td>Maximum variance</td>
<td>80.19</td>
<td>76.44</td>
</tr>
</tbody>
</table>

#### Scoring of all–night recordings

Each recording is diagnosed to be related to either class grounded on decisions that are attained for the corresponding patient set of one–minute segments with the best scheme proved. Fig. ?? shows class separation based on OSA detection of one–minute segment within a single all–night recording (horizontal axis corresponds to the cardinal recording number). The circle, cross and star label represents the apneic subjects, borderline class, and normal subjects, respectively. As seen for either splitting approaches, a complete separation between normal and pathological classes can be achieved using a minimum set of 30 apnoea segments of 1-minute length per a single all–night recording. Yet, the borderline recordings are spuriously located in class A or C, so their adequate interpretation remains an open issue.

#### 7.3 Discussion

- Proposed training approach for OSA detection has been tested without major restrictions on the preprocessing stage (artifact removing, denoising). Besides, attained estimation of HRV time series, based on QRS complex detection [38], provides enough classification accuracy, which makes it suitable for automatizing the suggested methodology for OSA diagnosing.

- In the case of DLM, despite being a TVAR model, its accuracy is one of the lowest tested decompositions. The model order average is not an adequate estimator due to the difference between model orders in the training group, which is over five points. Additionally, the dynamic imposed by the model requires a high order, near to the signal length, aiming to follow the signal trend as the other models. However, even with the above problem, the decomposition is able to find differences between control and apneic events with an acceptable accuracy.
Some issues should be considered before involving other TF representations within the discussed relevance-based spectral splitting framework. Namely, the Wavelet Packet Analysis that is free of any assumptions about the signal stationarity, had been used for evaluation of HRV sub-frequency regions in supraventricular tachyrhythmia patients [56]. But, when the fine and coarse resolution components are computed, the resulting sub-band coefficient number becomes different for each node, while the proposed methodology based on the spectral splitting scheme, requires an equal number of samples per time-evolving spectral component. Likewise, the Empirical Mode Decomposition, but its implementation contains heuristic and ad hoc elements that make it hard to analyze mathematically [57].

Several studies have established the discriminating capability of HRV frequency bands between normal and sleep apnea patients [52], thus, the set of considered short-time measures needs to be related to the suitable time-variant features estimated by spectral sub-band methods. Namely, frequency cepstral coefficients are introduced, which allow a better energy distribution identification, emphasizing the different frequency components improving OSA detection. It is worth noting that assessed feature set does not require further processing, by instance, smooth filtering over time windows over exceeding the assumed non-overlapping 1-minute HRV segment. This issue becomes important to get a well defined methodology of HRV processing, as demanded in [39]. In addition, since more efforts should be done to define the features carrying fundamental information for the OSA classification, as quoted in [37, 50], then the set of considered features can fulfill with this requirement because of their easier filter-banked interpretation.

The main problem founded with the comparison between the different selection features schemes, was that the initialization of the algorithms requires a fixed number of features to be preserved at the end of the process. The initial values fixed for CLeVer, mRMR and RFE are a random-kind, because of the uncertainty about how many features are needed in order to obtain a high classification accuracy.

Regarding the comparison among static relevance and dynamic relevance analysis, it can be seen that the subset of centroid coefficients achieves the better accuracy performance. This result is coherent with
chapter 7. features selection for apnea detection

the dynamic analysis, regardless of their exhibited lower values of relevance with the static analysis. At the same time, coefficients of centroid energy turn out to have the worse accuracy, even when they reached higher relevance weights in the relevance static analysis. The main issue with the energy features, is that this features have strong dynamics, in this line the feature could have a high relevance in some point at the time, but in another point could be non–significant at all. For this reason, the variance measure of the weights is important when the studied phenomena has so strong behavior.

present study introduces the spectral splitting over the HRV spectrogram to optimize the performance of the dynamic filter-banked feature set in terms of improving the OSA detection accuracy. Both considered approaches for band partitioning (heuristic and relevance–based) achieve a small feature number (10 cepstral coefficients for heuristic and 7 for latter splitting approach). In addition, since the extracted feature set is directly related to HRV sub–bands, the discussed training methodology provides, as additional benefit, an easier clinical interpretation of HRV-derived parameters.

since heuristic splitting approach does not take advantage of the distributed information over spectrogram, an unsupervised measure of the stochastic variability is introduced. As a result, a relevance matrix is achieved that determining the relevance related to each one of the spectral components along the time axis. As shown in table 7.3 both obtained spectral grids are similar, concerning the estimated band partition boundaries, for both considered bands of interest. Detailed visual inspection of estimated relevance matrix in figure 7.2 shows that the most relevant information, for OSA detection, is the LF sub–band partition ranging from 0.07 to 0.1 Hz, as well as the HF sub-band between 0.34 and 0.46 Hz. Both found relevant sub–bands can be supported by previous results obtained through different methodologies [37, 39].

the used time-series-enhancement by TF representations should be regarded as an important factor for adequate filter–banked feature generation. In this study, the dynamic filter–banked feature set is extracted from Fourier–based spectrogram that have been reported to be appropriate for OSA analysis [58]. The best achieved performed
(Ac \sim 80.6\%, Se \sim 76.2\%, Sp \sim 82.2\%, respectively) by using the proposed spectrogram–based feature set are comparable with other results reported by different outcomes \cite{59}. Generally, the spectrogram is desirable for signals with a slow time–varying spectrum, but suffers of the TF resolution compromise.

– Even when the classifier optimization stage is beyond the scope of the present study, the HRV segment classification boosting is performed by a simple parallel combining \textit{k–nn} classifier with median selection rule. As a result, assessed cepstral coefficients set reaches as much as an accuracy of 80\% top, for both considered approaches of spectral splitting.

Regarding to the type of relevance weight measure, it can be noted that the three different measures show different relevance behaviors (see Table 7.3); nevertheless, the classification results do not change significantly. Therefore, it is difficult to select one single measure as the most appropriate. One important difference between supervised an unsupervised measures, is that the measure based on maximum–variance is computed taking into account the influence of each variable over the whole set of features, and not only the relation between the data and its respective label class; so, this relevance measure could be directly associated with the signal dynamic, which is convenient for the concrete case of spectral splitting.
CHAPTER 7. FEATURES SELECTION FOR APNEA DETECTION

Figure 7.4: Selection of frequency bands using relevance analysis for OSA diagnosis.
CHAPTER 7. FEATURES SELECTION FOR APNEA DETECTION

Figure 7.5: Scalogram enhancement for OSA diagnosis

Figure 7.6: Class distribution
Chapter 8

Time–Frequency Decomposition Enhancement for Apnea Patient Detection

One promising alternative for OSA diagnosis is the pulse photoplethysmography signal (PPG) that is a simple, but useful, method for measuring the pulsatile component of the heartbeat. PPG measurement evaluates peripheral circulation, and is tie–related either to arterial vasoconstriction or vasodilatation generated by the autonomic nervous system, being modulated by the heart cycle. Furthermore, automatic detection of time–variant decreases in the amplitude fluctuations of PPG have shown to be useful for OSA diagnosis \cite{60, 41, 61}.

Nonetheless, since there is a large number of situations when PPG envelope is affected independently of the apnea status, then, a low ratio sensitivity/specificity is accomplished. Therefore, to better discriminate between apnea from other PPG envelope alterations, an improved set of representing features should be taken into account, particularly, stochastic modeling of dynamic features for OSA detection is to be further considered in this chapter.

The aim of this study is to select a set of relevant dynamic features, extracted from TFR of time–dependent PPG envelope signals to increase the apnea detector specificity. This work analyzes the set comprising filter banked dynamic features that includes spectral centroids as well as the cepstral coefficients. Specifically, a time–evolving version of the standard linear multivariate decomposition is discussed throughout this chapter to perform
stochastic dimensionality reduction of the dynamic features in hand.

8.1 Preprocessing

Artifact Removal

It has been established that PPG measurements are quite sensitive to patient and probe-tissue movement artifact. Removal of such motion artifact as well as its separation from proper quality, although highly variable, pulse recordings is a non-trivial signal processing exercise \[62\]. To cope with this drawback, the Hjorth artifact detector is used. The principle behind the detector is that when the PPG signal differs largely from an oscillatory signal, it is very likely an artifact. Hjorth parameter has been proposed as an estimation of the central frequency of a signal and as half of the bandwidth. Further details of used artifact removal procedure are explained in \[60\].

Labeling of PPG Envelope Recordings

The PPG envelope, \( y(t) \), is estimated based on the root mean square series of input PPG signal, \( y_{PPG}(t) \). So, the discrete version of PPG envelope, after mean removal by a moving average filter, can be written as follows \[60\]:

\[
y(n) = \sqrt{\frac{1}{N} \sum_{k=n-(N-1)}^{n} \left( y_{PPG}(k) - \frac{1}{M} \sum_{l=k-(M-1)}^{k} y_{PPG}(l) \right)^2}
\]

where the windows length values for the filter, \( M \), and the root mean square series, \( N \), are fixed as 25 and twice the mean cardiac cycle, respectively.

It is worth noting that the discussed automated system for OSA diagnosing is based on analysis of set of fragments that are partitioned from the PPG envelope recordings. The OSA diagnostic labeling of PSG recording database has been made by experts after clinical analysis of the considered children patient group. The recordings that in average last as much as 8 hours are firstly partitioned into fragments of two different considered lengths: 15 or 60 minutes. Each fragment of either length is labeled using a decision rule based on SaO\(_2\) signal which has been simultaneously measured in time. Moreover, because of computational load the fragments are partitioned again into 90 s segments. Each 90-seconds frame is given the same label of the respective PPG fragment from where the segment has been extracted. Then, labeling of
CHAPTER 8. TIME–FREQUENCY DECOMPOSITION ENHANCEMENT FOR APNEA PATIENTS

partitioned PPG envelope recordings is provided according to the following procedures:

1. **Fragment Labeling.** In general, pathological patients can have some time periods related to both apneas and oxygen desaturation, but, they can also exhibit some normal periods without any respiratory problems. So, regarding subject diagnosis, it is useful to consider PSG fragments as a whole entity, then, a subject classification is carried out based on the number of PSG fragments related to apneic periods. The length of considered fragments is a trade-off between fragments and subject classification. In this study, both 15-minutes and 1-hour PSG fragments are considered, as recommended in [41]. The assessed set of PSG fragments is labeled as follows:

   At the beginning, a baseline level $\beta$ is fixed for each patient corresponding to the SaO$_2$ signal mode of the entire night recording. Then, the total time intervals with SaO$_2$ signal below $\beta - 3\%$, $t_{\beta-3}$ is calculated for each PSG fragment. Polysomnographic fragments of either length, 15-minutes or 1-hour, are labeled according to the following criteria:

   $$
   \begin{align*}
   t_{\beta-3} < 0.9 \text{ minutes}, & \quad \text{control} \\
   0.9 \text{ minutes} < t_{\beta-3} < 3 \text{ minutes}, & \quad \text{doubt} \\
   t_{\beta-3} > 3 \text{ minutes}, & \quad \text{pathological}
   \end{align*}
   $$

   The above imposed criterion implies a minimum of 5% of the time with evident oxygen desaturation to be considered as pathological. The assumed threshold corresponds to a severe OSA criterion in children of 18 apneas/hour having a mean duration of 10 s. In case of control group, that threshold is fixed to be 5 apneas/hour. As a result, the following data set of labeled fragments per considered class is assessed: control (70), doubt (24), and pathological (11), when considering 1-hour PSG fragments, whereas the set of control (326), doubt (47) and pathological (47) is achieved for 15-minutes PSG fragments; each one also labeled according to Eq. (8.1).

2. **Segment Labeling**

   Since each considered fragment of either length (15 minutes or one-hour) turns out to be very long to provide computational stability when
implementing discussed time–adapted PCA approach, then, PPG signals should be partitioned into processing time windows of shorter duration (termed segments). Seeing that each signal partition should comprise enough heart beats (see Figure 8.1), and taking into account that artifacts rarely last more than 60 s, the segment length is fixed empirically to be 90 s. Further, every 90-seconds segment is given the same label as the respective PPG fragment, wherein the partition is included. Nonetheless, there is a need for further clustering procedure to ensure that the assessed set of PPG segments are properly labeled. After carried on bi-class clustering (one cluster per class, control or apneic), by using the algorithm discussed in \cite{63}, distanced far enough from both cluster centroids are removed from present analysis. So, the group of remaining segments adequately labeled becomes herein the training set. In the test set, all the database segments are considered. Table 8.1 summarizes the amount of 90-seconds segments accomplished for both cases of considered PPG signal length: after artifact removal (*), and after clustering (**), which becomes the considered training set.

Figure 8.1: Histogram of heart beat rate per minute for a given set of labeled PPG fragments

8.2 Continuous Decompositions

Figure 8.2 illustrates examples of estimated enhanced TFR being performed for normal and pathological partitions. It can be seen that there are some normal segments whose waveform resemble like pathological ones, and vice versa. Since the selection of the appropriate TFR is required, tuning of the
Table 8.1: Amount of 90-seconds partitions accomplished for both cases of labeled PPG signal length

<table>
<thead>
<tr>
<th>Labeled PPG signal of 60-minutes-length</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical OSA diagnosis</td>
<td># Segments(*)</td>
<td># Segments(***)</td>
</tr>
<tr>
<td>Normal</td>
<td>2618</td>
<td>1908</td>
</tr>
<tr>
<td>Pathological</td>
<td>416</td>
<td>293</td>
</tr>
<tr>
<td>Assembled set</td>
<td>3034</td>
<td>2201</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labeled PPG signal of 15-minutes-length</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>2046</td>
</tr>
<tr>
<td>Pathological</td>
<td>409</td>
</tr>
<tr>
<td>Assembled set</td>
<td>2455</td>
</tr>
</tbody>
</table>

respective parameters is achieved by the procedure developed for biosignals discussed in [55]. Based on above explained spectral PPG envelope properties, the STFT–based quadratic spectrogram is computed by sliding Hamming windows for the following set of TFR estimation parameters: 37.5 ms processing window length, 50 % of overlapping, and 512 frequency bins.

Figure 8.2: Example of time–frequency representations.

8.2.1 Time–Frequency decomposition enhancement

Tuning with Linear Grid  Another aspect worthy of explicit attention is the generation of TFR–based dynamic features to be under study. Specifically for the present database, procedures for computation of cepstral coef-
ficients and centroids are similar; in the sense that both split the frequency into a fixed number of bands [55]. Respect to the coefficients calculation, given in Eq. (3.16) and (3.17), the following working parameters are to be determined, the initial number of time-variant features, the number of filter banks, the filter impulse response and its overlap over frequency domain. Nonetheless, it should be remarked that the tuning of the initial number of dynamic features is not a critical issue for the proposed training methodology since this amount is to be refined next by the relevance analysis.

Therefore, in accordance to the accuracy reached for a standard k–nn (27 neighbors aprox.) classifier, as shown in Figure 8.3, the input data space includes the following 39 TFR–based dynamic features to be further studied: the first 22 spectral centroids and their respective energy (estimated by using Hamming filters with 30% overlap, linear response distribution, and fixing $\gamma = 1$), and the first 17 time series of vector cepstral coefficients (computed by 48 triangular response filters with 50% overlap).

Static Relevance Analysis Figure 8.4(a) illustrates the attained estimates of normalized relevance weights for the training set. The relevance weights, that are calculated when taking a partially divided set, are ordered by ordinal contour number. Next, Figure 8.4(b) depicts in detail the weights, ordered by decreasing relevance, computed when taking into account the whole set of stochastic features. All 39 referred stochastic features in Figure 8.4 are nonzero weighted. Then, the number of selected features for the
Figure 8.4: On computing relevance weights based on static relevance case of STFT enhancement is 39. According to relevance analysis, the most relevant group is the energy feature set.

Figure 8.5: On computing dynamic relevance weights based on non-supervised measurement

**Dynamic Relevance Analysis** Figure 8.5(b) shows in detail the weights, computed when taking into account the whole set of stochastic features, ordered by decreasing variability. According to dynamic relevance analysis, the most relevant group is the LFCC features set.
CHAPTER 8. TIME–FREQUENCY DECOMPOSITION ENHANCEMENT FOR APNEA PATIENT

8.2.2 Results of Time–Frequency Decomposition Enhancement

It should be remarked that in this work, and because of the reduced input data assemble, some recordings are used for both training and validation. Therefore, for testing the classifier the apparent accuracy is assessed, that is performed by using \( k\)-nn classifier \((k = 3)\), as shown in Table 8.2.

<table>
<thead>
<tr>
<th>Dynamic feature set</th>
<th>Classification for 60-m-length</th>
<th></th>
<th>Classification for 15-m-length</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( Se ) [%] ( Sp ) [%] Acc [%]</td>
<td>( Se ) [%] ( Sp ) [%] Acc [%]</td>
<td>( Se ) [%] ( Sp ) [%] Acc [%]</td>
<td></td>
</tr>
<tr>
<td>Energy of Centroids</td>
<td>81.82 94.29 92.59 95.74 54.60 59.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centroids</td>
<td>90.91 100 98.77 91.49 95.40 94.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFCC</td>
<td>100 85.71 87.65 93.62 95.40 95.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full set</td>
<td>100 100 100 97.98 93.56 93.35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The decision threshold is proposed to be adjusted based on the ROC curve for patient classification, as shown in Figure 8.6. The location where the ROC curve gets the better classification accuracy points out to the decision threshold.

Lastly, each patient is diagnosed based on those decisions made from the set of fragments measured for himself. A rule to determine when a patient with a given number of pathological fragments is considered as a pathological subject is needed. To do this, the percentage of time under pathological fragments is considered and the threshold is selected as the one maximizing \( Se \) and \( Sp \) ratio at the ROC curve.

![Figure 8.6: Performed ROC curves for both considered fragment lengths](image-url)
Table 8.3: Classification of patient for training based on partially divided set of dynamic features

<table>
<thead>
<tr>
<th>Dynamic feature set</th>
<th>Se [%]</th>
<th>Sp [%]</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy of Centroids</td>
<td>70.00</td>
<td>87.50</td>
<td>73.68</td>
</tr>
<tr>
<td>Centroids</td>
<td>80.00</td>
<td>87.50</td>
<td>83.33</td>
</tr>
<tr>
<td>LFCC</td>
<td>90.00</td>
<td>75.00</td>
<td>83.33</td>
</tr>
<tr>
<td>Full set</td>
<td>80.00</td>
<td>87.50</td>
<td>83.33</td>
</tr>
</tbody>
</table>

Table 8.3 summarizes the performed patient classification accuracy for both considered combining approaches of dynamic features (partial and full set). In accordance with the discussed approach of relevance analysis, the LFCC and Centroids subsets of dynamic features reach the best accuracy that is similar to the one achieved for the whole training set. As a result, both sets should be strongly considered for OSA diagnosing with the advantage that the each performed time-evolving parameter is related to a fixed spectral sub-band, and thus, leading to an easier clinical interpretation. It must be quoted that displayed outcomes of accuracy shown in Table 8.3 are performed just when considering 60 m length fragments. In case of 15 m length, and if taking into consideration the full set of dynamic features, the overall performance is the following: \( Se = 90\% \), \( Sp = 62.5\% \), and \( Acc = 77.78\% \), which is significantly lower than those assessed outcomes for training for 60 m length fragments.

8.3 Discussion

- The enhanced parameter estimation carried out by introducing TFR should be regarded as a remarkable factor for an adequate generation of any set of dynamic features. Here, feature enhancement is performed by means of nonparametric spectrogram–based TFR been reported to be appropriate for the analysis of non-stationary biological signals consisting of different frequency components.

- For the discussed methodology for OSA detection, the TFR enhancement can be performed by using more elaborated approaches: wavelet–based scalograms, projection pursuit, by using time frequency distribu-
- Regarding to feature extraction and selection, proposed methodology for dynamic relevance is based on time-adapted linear component approach. At this point, two main issues are considered: the 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 PCA version is discussed throughout this paper as an unsupervised method to perform the relevance analysis on a considered set of stochastic features. Additionally, the proposed methodology of relevance analysis can be extended to other techniques such as linear component decomposition, as shown in [64].

- Two different combining approaches for selecting the best set of contours are studied: firstly, when taking a partially divided set that relates dynamic features having the same principle of generation. Secondly, when the best features are chosen despite of their physical meaning. From performed accuracy, shown in Table 8.2 it can be concluded that even that the former case reaches comparable accuracy values, the latter selection approach is more commonly used because of the convenient physical interpretation of the selected feature set. Furthermore, it has been established that the set of LFCC dynamic features should be strongly considered for OSA diagnosis. Performed outcomes bring enough evidence that if using a subset of LFCC features a fragment classification accuracy can reach as much as a 93%, which provides an adequate scheme for ambulatory OSA diagnosis. Therefore, to take into account evolution of random biological variables along time, definitively, leads to an accuracy improvement of OSA detection. Nonetheless, more efforts should be done to define feature set carrying fundamental information for the OSA classification, as quoted in [37]. Though, performed outcomes look very promising in terms of classification accuracy, testing of the discussed methodology should be
provided using larger data sets.

- The discussed automated system for OSA diagnosis is based on the analysis of fragments, being partitioned from the PPG envelope recordings. In this regard, labeling of partitioned PPG envelope recordings is provided so to have time epochs identified as apneic or not apneic. However, in clinical applications usually the interest lies in having a subject diagnosis related to apnea, both in adults [63] and children [61], and not just a time screening of the apnea events. With this aim, a rule has being applied to the fragment labeling, providing subject specific diagnosis. Comparison with PSG clinical decision is provided, showing the potential of the methods here presented. As a result, PPG can be considered as a promising alternative to reduce the number of the PSG sleep recordings.
Chapter 9

Features Selection for Epilepsy Detection

The electroencephalographic (EEG) signals represent the clinical signs of the synchronous activity of the neurons in the brain, but in case of epileptic seizures, there is a sudden and recurrent mal–function of the brain that exhibits considerable short–term non–stationarity that can be detected analyzing these recordings [66]. Thus, for example, the possibility to automatically detect epileptic seizures from EEG signals is limited by the wide variety of frequencies, amplitudes, spikes, and waves that use to appear along time with no precise localization [67]. Thereby, the performance of automatic decision support systems depends on features that parametrize accurately the non–stationary behaviors. Thus, a current challenging problem is to detect a variety of non–stationary biosignal activities with a low computational complexity, to provide tools for efficient biosignal database management and annotation.

In this chapter, three approaches are proposed: a) Epileptic seizure detection using dynamic features extracted directly from the EEG decomposition, b) Epileptic seizure detection based on dynamic relevance over the EEG TF and WT maps, c) Epileptic seizure detection using the knowledge about the EEG–rhythms,by processing them individually with the best scheme, a) or b) .
CHAPTER 9. FEATURES SELECTION FOR EPILEPSY DETECTION

9.1 Preprocessing

According to [68], due to the activity of brain cells, brain-waves of interest are generated having a frequency of around 3 – 30 Hz. The signal should be normalized prior to any analysis on the EEG waves aiming to reject undesired signals. The filtering is performed by a four-pole elliptic band-pass filter. Since the signal is filtered, some components in the stopped band are remaining due to the spectral response of the filter.

9.2 Time-Varying Decompositions

In this work, for signal decomposition, three approaches are considered: the signal is decomposed directly through SSA, EDS and DLM. Tuning of the different schemes of considered signal decomposition throughout this training procedure is carried out by using the average classification accuracy for the epileptic seizures detection, which is estimated using a $k$-nn (5 neighbors approx.) classifier, followed by the 10-folds validation methodology as in Chapter 3.

Singular Spectrum Analysis: for SSA tuning, the model order is chosen according to the classification accuracy. The number of stochastic features for this decomposition is 15.

Dynamic Linear Model: the Bayesian information criterion (BIC) is used as order estimator, with the aim to find the recurrent coefficient of the embedded TVAR model, in terms of DLM form. The estimated mean of the order, given by BIC, is 20.

Exponentially Damped Sinusoidal: likewise, the model order is chosen according to the classification accuracy. Order 5 is chosen for this model.

9.2.1 Stochastic Features Subset Selection

Aiming to reduce the number of stochastic features, different algorithms are tested. The main idea is to chose those features that are significant for the phenomena identification.
CHAPTER 9. FEATURES SELECTION FOR EPILEPSY DETECTION

RFE and mRMR  Both techniques are tested with a fixed number of stochastic–features time–instants (300). The new features are selected from the vectorized version of the input matrix.

CLEVer  Due to the CLeVer algorithm is k–means–based, when the number of stochastic features is high enough, the centroids of each cluster are not always the same. In this line of analysis, and for the EEG data set, the experiment is repeated 10 times with the training set, to try to find the main features. After the selection, 8 stochastic features (\{1,2,3,4,5,6,7,9\}) are chosen for the classification step.

9.2.2 Relevance Analysis

Although the number of stochastic features in discrete decompositions is low, the number of time instants in each one is high enough. With this in mind, a relevance analysis is made for SSA and DLM, aiming to find the most relevant dynamic features as follows:

![Figure 9.1: Static relevance analysis for stochastic features selection](image)

(a) SSA  (b) DLM

Figure 9.1: Static relevance analysis for stochastic features selection

Static Relevance  Figure 9.1 shows the static relevance analysis for both cases in DB1, DLM and SSA. For DLM, the stochastic features chosen are those with a relevance weight greater than 0.9. For SSA, the stochastic features chosen are those with a relevance weight greater than 0.7. It can be seen, that for SSA the relevance weight in the static analysis is not very selective, indicating that almost all the subset is needed for a good classification.
CHAPTER 9. FEATURES SELECTION FOR EPILEPSY DETECTION

Dynamic Relevance

Figure 9.2 shows the dynamic relevance analysis for both cases in DB1, DLM and SSA. For DLM, the stochastic features chosen are those with a relevance weight greater than 0.4. In this case, the algorithm is selective enough and indicates that for SSA only one dynamic feature has the properties for the classification task.

9.2.3 Results of Time–Varying Decompositions

Tuning of the different schemes of considered signal decomposition throughout this training procedure is carried out by using the average classification accuracy for the automatic epilepsy seizure detection. Because of the high computational cost of stochastic feature-based training, dimension reduction of the input space is carried out by means of a time–evolving version of the standard PCA, as in [50], then, the different approaches are tested and compared, using a \( k-nn \) classifier. Table 9.1 summarizes the performed classification accuracy for the different approaches and its respective set of stochastic features. As seen, the dynamic relevance analysis is more accurate than the static relevance analysis. Nevertheless the results remain low at least for the five–class problem. For the two–class problem in DB1, the features selection through RFE shows relevant results, near to the proposed technique based on dynamic relevance. The main problem with RFE is the definition \( a-priori \) of the desired number of features, besides to the fact that 300 time–instant–features do not give any information about the original input data.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Acc [%]</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DB1 Five–Class problem</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>32.86 ± 4.47%</td>
<td>15</td>
</tr>
<tr>
<td>SSA - Static Relevance</td>
<td>31.20 ± 3.05%</td>
<td>14</td>
</tr>
<tr>
<td>SSA - Dynamic Relevance</td>
<td>40.86 ± 5.24%</td>
<td>1</td>
</tr>
<tr>
<td>SSA - RFE</td>
<td>52.46 ± 4.15%</td>
<td>15</td>
</tr>
<tr>
<td>SSA - mRMR</td>
<td>30.86 ± 3.11%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>SSA - CLeVer</td>
<td>28.53 ± 4.05%</td>
<td>1</td>
</tr>
<tr>
<td>DLM</td>
<td>78.00 ± 2.30%</td>
<td>20</td>
</tr>
<tr>
<td>DLM - RFE</td>
<td>25.73 ± 4.02%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - mRMR</td>
<td>43.13 ± 2.59%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - CLeVer</td>
<td>76.26 ± 2.31%</td>
<td>8</td>
</tr>
<tr>
<td>DLM - Static Relevance</td>
<td>47.13 ± 3.70%</td>
<td>4</td>
</tr>
<tr>
<td>DLM - Dynamic Relevance</td>
<td>72.93 ± 2.41%</td>
<td>7</td>
</tr>
<tr>
<td>EDS</td>
<td>64.46 ± 4.25%</td>
<td>5</td>
</tr>
<tr>
<td><strong>DB1 Two–Class problem</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>66.33 ± 2.33%</td>
<td>15</td>
</tr>
<tr>
<td>SSA - RFE</td>
<td>72.00 ± 4.15%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>SSA - mRMR</td>
<td>82.66 ± 2.78%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>SSA - CLeVer</td>
<td>70.00 ± 2.63%</td>
<td>3</td>
</tr>
<tr>
<td>SSA - Static Relevance</td>
<td>63.83 ± 5.98%</td>
<td>6</td>
</tr>
<tr>
<td>SSA - Dynamic Relevance</td>
<td>68.50 ± 5.90%</td>
<td>3</td>
</tr>
<tr>
<td>DLM</td>
<td>97.83 ± 1.58%</td>
<td>20</td>
</tr>
<tr>
<td>DLM - RFE</td>
<td>90.66 ± 4.01%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - mRMR</td>
<td>74.83 ± 6.40%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - CLeVer</td>
<td>93.83 ± 4.16%</td>
<td>5</td>
</tr>
<tr>
<td>DLM - Static Relevance</td>
<td>96.50 ± 2.28%</td>
<td>8</td>
</tr>
<tr>
<td>DLM - Dynamic Relevance</td>
<td>97.83 ± 1.12%</td>
<td>4</td>
</tr>
<tr>
<td>EDS</td>
<td>72.00 ± 5.37%</td>
<td>5</td>
</tr>
<tr>
<td><strong>DB2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA</td>
<td>72.20 ± 3.16%</td>
<td>15</td>
</tr>
<tr>
<td>SSA - RFE</td>
<td>71.12 ± 3.33%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>SSA - mRMR</td>
<td>77.00 ± 1.23%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>SSA - CLeVer</td>
<td>68.12 ± 2.01%</td>
<td>3</td>
</tr>
<tr>
<td>SSA - Static Relevance</td>
<td>69.50 ± 2.09%</td>
<td>2</td>
</tr>
<tr>
<td>SSA - Dynamic Relevance</td>
<td>74.50 ± 2.88%</td>
<td>1</td>
</tr>
<tr>
<td>DLM</td>
<td>89.04 ± 1.84%</td>
<td>20</td>
</tr>
<tr>
<td>DLM - RFE</td>
<td>73.54 ± 2.86%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - mRMR</td>
<td>71.19 ± 3.81%</td>
<td>300 non–stochastic</td>
</tr>
<tr>
<td>DLM - CLeVer</td>
<td>74.70 ± 2.90%</td>
<td>3</td>
</tr>
<tr>
<td>DLM - Static Relevance</td>
<td>83.83 ± 2.67%</td>
<td>4</td>
</tr>
<tr>
<td>DLM - Dynamic Relevance</td>
<td>84.04 ± 2.14%</td>
<td>6</td>
</tr>
<tr>
<td>EDS</td>
<td>79.62 ± 1.89%</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 9.1: Performance outcomes for time series analysis using a single tuned \(k\)-nn classifier.
9.3 Time–Frequency Decompositions

According to [67] and [69], and for the sake of simplicity, the time-frequency analysis has been carried out by a quadratic TF representation, such as the spectrogram. This representation, based on a classical Fourier Transform, introduces a time localization concept by using a tapering window function of short duration going along the signal, and adding a time dimension. Particularly, the analysis is performed within a range of 0 to 83 Hz. Therefore, the spectrogram is computed by using, as tapering function, Gaussian windows with lengths of 2.9 s and 0.05. The scalogram (WT) is computed with a Complex Morlet, with scales between 1 and 256. Figure ?? shows both representations.

9.3.1 Continuous Decomposition Enhancement by Spectral Splitting

It must be noted that for this database only the maximum variance approach is applied, due to its performance in the others outcomes presented in this work. Figure ?? shows the behavior of the relevance for each point in the TFR and WT maps, and Table 9.2 shows the frequency bands that are chosen. Each band includes 1 time series of vector cepstral coefficients that is computed by 1 triangular response filter.

Table 9.2: Frequency bands selection for time–frequency decompositions enhancement

<table>
<thead>
<tr>
<th>Relevance</th>
<th>Frequency Bands</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral splitting</td>
<td>TFR</td>
<td>0 – 10, 10 – 30, 30 – 40, 40 – 50, 50 – 83 [Hz]</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>1 – 100, 100 – 200, 200 – 256 [Scale]</td>
</tr>
<tr>
<td></td>
<td>DB2</td>
<td></td>
</tr>
<tr>
<td>Spectral splitting</td>
<td>TFR</td>
<td>0 – 13, 13 – 25, 25 – 43, 43.01 – 56.2, 56.2 – 60.1, 60.1 – 73.9, 73.9 – 83 [Hz]</td>
</tr>
<tr>
<td></td>
<td>WT</td>
<td>1 – 65.25, 65.25 – 127.5, 127.5 – 256 [Scale]</td>
</tr>
</tbody>
</table>

According to the sub–band selection by spectral splitting, both representations of the EEG are decomposed using cepstral coefficients. In this pro-
CHAPTER 9. FEATURES SELECTION FOR EPILEPSY DETECTION

(a) Spectrogram DB1

(b) Scalogram DB1

(c) Spectrogram DB2
CHAPTER 9. FEATURES SELECTION FOR EPILEPSY DETECTION

Figure 9.4: Selection of the frequency bands using relevance analysis for EEG recordings.

(a) Relevance map for spectrogram DB1
(b) Relevance map for scalogram DB1
(c) Relevance map for spectrogram DB2
(d) Relevance map for scalogram DB2
cess, for each time a triangular response filter is applied, obtaining for DB1, 5 stochastic features for spectrogram and 3 stochastic features for scalogram; and for DB2, 7 stochastic features for spectrogram and 3 stochastic features for scalogram.

### 9.3.2 Results of Time–Frequency Decompositions

Tuning of the different schemes of considered signal decomposition throughout this training procedure is carried out by using the average classification accuracy for the automatic epilepsy detection. Because of high computational cost of stochastic feature-based training, dimension reduction of the input space is carried out by means of a time–evolving version of the standard PCA, as in [50], then, the different approaches are tested and compared, using a $k$–nn classifier.

Table 9.3 summarizes the performed classification accuracy for the different approaches and its respective set of stochastic features. As seen, there is no statistical difference in terms of classification performed by each decomposition techniques (TFR or WT) and its respective reduced set in the case of five–class problem for DB1. Nevertheless, in the two–class problem

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>Approach</th>
<th># Features</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DB1 Five–Class Problem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFR</td>
<td>Baseline</td>
<td>512</td>
<td>78.93 ± 2.72</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>5</td>
<td>78.66 ± 2.35</td>
</tr>
<tr>
<td>WT</td>
<td>Baseline</td>
<td>256</td>
<td>67.26 ± 4.13</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>3</td>
<td>72.66 ± 4.43</td>
</tr>
<tr>
<td><strong>DB1 Two–Class Problem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFR</td>
<td>Baseline</td>
<td>512</td>
<td>96.33 ± 2.45</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>5</td>
<td>99.00 ± 1.00</td>
</tr>
<tr>
<td>WT</td>
<td>Baseline</td>
<td>256</td>
<td>83.88 ± 3.85</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>3</td>
<td>96.16 ± 3.33</td>
</tr>
<tr>
<td><strong>DB2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFR</td>
<td>Baseline</td>
<td>512</td>
<td>82.00 ± 3.32</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>7</td>
<td>94.04 ± 1.02</td>
</tr>
<tr>
<td>WT</td>
<td>Baseline</td>
<td>256</td>
<td>67.04 ± 2.66</td>
</tr>
<tr>
<td></td>
<td>Spectral splitting</td>
<td>3</td>
<td>65.12 ± 2.90</td>
</tr>
</tbody>
</table>
for DB1 and for DB2, a significant improvement is achievement by spectral splitting.

9.3.3 EEG–Rhythms Analysis

Rhythms extraction: in order to obtain clinically interpretable medical results, $\delta, \theta, \alpha$ and $\beta$ frequency band activities of EEG signals have to be investigated. The meaning of each one of these waveforms is explained in [71] as follows:

- $\alpha$: is the most important waveform of the EEG and comprises frequencies between $8 - 12$ Hz. This wave appears when the eyes are closed and the patient begins to rest.

- $\delta, \theta$: these waves should be seen when the epileptic seizures occurs and have higher amplitudes and lower frequencies respect to $\alpha$ waves ($0 - 4$ Hz and $4 - 8$ Hz, respectively).

- $\beta$: in addition, brain produces asynchronous waves, which have higher frequency ($13 - 30$ Hz) and lower magnitude, called $\beta$ waves.

Regarding to rhythms extraction from the EEG recordings, a four-level multi-resolution decomposition using Daubechies4 wavelets is implemented as recommended in [68]. Figure 9.3 shows the rhythms spectrum for data-set Z (normal activity) and data-set S (ictal activity). As can be seen, the spectra of the rhythms has some frequency–lag, due to the wavelet transform has not a linear frequency scale and in the coefficients remains some non–desirable components.

For EEG–rhythms analysis, the approach with the best performance, i.e. TFR decomposition, is implemented. The parameters of the TFR are the same than for the EEG signal, but treating each rhythm as an individual signal. The dynamic features obtained are mixed in a $k$–nn–classifier array, aiming to preserve the different dynamics.

Table 9.4 shows the results for different rhythms combination. Despite that the information of the rhythms covers almost all the spectrum, the information could be redundancy. The advantage of the TFR spectral splitting is that the information is summarized in the stochastic features with less probability of redundancy between them.
Figure 9.5: Example of rhythms spectrum for different data-sets.
Table 9.4: Classification accuracy for rhythms analysis

<table>
<thead>
<tr>
<th>Rhythms</th>
<th>Acc [%]</th>
<th>Rhythms</th>
<th>Acc [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DB1 Five–Class Problem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>70.66 ± 4.28</td>
<td>$\theta$</td>
<td>75.40 ± 3.61</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>68.33 ± 3.06</td>
<td>$\beta$</td>
<td>74.40 ± 2.08</td>
</tr>
<tr>
<td>$\delta, \theta$</td>
<td>84.33 ± 3.19</td>
<td>$\theta, \beta$</td>
<td>86.93 ± 2.49</td>
</tr>
<tr>
<td>$\delta, \theta, \beta$</td>
<td>88.46 ± 2.91</td>
<td>$\delta, \theta, \beta, \alpha$</td>
<td>89.33 ± 2.77</td>
</tr>
<tr>
<td><strong>DB1 Two–Class Problem</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>87.83 ± 3.77</td>
<td>$\theta$</td>
<td>89.33 ± 2.62</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>84.33 ± 5.22</td>
<td>$\beta$</td>
<td>93.33 ± 1.92</td>
</tr>
<tr>
<td>$\delta, \theta$</td>
<td>94.00 ± 3.35</td>
<td>$\theta, \beta$</td>
<td>96.33 ± 1.72</td>
</tr>
<tr>
<td>$\delta, \theta, \beta$</td>
<td>98.88 ± 1.10</td>
<td>$\delta, \theta, \beta, \alpha$</td>
<td>99.00 ± 0.88</td>
</tr>
<tr>
<td><strong>DB2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>72.75 ± 2.03</td>
<td>$\theta$</td>
<td>70.04 ± 3.12</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>81.70 ± 2.47</td>
<td>$\beta$</td>
<td>84.54 ± 2.60</td>
</tr>
<tr>
<td>$\delta, \theta$</td>
<td>77.70 ± 2.92</td>
<td>$\theta, \beta$</td>
<td>86.83 ± 3.24</td>
</tr>
<tr>
<td>$\delta, \theta, \beta$</td>
<td>87.87 ± 1.80</td>
<td>$\delta, \theta, \beta, \alpha$</td>
<td>90 ± 3.88</td>
</tr>
</tbody>
</table>

### 9.4 Discussion

This chapter proposes a comparison between different EEG signal decompositions and its respective enhancements through spectral–splitting for epilepsy events detection. Additionally, a methodology to evaluate the rhythms or brainwaves contribution to neuronal activity related to either normal or epileptic seizure events is proved. The methodology lies in the hypothesis that by providing an appropriate spectral sub–band division, it is possible to detect ictal events when the rhythms information is not available.

- Regarding to the discrete decompositions, it is clear that the overlapping between classes is the major impediment for a good classification task. Nevertheless, the fact that the static relevance and the dynamic relevance analysis for dimension reduction in the case of DLM present so different results, gives an indication that the weights of the decompositions have strong changes across the time axis. It is important to remark the RFE selection technique classification accuracy. The initial number of features parameter is not easy to fix, due to the fact that with a large–size features, is difficult to determine how many of this features are needed to obtain a good classification. In addition, if the number of features is suitable, the interpretation of time–instant
The overlapping of the classes could be explained because this problem usually is addressed like a three-classes problem [67]. The EEG segments are sorted into three different classes. Z and O types of EEG segments are combined into a single class; N and F types are also combined into a single class; and type S is the third class. This scenario with only three categories is close to the real medical applications.

Regarding to the EEG signal decompositions, the representation based on TF provides a better accuracy in comparison with the representation based on WT. It could be explained due to the fact that The Wavelet Transform 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 [72]. Nevertheless, for EEG recordings both kinds of frequencies are present at all the time interval, then, some of the spectral components are lost.

The spectral splitting over the TF and WT maps improves the results, comparing with the baseline (continuous decomposition as a whole set of stochastic features). As can be seen in Figure ??, the dynamic of the EEG signals is distributed along the time axis and can be expressed in a briefly way by the cepstral coefficients.

Present study discusses the introduction of spectral splitting over the EEG decomposition to optimize the epileptic seizures detection, in terms of improving the classification accuracy with the lowest possible complexity (fewest stochastic feature set but not increasing the computational effort) [73]. Both considered approaches of EEG decomposition achieve an small feature number (7 cepstral coefficients as maximum). In addition, if the rhythms are unknown, the discussed training methodology provides additional benefit of easier clinical interpretation of EEG–derived parameters through by the relevance map, principally in the two–class problem.

According to Table[9.4] in almost all cases, the rhythm $\beta$, i.e the rhythm with the higher frequency, has a significant impact in the classification
accuracy. This behavior is due to the fact that this brainwave represents the non–synchronous waves that occur when an epileptic seizure has place [71].

Concerning to the comparison between the best approach within and without the knowledge of the EEG rhythms, the classification accuracy is improved by the spectral splitting in the case of the two–class problem (99 ± 0.8 % vs 99 ± 1% for DB1 and 94 ± 1 % vs 90 ± 4% for DB2) is achieved. Nevertheless, in the five–class problem, the knowledge of the rhythms improves the accuracy (78.66 ± 2.35% vs 89.33 ± 2.7% ), due to the fact, that the relevance map is not clear and the sub–band selection could not be suitable for so complex problem.
Chapter 10

Respiratory Frequency Extraction in Treadmill Exercise Test

HRV can provide information on the relative inputs of the two autonomic nervous system mutually antagonistic components: the sympathetic one, which acts to increase heart rate, and the parasympathetic component, which acts to slow the heart and to dilate blood vessels [74]. The former component activity is measured by the power in the termed Low Frequency (LF) spectral band \((f \in [0.04, 0.15] \text{ Hz})\), while the latter activity is measured by high frequency (HF) band \((f \in [0.15, 0.40] \text{ Hz})\) [73]. The fact that respiratory frequency is not restricted to the classical HF band during exercise stress testing, makes it necessary to redefine the HF band [75]. During exercise, HRV signals are highly non-stationary and may be influenced by false detections of QRS complexes due to muscular noise and motion artifacts present in the ECG signal. Moreover, the respiratory frequency during exercise is in itself a highly dynamic quantity and changes with effort and work load [76]. In this chapter, several strategies are implemented, aiming to find the respiratory frequency and follows its dynamic along the time.

10.1 Preprocessing

Heart Rate Variability (HRV): Aiming to extract the HRV signal from the R-R sequence during exercise stress testing, the model TVIPFM is used
as recommended in [77] as follows:

\[ x_{HRV}(t) = x_{HR}(t) - x_{HRM}(t) = \frac{m(t)}{T(t)}, \]  

(10.1)

where \( x_{HRV}(t) \) is the HRV signal, \( x_{HR}(t) = \frac{1+m(t)}{T(t)} \) is the instantaneous heart rate, \( x_{HRM} = \frac{1}{T(t)} \) is a time-varying mean heart rate, \( T(t) \) is the time-varying mean heart period, and \( m(t) \) is the representation of the influence of the autonomic nervous system over the sino-atrial (SA) node activity. In this model, \( m(t) \) is assumed a modulating signal, causal and band-limited.

**Segmentation of the HRV Signal:** Due to the fact that the analysis is made over the stress test, a part of the signal corresponding to the initial state and the after exercise state have to be removed.

### 10.2 Signal Decomposition

#### 10.2.1 Time–Frequency Decomposition

The first approach for the addressed problem, consists in finding the maximum frequency at each time for the HRV-TFR, assuming that the most relevant band for the respiratory frequency is between 0.4 and 0.8 Hz. Figure 10.1 shows two examples where the maximum frequency in the selected band is found. Despite that, the algorithm is able to find an approximation of respiratory frequency is present, the main problem is present when the respiratory frequency is up of 0.8 Hz as in the Register 1.

**Instantaneous Frequency**

Aiming to improve the respiratory frequency detection, the instantaneous frequency explained in [78] with some modifications for HRV signals is implemented. The algorithm try to find peaks in the spectrum of the signal. For the algorithm considerations, a region is selected when a certain percentage of the spectral power is contained in an interval centered around the largest peak, otherwise the spectrum is omitted from averaging. According to [78], estimation of the respiratory frequency as the largest peak of comes with the risk of choosing the location of a spurious peak. This risk is, however,
considerably reduced by narrowing down the search interval to only include frequencies in an interval of centered around a reference frequency $f_\omega$.

For this work, the search interval was considered as the band between 0.3 and 0.9 Hz.

### 10.2.2 SSA

Regarding to extract the maximum information about the frequency respiratory behavior, a cascade decomposition is carried out, combining both discrete and continuous decompositions.

In the first step of this process, the signal is decomposed through SSA. The window length is fixed in a half of the signal length. The number of eigen-triples is fixed in 35, assuming that the most relevant information is already contained in this set. Figure 10.2 shows the original signal and the reconstructed one.

The second step consists in selecting the non-correlated eigen-triples for each recording. The eigen-triples are ordered according to the singular values, then, when high correlated components are found, the first eigen-triple is chosen, due to the fact that it is the most relevant. Additionally, the low frequencies are eliminated and the component related to the hardware used in the stress test (around 1.1 Hz $f_{\omega}$), aiming to find the respiratory frequency in the HF band of the HRV spectrum. Figure 10.3 shows the matrix of the coefficients of correlation between the 35 eigen-triples. Figure 10.4 shows the time frequency representation based on the STFT, of the reconstructed
Figure 10.2: Examples of original and reconstructed signals - HRV.
CHAPTER 10. RESPIRATORY FREQUENCY EXTRACTION IN TREADMILL EXERCISE

signal and the filtered one. As can be seen, the filtered signal does not have
the low frequency components.

![Figure 10.3: Examples of correlation matrix of the eigen triples.](image)

Figure 10.3: Examples of correlation matrix of the eigen triples.

To ensure that the filtered signal does not have low-frequency compo-
nents, the instantaneous frequency is computed from the TFR. Figure 10.5
shows the instantaneous frequency for each one of the selected eigen–triples.

Finally, the results of the last step are compared with the respiratory
frequency given by the jeager. Figure 10.6 shows the comparison with the
original data and the resulting one.

Table 10.1 shows the relative error between the real respiratory frequency
and the instantaneous frequency found with the SSA algorithm for 10 cases.
It can be seen that the algorithm is able to find at least a band where the res-
piratory frequency is present. Nevertheless, it is possible to find an approach
more appropriate in frequency searching. Under this assumption the empirical
mode decomposition is explored, aiming to improve the band frequency
selection.

10.2.3 Empirical Mode Decomposition (EMD)

The EMD is a procedure decomposing the signal \( x(t) \) into a finite set of
frequency and amplitude modulated signals, being the representation basis
called Intrinsic Mode Functions IMF. Each IMF is computed by an itera-
tive process [10]: i) The extrema of the signal is detected (maximum and
minimum), ii) The upper and lower envelopes are computed, connecting the
maximum and minimum independently by cubic spline interpolation, iii) The
local mean of the envelopes is calculated, iv) If the local mean is zero, it is
CHAPTER 10. RESPIRATORY FREQUENCY EXTRACTION IN TREADMILL EXERCISE

Figure 10.4: Examples of reconstructed and filtered signals - HRV.

Figure 10.5: Examples of eigen–triples instantaneous frequency.
Figure 10.6: Comparison between the processed and the original signals - HRV.

<table>
<thead>
<tr>
<th>Recording</th>
<th>Relative Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.75</td>
</tr>
<tr>
<td>2</td>
<td>179.59</td>
</tr>
<tr>
<td>3</td>
<td>94.00</td>
</tr>
<tr>
<td>4</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>23.07</td>
</tr>
<tr>
<td>8</td>
<td>7.26</td>
</tr>
<tr>
<td>16</td>
<td>0.29</td>
</tr>
<tr>
<td>21</td>
<td>17.84</td>
</tr>
</tbody>
</table>

Table 10.1: Relative Error
CHAPTER 10. RESPIRATORY FREQUENCY EXTRACTION IN TREADMILL EXERCISE

subtracted of the original signal; the result is the first IMF, v) If the following conditions are not met, a sifting process is carried out. If the conditions are met, this IMF is considered, then, the remainder is the new signal and the process begins again.

Conditions for IMF After the signals are decomposed by EMD, the instantaneous frequency is found and the relative error between the respiratory frequency given in the database and the instantaneous frequency is computed. Figure 10.7 shows both frequencies, the red line represents the real respiratory frequency and the cyan line shows the frequency found.

![Figure 10.7: Results for EMD and instantaneous frequency](a) Recording 8 (b) Recording 16

10.3 Results

Table 10.2 summarizes the results obtained for TFR and EMD plus TFR. The effect of include or no include more IMF is showed too.

10.4 Discussion

This chapter proposes a comparison between different HRV signal decompositions aiming to find the best approximation to the respiratory frequency. The methodology lies in the hypothesis that providing an appropriate spectral decomposition, it is possible to detect the respiratory frequency when the
real data is not available. The decompositions based on TFR and EMD follow by TFR present the best results regarding to the relative error between the frequency band founded and the real respiratory frequency.

- The cascade filtering process is needed in this type of biosignals. In this case, the components with highest power are found in the low frequency band. When these components are added at the analysis, the components of high frequency are hidden in almost all the registers, due to the LF components; and it is usual that the only component visible in the HF band is the oscillatory component in the range of 1.1 Hz identified as a signal derived of the test machine.

- As can be seen in Figure 10.3 some correlations between signals are negative, i.e. when the analyzed component has an increasing behavior, the negative correlated one has a decreasing behavior. If these signals are mixed, the result is the cancellation of both components and the spectrogram is not able to recover the information lost. With this motivation, the negative correlations are removed in the second step of the methodology.

- After the filtering process, one component seems to appear in the HF band. This component could be related to the running cadence coupling. This frequency is usually higher than the respiratory frequency and has an increasing behavior due to the protocol of the data acquisition explained in Chapter 5.

- Regarding to the SSA process, although the algorithm is able to find a band containing the respiratory frequency, in almost all cases is not able to track the dynamic changes of this component along the time, due to

<table>
<thead>
<tr>
<th>Decomposition</th>
<th>[%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFR</td>
<td>5.0699 ± 3.8869</td>
</tr>
<tr>
<td>IMF 1</td>
<td>9.8830 ± 11.9459</td>
</tr>
<tr>
<td>IMF 2</td>
<td>4.8676 ± 3.6619</td>
</tr>
<tr>
<td>IMF 3</td>
<td>12.8063 ± 5.8200</td>
</tr>
<tr>
<td>IMF (1 + 2)</td>
<td>3.8423 ± 3.5187</td>
</tr>
</tbody>
</table>

Table 10.2: Relative error for respiratory frequency detection
the non–linearity of the process. Another approaches are recommended for signal tracking as the non–linear version of the Kalman filter.
Part IV

Conclusions and Future Work
Chapter 11

Conclusions

This work presented a methodology for stochastic features subset selection based on a time-varying relevance function, aiming to improve the classification accuracy and the representation of biosignals and to reduce the dimensionality of the input matrix. Different decomposition techniques were implemented, aiming to find a more suitable set of stochastic features representing the dynamics of real-data biosignal recordings. Those decompositions were formulated into a generalized structure, i.e., the signals were decomposed through a linear combination of time-varying coefficients, weighted with the proper basis of the representation. According to the number of used coefficients, a high-dimensional representation space was obtained.

With this idea in mind, a stochastic features selection algorithm was proposed. A selection criteria were established for three cases: firstly the static relevance, i.e., when the relevance weight of each feature is represented like the average of the relevance weight along the time.

Secondly the dynamic relevance, i.e., when the relevance weight of each feature is represented by its variance, under the hypothesis that a less varying weight could represent in a better way the studied phenomena. Additionally, two different combining approaches for selecting the best set of stochastic features are studied: i) when taking a partially divided set comprising a single type of stochastic features, generated by the same principle, ii) when the best dynamic features are chosen among the whole set of features, no matter on their physical meaning. Achieved performance, for both cases, leads to the conclusion that latter approach for selecting dynamic features provides a higher accuracy than the former approach, although the physical meaning of the selected feature set.
Finally, a last approach consisting in the dimension reduction of highly correlated feature spaces is proposed, i.e. time–frequency and time–scale representation, through relevance maps with information along time and frequency/scale axes.

As a result, 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 the evaluation measure of time–variant transformation, and therefore, distinguishing relevant stochastic features.

**Future Work:** Following the research line described in this thesis, two main projects could be taken up, which involve the proposed improvements for the dynamic feature selection algorithm:

1. Further efforts on testing other relevance measures based on non–linear analysis, should be focused on extended studies to corroborate the potential of another relevance approaches along with biosignals analysis and pathology diagnosing.

2. The influence of different classification approaches (neural networks, hidden markov models, etc.) should be further studied to improve the performance of the proposed methodology.
Part V
Appendix
Appendix A

List of Publications


APPENDIX A. LIST OF PUBLICATIONS

by partial least square-based extraction of dynamic features,” in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 31 2010-sept. 4 2010, pp. 6321–6324


Bibliography


BIBLIOGRAPHY


