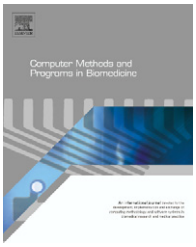
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 108 (2012) 250–261

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Unsupervised feature relevance analysis applied to improve ECG heartbeat clustering

J.L. Rodríguez-Sotelo[a], *D. Peluffo-Ordoñez*[b], *D. Cuesta-Frau*[c,*], *G. Castellanos-Domínguez*[b]

[a] *Universidad Autónoma de Manizales, Grupo Automática, D. Ing. Electrónica y Automatización, Antigua estación de ferrocarril, Manizales, Colombia*

[b] *Universidad Nacional de Colombia, D. Ing. Eléctrica, Electrónica y Computación, Km. 9, Vía al aeropuerto, Campus la Nubia, Manizales, Colombia*

[c] *Technological Institute of Informatics, Polytechnic University of Valencia, Campus Alcoi, Plaza Ferrándiz y Carbonell, 2, 03801 Alcoi, Spain*

ARTICLE INFO

Article history:
Received 30 August 2011
Received in revised form
3 April 2012
Accepted 17 April 2012

Keywords:
Electrocardiogram analysis
Feature selection
Relevance analysis
Heartbeat classification
Q – α algorithm

ABSTRACT

The computer-assisted analysis of biomedical records has become an essential tool in clinical settings. However, current devices provide a growing amount of data that often exceeds the processing capacity of normal computers. As this amount of information rises, new demands for more efficient data extracting methods appear.

This paper addresses the task of data mining in physiological records using a feature selection scheme. An unsupervised method based on relevance analysis is described. This scheme uses a least-squares optimization of the input feature matrix in a single iteration. The output of the algorithm is a feature weighting vector.

The performance of the method was assessed using a heartbeat clustering test on real ECG records. The quantitative cluster validity measures yielded a correctly classified heartbeat rate of 98.69% (specificity), 85.88% (sensitivity) and 95.04% (general clustering performance), which is even higher than the performance achieved by other similar ECG clustering studies. The number of features was reduced on average from 100 to 18, and the temporal cost was a 43% lower than in previous ECG clustering schemes.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The computer-assisted analysis of biomedical records has become an essential tool in clinical settings. The widespread access to portable medical devices or new personal devices such as cell phones, smartphones, pdas, tablets, and wearable devices, is boosting the amount of biomedical data available. These devices provide a growing amount of data that often exceeds the processing capacity of affordable computers. As this amount of biosignal data rises, new demands for more efficient information extracting methods appear [1].

A number of algorithms have been proposed for knowledge discovery and management in medical databases [2]. These methods are aimed at turning the huge amount of information in these databases into a more manageable source. This objective can be achieved by removing useless data such as noise or outliers [3], selecting only a data subset [4], performing a data mining process to capture data patterns or relationships [5], or by obtaining an efficient lower dimension representation of the data [6].

Feature selection algorithms are dimensionality reduction methods often associated to data mining tasks of classificationor clustering [1]. These methods provide a reduced set

* *Corresponding author*. Tel.: +34 651028565.
E-mail addresses: jlrodriguez@autonoma.edu.co (J.L. Rodríguez-Sotelo), dcuesta@disca.upv.es (D. Cuesta-Frau), cgcastellanosd@unal.edu.co (G. Castellanos-Domínguez).

0169-2607/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2012.04.007

of the input features while preserving the relevant discriminatory information. Feature matrix projection methods and its derivatives, specially principal component analysis (PCA) based methods, are probably the most popular of such relevance analysis schemes. PCA related methods have been applied successfully to many biomedical signals: ECG [7–9], EEG [10,11], RR series [12], respiration [13]; and in diagnosis applications related to the Alzheimer's disease [14], Parkinson's disease [15], and many more such as in [16,17], or [18], among others.

We describe in this paper a new feature matrix projection scheme for unsupervised relevance analysis. This method can be applied to any biomedical signal where a clustering process is involved. It outperforms the conventional PCA methods in terms of generalization, complexity, and feature discriminatory information. The strong asymmetry frequently found in biomedical data sometimes prevent PCA methods from being used, since they assume a uniform data class distribution [19]. In addition, PCA methods are also very sensitive to the presence of noise and outliers [20].

The algorithm proposed is based on a least-squares optimization scheme of the feature matrix that iteratively converges to a local maximum of a relevance function [21]. The output of the method is a feature significance score that enables the exclusion of superfluous features and a weighted representation of the remaining ones. We have further improved the original method by reducing the number of iterations to 1. Consequently, the computational requirements to obtain the final feature set is lower, whereas its information content remains almost unchanged.

The performance of the method presented was assessed using a heartbeat clustering test on ECG records. ECGs are one of the most used biomedical signals, and long-term records, where data mining might be a powerful tool, are quite common [4]. Labelled records from the MIT/BIH database [22], containing different types of arrhythmias, were used in the experiments.

A general feature extraction procedure was applied to the segmented heartbeats. The feature set went then through the relevance analysis stage proposed before entering the clustering process. The objective of the experiments was to assess the quality of the heartbeat partition obtained using this method of relevance analysis, both in terms of accuracy and temporal cost.

2. Method

The objective of a feature relevance analysis is to find a simpler representation of the input features that preserves most of their significant information according to a given criterion function [23]. A tradeoff must be found between the quantity and/or quality of the input features, and the separability of the objects from which they were extracted. Feature selection has many important potential benefits that makes it suitable in a number of clustering applications [24].

Our method employs a feature relevance analysis scheme to filter the input data, and preserve and score only those features that maximize the information content. Although this method can be used in any case where a clustering process of a feature vector set takes place, we focused our attention on ECG data. ECG records are one of the most common biomedical signals, and high-resolution and/or long-term ECGs are probably the best candidates for data mining approaches [4].

2.1. *Heartbeat feature selection algorithm*

The aim of the method is to compute weighting feature values that enable an effective data dimensionality reduction along with a proper feature relevance ranking. To accomplish this goal, a distance-based least-squares optimization scheme is proposed.

2.1.1. *Definitions*

Let $\mathbf{x}_j \in \mathbb{R}^p$ be a heartbeat feature vector, $1 \le j \le n$. Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ be a matrix that results from arranging all $\mathbf{x}_j$ vectors in rows. Let $\mathbf{Y} \in \mathbb{R}^{n\times p}$ be a linear projection of $\mathbf{X}$ given by $\mathbf{Y} = \mathbf{XV}$, where $\mathbf{V} \in \mathbb{R}^{p\times p}$ is an unknown orthogonal projection matrix. If the number of features is reduced from p to q, we have to introduce the truncated version of the previous matrices, namely $\hat{\mathbf{V}} \in \mathbb{R}^{p\times q}$, $\hat{\mathbf{Y}} \in \mathbb{R}^{n\times q}$, $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{V}}$, and the least squares estimated matrix $\hat{\mathbf{X}} \in \mathbb{R}^{n\times p}$, $\hat{\mathbf{X}} = \hat{\mathbf{Y}}\hat{\mathbf{V}}^\top$.

Finally, let $\boldsymbol{\alpha} \in \mathbb{R}^p$ be an unknown feature weighting vector that can be arranged in an affinity symmetric positive definite matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\top$, with $\|\boldsymbol{\alpha}\| = 1$.

2.1.2. *Mathematical background*

The dissimilarity between the original feature matrix $\mathbf{X}$ and the reconstructed version $\hat{\mathbf{X}}$ can be quantified by the euclidean norm, that is, $||\mathbf{X} - \hat{\mathbf{X}}||$. Since this dissimilarity has to be minimized in terms of $\hat{\mathbf{V}}$, and $\mathbf{A}$, the initial objective function can be expressed as:

$$\min_{\hat{\mathbf{V}}} ||\mathbf{X} - \hat{\mathbf{X}}||_{\mathbf{A}}^2 = \min_{\hat{\mathbf{V}}} ||\mathbf{X} - \hat{\mathbf{Y}}\hat{\mathbf{V}}^\top||_{\mathbf{A}}^2 \tag{1}$$

where we use a squared version for simplicity. Applying the inner product definition and trace properties to Eq. (1), it can be rewritten as:

$$\mathrm{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X}^\top - \hat{\mathbf{X}}^\top\mathbf{A}\mathbf{X}^\top - \mathbf{X}^\top\mathbf{A}\hat{\mathbf{X}}^\top + \hat{\mathbf{X}}^\top\mathbf{A}\hat{\mathbf{X}}^\top) \tag{2}$$

Another term of interest is $||\mathbf{X}||_{\mathbf{A}}^2$, given by:

$$||\mathbf{X}||_{\mathbf{A}}^2 = \mathrm{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X}) = \mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{X}^\top) = \sum_{i=1}^{p} \gamma_i\lambda_i \tag{3}$$

where λ_i and $\mathbf{v}_i$ denote respectively the i-th eigenvalue and eigenvector of $\mathbf{X}^\top\mathbf{X}$ and $\gamma_i = \lambda_i^{-1}\mathbf{v}_i^\top\mathbf{X}^\top\mathbf{A}\mathbf{X}\mathbf{v}_i$. Arranging Eq. (2) according to formulation for Eq. (3), it results in the following new expression:

$$\sum_{i=1}^{p}\gamma_i\lambda_i - 2\sum_{i=1}^{q}\gamma_i\lambda_i + \sum_{i=1}^{q}\gamma_i\lambda_i = \sum_{i=q+1}^{p}\gamma_i\lambda_i \tag{4}$$

Therefore, there is a relationship between Eqs. (3) and (4) that can be expressed as:

$$||\mathbf{X}||_{\mathbf{A}}^2 = \sum_{i=1}^{q} \gamma_i \lambda_i + ||\mathbf{X} - \hat{\mathbf{X}}||_{\mathbf{A}}^2 = \mathrm{tr}(\hat{\mathbf{X}}^\top \mathbf{A}\mathbf{X}) + ||\mathbf{X} - \hat{\mathbf{X}}||_{\mathbf{A}}^2$$

Since $||\mathbf{X}||_{\mathbf{A}}^2$ is constant, this becomes a dual optimization problem; minimizing the error function $||\mathbf{X} - \hat{\mathbf{X}}||_{\mathbf{A}}^2$ is the same as maximizing its complement $\mathrm{tr}(\hat{\mathbf{X}}^\top \mathbf{A}\mathbf{X})$. In addition, this term can be expanded in a more suitable form for optimization:

$$\mathrm{tr}(\hat{\mathbf{X}}^\top \mathbf{A}\mathbf{X}) = \mathrm{tr}(\hat{\mathbf{V}}\hat{\mathbf{Y}}^\top \mathbf{A}\mathbf{X}) = \mathrm{tr}(\hat{\mathbf{V}}\hat{\mathbf{V}}^\top \mathbf{X}^\top \mathbf{A}\mathbf{X}) = \mathrm{tr}(\hat{\mathbf{V}}^\top \mathbf{X}^\top \mathbf{A}\mathbf{X}\hat{\mathbf{V}})$$

and, thus, the new criterion function becomes:

$$\min_{\hat{\mathbf{V}}} ||\mathbf{X} - \hat{\mathbf{X}}||_{\mathbf{A}}^2 \Rightarrow \max_{\hat{\mathbf{V}},\mathbf{A}} \mathrm{tr}(\hat{\mathbf{V}}^\top \mathbf{X}^\top \mathbf{A}\mathbf{X}\hat{\mathbf{V}}) \quad (5)$$

Since $\mathbf{A} = \mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\top$, letting for an initial assumption $\boldsymbol{\alpha} = \mathbf{1}_p$ be and introducing a new arbitrary orthonormal $n \times n$ matrix $\mathbf{Q}$, such that:

$$\mathrm{tr}(\hat{\mathbf{V}}^\top \mathbf{X}^\top \mathbf{X}\mathbf{X}^\top \mathbf{X}\hat{\mathbf{V}}) = \mathrm{tr}(\mathbf{Q}^\top \mathbf{X}\mathbf{X}^\top \mathbf{X}\mathbf{X}^\top \mathbf{Q}) = \mathrm{tr}(\mathbf{Q}^\top \mathbf{A}\mathbf{A}\mathbf{Q}) = \sum_{i=1}^{q} \lambda_i^2 \quad (6)$$

Eq. (5) results in a new expression:

$$\max_{\mathbf{Q},\mathbf{A}} \mathrm{tr}(\mathbf{Q}^\top \mathbf{A}\mathbf{A}\mathbf{Q}) \quad (7)$$

which coincides with that in [21]. We customized the method described in such work, termed the $Q-\alpha$ algorithm after the two unknowns of the expression, to solve efficiently the optimization problem. Eq. (6) shows that the maximization is closely related to the combined energy of the first q leading eigenvalues of $\mathbf{A}$. The calculation for all possible feature subsets is computationally intractable. Instead, a suboptimal method was proposed in [21].

Since neither $\boldsymbol{\alpha}$ nor $\mathbf{Q}$ are known, an iterative method was employed to solve the optimization problem stated in Eq. (7). It starts with an initial guess for $\mathbf{Q}$, and iteratively computes $\boldsymbol{\alpha}$ and the largest q eigenvectors until the method converges to a local maxima or a specific criterion is met.

However, we propose a faster scheme. Computation of vector $\boldsymbol{\alpha}$ can be reduced just to one iteration with no significant decrease of accuracy. Maximizing $\mathrm{tr}(\mathbf{Q}^\top \mathbf{A}\mathbf{A}\mathbf{Q})$ is equivalent to maximize:

$$\mathrm{tr}(\mathbf{A}\mathbf{A}) = \mathrm{tr}(\mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\top \mathbf{X}\mathrm{diag}(\boldsymbol{\alpha})\mathbf{X}^\top)) \quad (8)$$

Since Eq. (8) is bilinear regarding $\boldsymbol{\alpha}$, the objective function can be rewritten as $\boldsymbol{\alpha}^\top \mathbf{H}\boldsymbol{\alpha}$, where $\mathbf{H}_{ij} = \mathrm{tr}(\mathbf{x}_i^\top \mathbf{x}_i \mathbf{x}_j^\top \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j^\top \mathrm{tr}(\mathbf{x}_i^\top \mathbf{x}_j) = (\mathbf{x}_i \mathbf{x}_j^\top)^2$. It can be inferred that an approximate relevance vector $\hat{\boldsymbol{\alpha}}$ corresponds to the eigenvector of the largest eigenvalue of $(\mathbf{X}^\top \mathbf{X}) \cdot^2$ (where notation $(\chi) \cdot^2$ stands for the square of each one of the elements of matrix χ).

Convergence of the original algorithm is discussed in detail in [21], where authors suggest a number of iterations between 5 and 10. Nevertheless, an indicator of the algorithm convergence could be a negligible change of $\boldsymbol{\alpha}$ at two consecutive iterations, $||\boldsymbol{\alpha}^{(r)} - \boldsymbol{\alpha}^{(r-1)}|| < \delta$, where $\delta \geq 0$, and r is the iteration index, or an accumulated variance criterion [28]. In our approach, we used this last criterion to choose q as the value for which the $\hat{\boldsymbol{\alpha}}$ accumulated variance was 98%.

The method proposed not only discards features but also scores the remaining ones. It is necessary to define a new weighting matrix as $\mathbf{W} = \mathrm{diag}(\sqrt{\hat{\boldsymbol{\alpha}}})$. Moreover, truncated matrices do not account for weights and new ones have to be defined as $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}$ and $\mathbf{Y} = \widetilde{\mathbf{X}}\widetilde{\mathbf{V}}$, where $\widetilde{\mathbf{V}}$ are the principal components of $\widetilde{\mathbf{X}}$. Matrix $\mathbf{Y}$ corresponds to the resulting set of features. The steps required by this method are summarized in Algorithm 1.

Algorithm 1 *(Heartbeat feature selection algorithm proposed based on a relevance analysis).*

Input: Heartbeat feature $n \times p$ matrix $\mathbf{X}$ (n heartbeats, p features)
Compute $(\mathbf{X}^\top \mathbf{X}) \cdot^2$
$\hat{\boldsymbol{\alpha}} \leftarrow$ eigenvector corresponding to the largest eigenvalue of $(\mathbf{X}^\top \mathbf{X}) \cdot^2$
$\mathbf{W} \leftarrow \mathrm{diag}(\hat{\boldsymbol{\alpha}})$
$\widetilde{\mathbf{X}} \leftarrow \mathbf{X}\mathbf{W}$ {Weight original data}
$\widetilde{\mathbf{V}} \leftarrow$ principal components of $\widetilde{\mathbf{X}}$
$\mathbf{Y} \leftarrow \widetilde{\mathbf{X}}\widetilde{\mathbf{V}}$ {Project data}
Output: New set of features $\mathbf{Y}$

3. Experimental set-up

The method proposed, based on the unsupervised feature selection scheme described in [21], is aimed at improving both the temporal cost and accuracy of heartbeat clustering methods. Therefore, the experiments were devised to assess the influence of such feature selection stage on the partition quality obtained, using a variety of features and heartbeat types. In the following subsections we describe all the stages involved in this experimental study.

3.1. *Experimental dataset*

The experimental dataset consisted of labelled ECG records from the Physionet database [22]. The experimental records were drawn from the MIT/BIH arrhythmia set. This set contains 48 recordings of 30 min duration.

All the heatbeats were already labelled. However, the AAMI standard [29] recommends the consideration of the following heartbeat types: normal beat (labelled as *N*), Supraventricular ectopic beat (*S*), Ventricular ectopic beat (*V*), Fusion beat (*F*), and unknown beat class (*Q*). Any of these types might be present in any record. A complete description of all the records is included in Table 1, showing the equivalence between the AAMI and MIT/BIH labels in the two first rows.

A frequent difficulty found in ECG heartbeat clustering is that some records may exhibit a strong unbalanced number of heartbeats per class. This is the case for MIT/BIH database. For example, record 100 contains 2237 *N* heartbeats, but only 33 *S* and 1 *V* heartbeats. This is a challenge for any clustering algorithm [30].

Table 1 – Set of recordings of the MIT/BIH database used in the experiments.

AAMI	N					S				V		F		Q	
MIT	N	L	R	*e*	*j*	A	*a*	*J*	S	V	E	F	*f*	P	Q
Code	1	2	3	34	11	8	4	7	9	5	10	6	38	12	13
100	2237					33				1					
101	1858					3									2
102	99									4			56	2026	
103	2080					2									
104	163									2			666	1378	18
105	2524									41					5
106	1506									520					
107										59				2076	
108	1738				1	4				16		2			
109		2490								38		2			
111		2121								1					
112	2535					2									
113	1787						6								
114	1818					10		2		43		4			
115	1951														
116	2300					1				109					
117	1532					1									
118			2164			96				16					
119	1541									444					
121	1859					1				1					
122	2474														
123	1513									3					
124			1529		5	2		29		47		5			
200	1742					30				825		2			
201	1623				10	30	97	1		198		2			
202	2059					36	19			19		1			
203	2528						2			444		1			4
205	2569					3				71		11			
207		1457	85			106				105	105				
208	1585								2	992		372			2
209	2619					382				1					
210	2421						22			194	1	10			
212	922		1824												
213	2639					25	3			220		362			
214		2000								256		1			2
215	3194					2				164		1			
217	244									162			260	1540	
219	2080					7				64		1			
220	1953					94									
221	2029									396					
222	2060				212	208		1							
223	2027			16		72	1			473		14			
228	1686					3				362					
230	2253									1					
231	314		1252			1				2					
232			396		1	1381									
233	2229					7				830		11			
234	2698							50		3					
Total	74,989	8068	7250	16	229	2542	150	83	2	7127	106	802	982	7020	33

The first row corresponds to the labels used according to the AAMI standard, and the second row lists the labels used in the MIT/BIH database. The third row corresponds to the numerical code of these last labels. The first column is the name of the records, whereas the others contain the number of heartbeats of each type.

3.2. *Segmentation*

The ECG records underwent a double segmentation process. In one hand, the records were split into consecutive non-overlapping epochs, and, on the other hand, all the heartbeats in each epoch were segmented.

Biomedical signals are usually non-stationary [31]. To better account for ECG attribute changes, we experimented with a variable number of epochs for each record: from 1 to 10. Each epoch was processed independently, and the partial results were combined to obtain the final record result.

Regarding heartbeat segmentation, since all the heartbeats in the MIT/BIH database are labelled, not only the heartbeat type, but also the QRS-complex location, it was possible to accurately extract the heartbeats included in each ECG epoch, and proceed with the feature extraction stage for each one.

3.3. *Feature extraction*

An important issue in signal clustering is how to represent the time sequences to partition. This representation greatly influences the performance of the subsequent methods. In practice, these sequences are usually chosen by researchers based on previous similar works [9,29,32–34] or knowledge.

In this way, a preselected subset of 100 features was extracted from the set of all possible quantifiable heartbeat attributes. Since the experiments are focused on cardiac arrhythmias, the specific features were chosen according to their performance in analogous studies [30,33–35]. This set was intentionally overdimensioned in order to enable the relevance analysis method to process most of the features described in the scientific literature. Additionally, dependent features often enable better classication performances [36]. The initial input feature vectors $\mathbf{x}_j = \{x_{1j}, x_{2j}, \ldots, x_{pj}\}$, with $p = 100$ were composed of:

- *Heart rate variability (HRV) derived features* (x_{1j}, x_{2j}, x_{3j}): Atrial (S) and ventricular (V) ectopic beats manifest abrupt changes on fiducial point intervals. In addition, morphology is sometimes very similar to normal heartbeats [30] and therefore temporal attributes are essential to detect such heartbeats. These features are computed as in [29]:

$$\begin{aligned} x_{1j} &= l_j - l_{j-1}, && (RR\,\text{interval}) \\ x_{2j} &= l_{j-1} - l_{j-2}, && (pre\text{-}RR\,\text{interval}) \\ x_{3j} &= l_{j+1} - l_j, && (post\text{-}RR\,\text{interval}) \end{aligned}$$

  where l_j accounts for the location of the R-wave of the j-th heartbeat.
- *Heartbeat prematurity features* (x_{4j}, x_{5j}, x_{6j}): Prematurity features account for changes in heart rate [30], and contribute to the identification of S beats. The two first features, x_{4j} and x_{5j}, can be readily calculated as $x_{4j} = x_{1j} - x_{2j}$ and $x_{5j} = x_{3j} - x_{1j}$. Not only is the absolute value of these features informative, but also their sign changes. Feature x_{6j} rates the number of consecutive S beats and is also sensitive to an increase of the heart rate. This parameter was first introduced by Rodríguez-Sotelo et al. [30], and is defined as follows:

$$x_{6j} = \left(\frac{x_{3j}}{x_{1j}}\right)^2 + \left(\frac{x_{2j}}{x_{1j}}\right)^2 - \left(\frac{1}{3}\sum_{k=1}^{3} x_{kj}^2 \log(x_{kj})^2\right). \quad (9)$$

  The first and second squared terms in Eq. (9) are sensitive to abrupt heart rate changes. The last term corresponds to an unnormalized Shannon entropy factor, which increases the value of x_{i6} whenever heart rate increases steadily.
- *Morphological features*(x_{7j}, x_{8j}, x_{9j}): Morphology is one of the most powerful heartbeat discriminating attribute. It can be captured in the time domain using a number of well known techniques [4]. Feature x_{7j} quantifies the morphological dissimilarity between current QRS-complex, and a linearly averaged QRS-complex of the last 10 complexes [37], by means of a dynamic time warping (DTW) approach [4]. Feature x_{8j} is aimed at segmenting ventricular arrhythmias that exhibit abnormal QRS-complexes, such as ventricular extrasystoles (V) or branch blocks (N) [34]. It is defined as:

$$x_{8j} = \left|\frac{\max\{\mathbf{QRS}_j[t]\}}{\min\{\mathbf{QRS}_j[t]\}}\right|,$$

  being $\mathbf{QRS}_j[t]$ the discrete time representation of the current QRS-complex [34]. Finally, feature x_{9j} represents the energy of a QRS-complex, computed as $x_{9j} = \sum_{k=0}^{L_j} \mathbf{QRS}_j[t]^2$, where L_j is the length of the j-th QRS-complex.
- *Basis functions features*(x_{10j}, …, x_{100j}): The representation of heartbeats by means of the coefficients of certain basis functions has also been used successfully in electrocardiogram clustering and compression applications [33]. Hence, features x_{10j} to x_{19j} correspond to the Hermite coefficients obtained as described in [33]. Features x_{20j} to x_{90j} represent the 4th-level coefficients of a Daubechies-2, (*dB2*), Wavelet heartbeat decomposition, used in the same way as in [35] (approximation, A, and detail, D, $x_{20j-25j} \Leftarrow A_4$, $x_{26j-31j} \Leftarrow D_4$, $x_{32j-41j} \Leftarrow D_3$, $x_{43j-58j} \Leftarrow D_2$, $x_{59j-90j} \Leftarrow D_1$). In addition, the variance and maximum values of the previous coefficients were also included as features ($x_{91j-95j} = \text{var}\{A_4, D_4, D_3, D_2, D_1\}$, $x_{96j-100j} = \max\{A_4, D_4, D_3, D_2, D_1\}$).

3.4. *Feature selection*

The feature selection stage was carried out using the new method described in Section 2.1. All the feature vectors $\mathbf{x}_j = \{x_{1j}, x_{2j}, \ldots, x_{100j}\}$, obtained as the output of the feature extraction stage (Section 3.3), and arranged in a matrix $\mathbf{X}$, underwent the process based on the $Q-\alpha$ algorithm for feature weighting and dimensionality reduction. The selected features, corresponding to a score $\widehat{\alpha_i} > 0$ and included by the variance criterion, were projected and arranged in a new set $\mathbf{Y}$ to be used as representatives of each heartbeat in the final clustering stage.

3.5. *Feature vector clustering*

Vectors in $\mathbf{Y}$ were clustered in order to find an unsupervised partition of the heartbeats in each ECG record. The objective was to find a partition that matched the heartbeat types. This process was carried out in three steps:

- *Estimate the number of clusters*. This number was obtained by means of a spectral analysis of an affinity measure, as described in [38].
- *Initialize cluster centroids*. The first cluster centroids were found based on the J–H-means clustering algorithm [39], where the objective function used was the minimum sum-of-squares (MSS). The details of this method can be found in [30].
- *Compute the partition*. The final partition was obtained using a clustering algorithm based on the Gaussian expectation maximization clustering (GEMC) algorithm [35].

Any clustering method needs to optimize an objective function in terms of intracluster and intercluster dissimilarity [4]. We chose a feature vector dissimilarity measure based on a nonlinear approach, DTW. This method has proven to be very effective in similar tasks [4,34].

The accuracy of the resulting heartbeat clustering was assessed in terms of sensitivity (*Se*), specificity (*Sp*), and clustering performance (C_P), defined as:

- $Se = \frac{T_P}{T_P+F_N}$. Sensitivity. Ratio of correctly not included heartbeats in a cluster.
- $Sp = \frac{T_N}{T_N+F_P}$. Specificity. Ratio of correctly included heartbeats in a cluster.
- $C_P = \frac{T_N+T_P}{T_N+F_P+T_P+F_N}$. Clustering performance. Ratio of correctly grouped heartbeats.

where T_N and T_P are the true negatives and positives respectively, and F_N and F_P the corresponding false negatives and positives.

3.6. Comparative experiments

The experiments were repeated removing or replacing the relevance analysis method with another feature selection stage for comparative purposes. Two methods based on classical PCA and weighted PCA (WPCA) [9] were applied instead of the algorithm proposed. The complete iterative version of the $Q-\alpha$ algorithm was also used (represented by $\boldsymbol{\alpha}$ instead of $\hat{\boldsymbol{\alpha}}$). Additionally, the results can be compared with similar works that also used the BIT/MIH database, such as [29,30,33,34].

4. Results

4.1. Overall results

The method described was applied to all the records in the experimental database, namely, using 6 epochs for record, and the approximated $\hat{\boldsymbol{\alpha}}$ solution. The results in terms of *Se*, *Sp*, and C_P mean–variance, $(\mu - \sigma^2)$, are shown in Table 2. Each column corresponds to a different heartbeat type: *N*, *S*, *V*, *F*, and *Q*. Globally, the *Se* achieved was 85.88% with a variance of 11.01. Results for other methods are also included in Table 2.

4.2. Results using record segmentation

The experiments were repeated processing the records in segments instead, and then combining the partial results to obtain the final ones in each case. Thus, experiments were repeated for records split in 2–10 equal length epochs. The results for computational cost and clustering performance are graphically shown in Fig. 1. The results for methods based on PCA and WPCA are also included in Fig. 1 for comparative purposes of accuracy and temporal cost.

4.3. Weighting vector α results

Fig. 2 shows an example of the $\boldsymbol{\alpha}$ vector corresponding to the relevance analysis of the last 5 min of record 217, as well as the comparative results for PCA-based methods. The rest of results for $\boldsymbol{\alpha}$ have been obviously omitted but the example provided is very representative of the general trend.

4.4. Temporal cost results

The method proposed was implemented on a standard PC computer (CPU: Intel(R) Core(TM)2 Duo CPU T6670 2.2 GHz. RAM: 4 Gb. OS: 32 bits, Windows Vista) using the Matlab© environment. The average temporal cost for processing each record in the MIT/BIH database was 6.1896 s with a variance of 2.9353.

5. Discussion

5.1. Influence of heartbeat type

- Normal heartbeats (*N* type). All the compared methods performed equally well. *Se* and *CP* where higher than 99% in all methods, being *Sp* slightly lower, specially for the methods based on PCA. Type *N* are the most regular and numerous heartbeats and therefore they result in densely populated clusters clearly separated from the other types. Moreover, outliers have little influence on the quantitative performance due to the high number of heartbeats.
- Supraventricular ectopic heartbeats (*S* type). The performance was still high although lower than in the previous case. This is so because *S* heartbeats are not as numerous as *N* heartbeats and the influence of erroneously clustered heartbeats is higher. Variance is also relatively high, which means that the performance depends on the specific record processed. Still, the methods based on the $Q-\alpha$ algorithm perform better than those based on PCA.
- Ventricular ectopic heartbeats (*V* type). This type of heartbeats is also relatively scarce but methods performed better in this case than for *S* heartbeats because temporal parameters add separability to *V* heartbeats (specially prematurity) [34]. While C_P was higher than 98% for the method proposed, it decreases to 93% for the PCA approach. *Se* is also significantly lower.
- Fusion heartbeats (*F* type). Fusion *F* beats are very difficult to characterize, since their morphology is often very similar to that of other types [29,30]. Owe to this low separability, the performance in terms of *Se* and C_P was specially low, from 70.73% down to a poor 42.44%, but similar to other previous works [29,30,33,34].
- Unknown heartbeats (Q type). These heartbeats were correctly classified in most of the cases, but since they correspond to unknown specific types, in some instances there could be an overlapping with other type features, accounting for a lower performance than for *N*, *S*, and *V* types.

In general, the method performance regarding heartbeat type varies greatly. For instance, for the $\hat{\boldsymbol{\alpha}}$ weighting in the first row, $Se_N = 99.06\%$, $Se_V = 96.31\%$, whereas $Se_F = 50.59\%$. Being these results very similar to those obtained in previously published works, the uneven number of heartbeat types, namely, the unbalanced classes, causes any method to achieve a lower performance.

Table 2 – Global clustering results for all groups of heartbeats in the arrhythmia subdatabase.

	N $\mu - \sigma^2$	S $\mu - \sigma^2$	V $\mu - \sigma^2$	F $\mu - \sigma^2$	Q $\mu - \sigma^2$
α					
Se	99.25 – 2.48	91.11 – 15.66	96.11 – 8.24	70.73 – 32.05	91.98 – 17.83
Sp	95.77 – 9.12	99.36 – 2.19	99.87 – 0.24	99.59 – 0.77	99.79 – 0.47
CP	99.16 – 2.5	96.18 – 6.15	98.29 – 3.37	93.29 – 10.62	99.91 – 0.21
$\hat{\alpha}$					
Se	99.06 – 3.5	93.52 – 16.54	96.31 – 7.03	50.59 – 35.33	89.9 – 26.77
Sp	95.32 – 9.49	99.39 – 1.62	99.84 – 0.31	99.51 – 0.99	99.4 – 0.73
CP	99.25 – 2.01	96.24 – 6.67	98.13 – 3.86	81.64 – 32.51	99.96 – 0.06
PCA					
Se	99.27 – 2.55	91.26 – 15.26	91.5 – 21.52	50.57 – 34.63	79.91 – 44.67
Sp	92.19 – 19.06	99.5 – 1.51	99.86 – 0.26	99.6 – 0.76	99.85 – 0.34
CP	99.14 – 2.46	95.68 – 7.47	93.71 – 20.77	81.05 – 33.04	79.93 – 44.68
WPCA					
Se	99.24 – 2.79	90.08 – 18.57	91.62 – 21.61	42.44 – 39.57	87.86 – 26.75
Sp	92.17 – 19.37	99.58 – 1.22	99.85 – 0.28	99.53 – 0.92	99.79 – 0.47
CP	99.12 – 2.58	96.16 – 6.07	93.77 – 20.79	68.87 – 41.56	99.91 – 0.21

N, S, V, F, and Q account for the AAMI heartbeat types. μ represents the average for *Se*, *Sp*, and *CP*, and σ^2 their variance.

Overall performance was *Se* = 85.88% for the $\hat{\alpha}$ method proposed. The performance using other methods was lower than that, except for the complete case of α (89.84%). For example, using no feature selection stage, that is, the 100 features, *Se* was 84.59%, with a weighted PCA approach, *Se* was 85.24%, and with a standard PCA scheme, *Se* = 85.35%, under the same conditions as for the method proposed. These results confirm that a feature selection stage can be a key component of any clustering approach, unless it was excessively time-consuming or complex.

5.2. *Influence of number of epochs*

When the records were processed using a number of epochs greater than 1, the results improved in general (Fig. 1). This is due to the dynamic nature of the records, that is, the ECGs, specially those of long duration, exhibit a great non-stationarity, which in turn is translated to the features. For example, some heartbeat types may appear more frequently in some signal zones than in others due to different patient activity or medication.

Using an unsupervised feature selection method as the one described in this paper, when applied to an entire record, an averaging effect may result in a poorer heartbeat type separation capability. Processing a record epoch by epoch yields a better feature adaptation to the specific properties of the record zone, and therefore, a more accurate heartbeat representation. However, a tradeoff among the number of epochs, processing cost, and accuracy must be considered.

The dynamic nature of an ECG also makes it difficult to define a general set of optimal features. From the initial set of 100 features, the specific features selected varied enormously

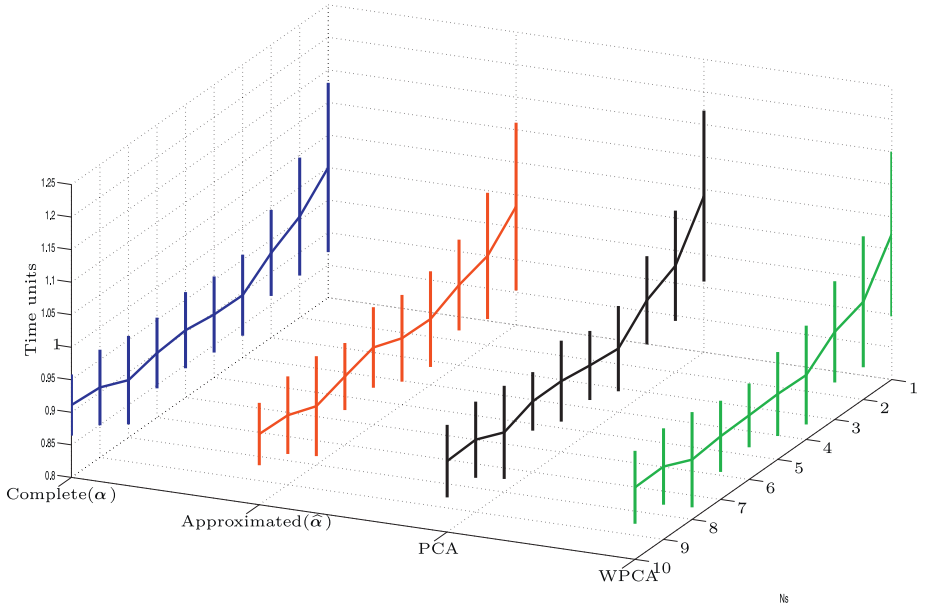


Fig. 1 – Results of computational cost for the performance of several feature selection methods depending on the number of record epochs. Amplitude of vertical lines at each point represent the variance of the result.

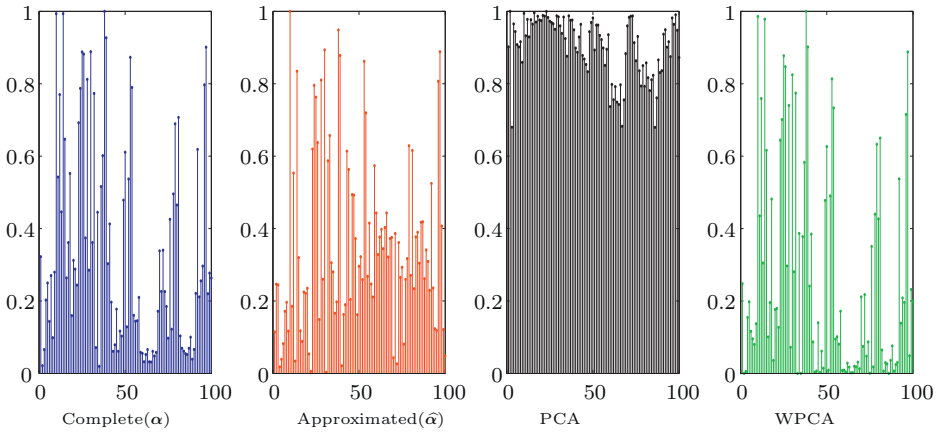


Fig. 2 – Results of computed feature relevance for the last 5 min of MIT/BIH record 217 using the method proposed and PCA-based methods.

from epoch to epoch, and from record to record. This is why we do not try to list a set of recommended features but instead, suggest to let the method choose them in an unsupervised way. Additionally, the algorithm also determines the number of selected features automatically from the accumulated variance criterion.

As an example of this non-stationary behaviour of an ECG, and consequently, of the heartbeat features, Fig. 3 shows graphically the processing of a record using the method proposed. The plot on the top corresponds to MIT/BIH record 207, segmented into 6 epochs, S_1–S_6, of 5 min length. This record contains heartbeat types *N*, *S*, and *V* (according to AAMI labelling). The resulting weight vectors $\hat{\boldsymbol{\alpha}}$ for epochs S_1, S_5 and S_6 (zoomed and shown in the three central plots) are depicted in the three subplots at the bottom. It can be seen that in S_1, the relevant variables correspond to the Wavelet transform features. In S_5, the Hermite coefficients had higher weights because these coefficients characterize appropriately the morphology of beats type *L* (MIT/BIH labelling) and *V*. Finally, in S_6, the weight for each one of the first 3 attributes (*HRV* features) increased.

5.3. *Influence of feature weighting*

From a clustering point of view, the effect of a feature weighting scheme is to enhance the separability of the heartbeat types, in other words, to increase the intercluster distance. As an example, Fig. 4 depicts the weighting results for the three first principal components of MIT/BIH record 217. As can be seen, all the methods improve the heartbeat separation obtained from the raw features (first plot on the left). In this example, the method based on $\boldsymbol{\alpha}$ provides the highest separability of all.

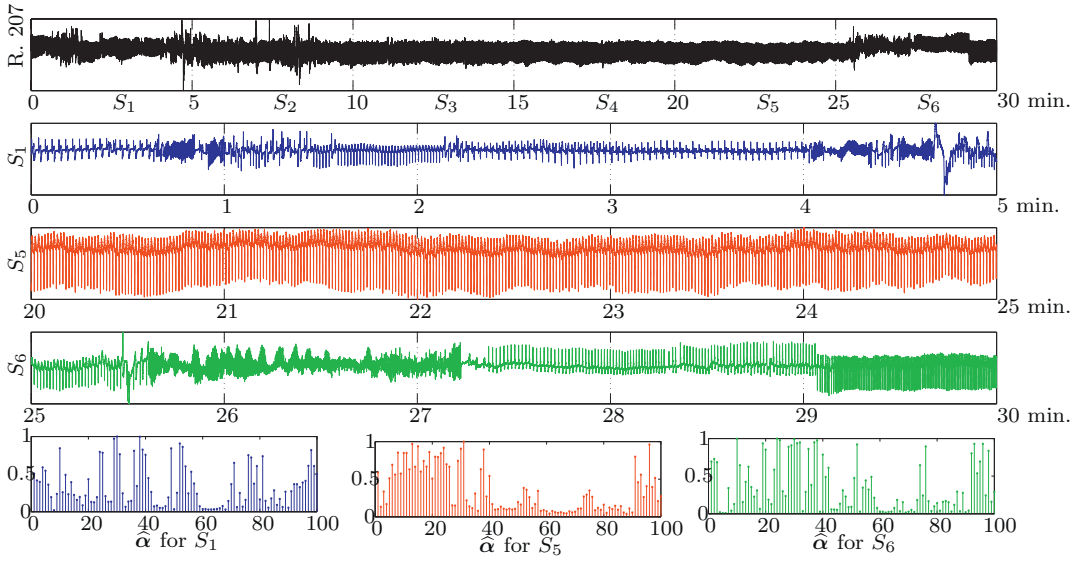


Fig. 3 – Results of the feature relevance analysis with the method proposed. Top plot corresponds to MIT/BIH record 207, which is split into 6 epochs of 5 min, S_1 to S_6. The three following plots show epochs S_1, S_5 and S_6. The bottom plots represent graphically the weights obtained for each $\hat{\alpha}$.

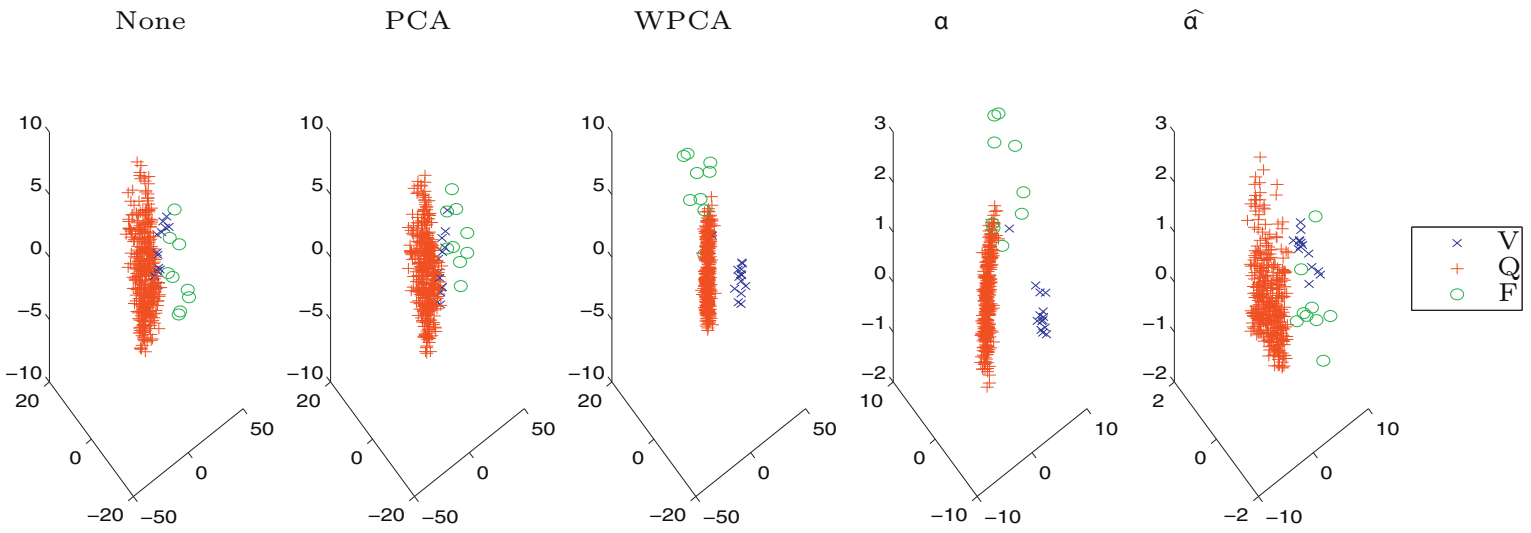


Fig. 4 – Feature separability example for the first 3 principal components for MIT/BIH recording 217, after weighting the feature matrix with different methods.

5.4. *Temporal cost*

Computational cost is a big issue in ECG signal clustering or classification [4]. Some methods yield a good accuracy but the temporal requirements make them of limited clinical use in real time applications or long term record analysis.

The method proposed ($\hat{\boldsymbol{\alpha}}$) took on average 3.77 s per record. Other methods enumerated in Section 3.6, took 6.18 s, 7.54 s, 8.79 s, and 6.63 s (α, WPCA, PCA, or no feature selection approach, respectively). In comparison to no feature selection stage included, the algorithm described exhibited a temporal cost 43% lower. It has to be noted that the speed-up is masked by the rest of stages such as feature extraction or clustering, that remain the same for all methods and also involve a significant computational burden. Moreover, the duration of the records employed in the experiments was 30 min, and since the clustering applied has a quadratic time complexity ($O(n^2)$), the temporal cost differences could be much greater for longer records.

The influence of record split on temporal cost is depicted in Fig. 1. It has a positive effect on clustering accuracy and computational complexity. The processing time decreases down to a minimum at 4 epochs, and then remains stable. Furthermore, in most cases, the processing time of all epochs is considerably shorter than the time of analysis of a whole record. The windowing approach significantly reduces the computational cost due to the quadratic nature of the clustering, and a higher $\hat{\boldsymbol{\alpha}}$ sparsity for lower n.

6. Conclusion

This paper describes a new unsupervised heartbeat feature selection method based on a relevance analysis. We customized a general scheme derived from the $Q-\alpha$ algorithm [21] to be applied in heartbeat clustering approaches. In contrast to other similar studies or methods, the approach proposed exhibits a lower temporal cost at a similar classification accuracy. The core of the method described is the so-called $Q-\alpha$ algorithm [21]. This is an iterative method that converges rapidly (on average only 5 or less iterations were needed) to maximize a relevance function related to the input heartbeat features. In addition, this method automatically computes the optimal value for the number of selected features, q. The $Q-\alpha$ algorithm can be easily implemented in a computer application since it only involves basic matrix operations. This method achieved the highest performance, a global 89.84% sensitivity.

However, we proposed a new scheme of this algorithm where only one iteration is needed, obtaining an approximation of the weight vector $\boldsymbol{\alpha}$, denoted $\hat{\boldsymbol{\alpha}}$. This scoring is then applied to the p heartbeat feature vectors to obtain the representative clustering set. Although the sensitivity decreased with our approach (85.88%), the computational cost was significantly lower.

Researchers or programmers can use any of the two methods. When accuracy is the key issue and time constraints are relaxed, the complete $\boldsymbol{\alpha}$ calculation can be carried out as described in detail in [21]. Otherwise, the method proposed achieves also a high accuracy but at much lower temporal cost and algorithm complexity. In addition, both methods are unsupervised, and therefore they can better suit real clinical applications where pre-labelled sets are usually not available.

This study also demonstrated that more features does not necessarily mean higher accuracy. With an average of 18 selected features, overall *Se* was 85.88%, whereas with all the 100 features, *Se* was lower, 84.5895%. Despite the feature relevance non-stationarity, a voting scheme could be applied to discard or preserve in advance a number of those 100 features and further improve the efficiency of the algorithm.

The temporal cost of the whole process was also reduced. Although it could be argued that a feature selection stage adds computational cost, the feature dimensionality reduction achieved, and an efficient feature selection design, outperforms the additional matrix operations involved. Besides, the record segmentation proposed contributes to a lower time complexity.

Conflict of interest

None declared.

Acknowledgements

This research was carried out under grant "*Centro de Investigación e Innovación de Excelencia - ARTICA*", funded by COLCIENCIAS (Colombia), the program "*Apoyo a tesis de posgrado cod. 10242*", from the Universidad Nacional de Colombia, and also supported by the Spanish Ministry of Science and Innovation, research project TEC2009-14222.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cmpb.2012.04.007.

REFERENCES

[1] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Transactions on Knowledge and Data Engineering 17 (4) (2005) 491–502, http://dx.doi.org/10.1109/TKDE.2005.66.

[2] T.-H. Cheng, C.-P. Wei, V.S. Tseng, Feature selection for medical data mining: comparisons of expert judgment and automatic approaches, in: Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, 2006, pp. 165–170.

[3] V. Hodge, J. Austin, A survey of outlier detection methodologies, Artificial Intelligence Review 22 (2) (2004) 85–126, http://dx.doi.org/10.1023/B:AIRE.0000045502.10941.a9.

[4] D. Cuesta, J. Pérez-Cortés, G. Andreau-García, Clustering of ECG signals in computer-aided Holter analysis, Computer Methods and Programs in Biomedicine 72 (2003) 179–196.

[5] J. Soni, U. Ansari, D. Sharma, S. Soni, Predictive data mining for medical diagnosis: an overview of heart disease prediction, International Journal of Computer Applications 7 (8) (2011) 43–48.

[6] S.N. Ghazavi, T.W. Liao, Medical data mining by fuzzy modeling with selected features, Artificial Intelligence in Medicine 43 (2008) 195–206, http://dx.doi.org/10.1016/j.artmed.2008.04.004.
[7] V. Monasterio, P. Laguna, J.P. Martínez, Multilead analysis of T-wave alternans in the ECG using principal component analysis, IEEE Transactions on Biomedical Engineering 6 (7) (2009) 1880–1890.
[8] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, J.M. Roig, Principal component analysis in ECG signal processing, EURASIP Journal of Applied Signal Process 2007 (2007), 98–98.
[9] J.L. Rodríguez-Sotelo, E. Delgado-Trejos, D.H. Peluffo-Ordoñez, D. Cuesta-Frau, C.G. Castellanos-Domínguez, Weighted-PCA for unsupervised classification of cardiac arrhythmias, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 1906–1909, http://dx.doi.org/10.1109/iembs.2010.5627321.
[10] T. Hoya, G. Hori, H. Bakardjian, T. Nishimura, T. Suzuki, Y. Miyawaki, A. Funase, J. Cao, Classification of single trial EEG signals by a combined principal + independent component analysis and probabilistic neural network approach, in: International Symposium on Independent Component Analysis and Blind Signal Separation, 2003.
[11] J.E. Arruda, H.A. McGee, H. Zhang, C.J. Stanny, The effects of EEG data transformations on the solution accuracy of principal component analysis, Psychophysiology 48 (3) (2011) 370–376, http://dx.doi.org/10.1111/j.1469-8986.2010.01067.x.
[12] J. Hayano, Y. Sakakibara, A. Yamada, M. Yamada, S. Mukai, T. Fujinami, K. Yokoyama, Y. Watanabe, K. Takata, Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects, The American Journal of Cardiology 67 (2) (1991) 199–204, http://dx.doi.org/10.1016/0002-9149(91)90445-Q.
[13] J.A. Smith, R. Drake, A. Simpson, A. Woodcock, A. Pickles, A. Custovic, Dimensions of respiratory symptoms in preschool children: population-based birth cohort study, American Journal of Respiratory Critical Care Medicine 177 (12) (2008) 1358–1363, http://dx.doi.org/10.1164/rccm.200709-1419OC.
[14] R.G.N.D.N.F.M. Pagani, D. Salmaso, Principal component analysis in mild and moderate alzheimer's disease—a novel approach to clinical diagnosis, Psychiatry Research: Neuroimaging 173 (2009) 8–14.
[15] L. Rocchi, L. Chiari, A. Cappello, F.B. Horak, Identification of distinct characteristics of postural sway in Parkinson's disease: a feature selection procedure based on principal component analysis, Neuroscience Letters 394 (2) (2006) 140–145.
[16] S. Karamchandani, U.B. Desai, S.N. Merchant, G.D. Jindal, Principal component analysis based backpropagation algorithm for diagnosis of peripheral arterial occlusive diseases, in: CCECE, IEEE, 2009, pp. 482–485.
[17] Z. Luo, X. Wu, S. Guo, B. Ye, Diagnosis of breast cancer tumor based on pca and fuzzy support vector machine classifier, International Conference on Natural Computation 4 (2008) 363–367, doi:http://doi.ieeecomputersociety.org/10.1109/ICNC.2008.932.
[18] O. Tadevosyan-leyfer, M. Dowd, R. Mankoski, B. Winklosky, S. Putnam, L. McGrath, H. Tager-flusberg, S.E. Folstein, A principal components analysis of the autism diagnostic interview-revised, Journal of the American Academy of Child and Adolescent Psychiatry 42 (7) (2003) 864–872, http://dx.doi.org/10.1097/01.CHI.0000046870.56865.90.
[19] S. Serneels, T. Verdonck, Principal component analysis for data containing outliers and missing elements, Computational Statistics and Data Analysis 52 (2008) 1712–1727, http://dx.doi.org/10.1016/j.csda.2007.05.024.
[20] H.-P. Kriegel, P. Kröger, E. Schubert, A. Zimek, A general framework for increasing the robustness of PCA-based correlation clustering algorithms, in: Proceedings of the 20th International Conference on Scientific and Statistical Database Management, SSDBM 08, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 418–435.
[21] L. Wolf, A. Shashua, Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weight-based approach, Journal of Machine Learning 6 (2005) 1855–1887.
[22] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220, circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.
[23] L.-J. Zhang, Z.-J. Li, H.-W. Chen, An effective gene selection method based on relevance analysis and discernibility matrix, in: Z.-H. Zhou, H. Li, Q. Yang (Eds.), Advances in Knowledge Discovery and Data Mining, Vol. 4426 of Lecture Notes in Computer Science, Springer Berlin, Heidelberg, 2007, pp. 1088–1095.
[24] L. Talavera, Dynamic feature selection in incremental hierarchical clustering, in: Proceedings of the 11th European Conference on Machine Learning, ECML'00, Springer-Verlag, London, UK, 2000, pp. 392–403.
[28] K.D. Toennies, A. Celler, S. Blinder, T. Moeller, R. Harrop, Scatter segmentation in dynamic spect images using principal component analysis, in: Proceedings of SPIE (Medical Imaging), San Diego, Part I, 2003, pp. 507–516.
[29] P. De-Chazal, M. O'Dwyer, R. Reilly, Automatic classification of heartbeats using ecg morphology and heartbeat interval features, IEEE Transaction on Biomedical Engineering 1 (7) (2004) 1196–1206.
[30] J. Rodríguez-Sotelo, D. Cuesta-Frau, G. Castellanos-Domínguez, Unsupervised classification of atrial heartbeats using a prematurity index and wave morphology features, Medical and Biological Engineering and Computing 7 (7) (2009) 731–741.
[31] P. Micó, M. Mora, D. Cuesta-Frau, M. Aboy, Automatic segmentation of long-term ECG signals corrupted with broadband noise based on Sample Entropy, Computer Methods and Programs in Biomedicine 98 (2010) 118–129.
[32] A. Quiceno-Manrique, J. Alonso-Hernandez, C. Travieso-Gonzalez, M. Ferrer-Ballester, G. Castellanos-Dominguez, Detection of obstructive sleep apnea in ECG recordings using time-frequency distributions and dynamic features, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2009, pp. 5559–5562.
[33] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, L. Sörnmo, Clustering ecg complexes using Hermite functions and self-organising maps, IEEE Transactions on Biomedicine 48 (2000) 838–847.
[34] D. Cuesta, M. Biagetti, R. Quinteiro, P. Mico-Tormos, M. Aboy, Unsupervised classification of ventricular extrasystoles using bounded clustering algorithms and morphology matching, Medical and Biological Engineering and Computing 5 (3) (2007) 229–239.
[35] D. Cvetkovic, E.D. Ubeyli, I. Cosic, Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study, Digital Signal Processing 18 (5) (2008) 861–874, http://dx.doi.org/10.1016/j.dsp.2007.05.009.

[36] P. Long, V. Vega, Boosting and microarray data, Machine Learning 52 (2003) 31–44.
[37] L. Sörnmo, P. Laguna, Bioelectrical signal processing in cardiac and neurological applications, Academic Press Series in Biomedical Engineering, Academic Press, 2005, ISBN 0124375529.
[38] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, in: Advances in Neural Information Processing Systems, vol. 14, MIT Press, 2001, pp. 849–856.
[39] P. Hansen, N. Mladenovic, *J*-means: a new local search heuristic for minimum sum of squares clustering, Pattern Recognition 34 (2001) 405–413.