Hindawi
Complexity
Volume 2018, Article ID 1435203, 18 pages
https://doi.org/10.1155/2018/1435203

*Research Article*

# Complexity-Based Discrepancy Measures Applied to Detection of Apnea-Hypopnea Events

**R. E. Rolón**,[1,2] **I. E. Gareis**,[1,3] **L. E. Di Persia**,[1] **R. D. Spies**,[4] **and H. L. Rufiner**[1,3]

[1]*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), UNL, CONICET, FICH, Santa Fe, Argentina*
[2]*Facultad Regional Paraná, Universidad Tecnológica Nacional, Paraná, Entre Ríos, Argentina*
[3]*Laboratorio de Cibernética, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Argentina*
[4]*Instituto de Matemática Aplicada del Litoral, IMAL, UNL, CONICET, FIQ, Santa Fe, Argentina*

Correspondence should be addressed to R. E. Rolón; rrolon@sinc.unl.edu.ar

Received 2 February 2018; Accepted 30 May 2018; Published 26 August 2018

Academic Editor: Dimitri Volchenkov



In recent years, an increasing interest in the development of discriminative methods based on sparse representations with discrete dictionaries for signal classification has been observed. It is still unclear, however, what is the most appropriate way for introducing discriminative information into the sparse representation problem. It is also unknown which is the best discrepancy measure for classification purposes. In the context of feature selection problems, several complexity-based measures have been proposed. The main objective of this work is to explore a method that uses such measures for constructing discriminative subdictionaries for detecting apnea-hypopnea events using pulse oximetry signals. Besides traditional discrepancy measures, we study a simple one called Difference of Conditional Activation Frequency (DCAF). We additionally explore the combined effect of overcompleteness and redundancy of the dictionary as well as the sparsity level of the representation. Results show that complexity-based measures are capable of adequately pointing out discriminative atoms. Particularly, DCAF yields competitive averaged detection accuracy rates of 72.57% at low computational cost. Additionally, ROC curve analyses show averaged diagnostic sensitivity and specificity of 81.88% and 87.32%, respectively. This shows that discriminative subdictionary construction methods for sparse representations of pulse oximetry signals constitute a valuable tool for apnea-hypopnea screening.

## 1. Introduction

Although it is widely used and accepted, the notion of complexity has very often avoided a rigorous formalization. It is therefore not surprising that no universally accepted measure exists yet for quantifying such a concept. In particular, within information theory, the complexity of any element of a code, or of any feature of a signal representation in the context of signal processing, is known to be strongly related to the information it carries or, more precisely, to the value of its entropy. It is important to point out however that, in the context of signal classification, the more informative features (in terms of classification) are not necessarily the ones with larger entropy. Hence, more "ad hoc" measures are needed. In fact, any appropriate complexity measure corresponding to a given feature should be instead, strongly related to the amount of information about class membership provided by such a feature. One could then think of using as measure of complexity the conditional entropy of the class given the feature. However, features providing the most discriminative information regarding a class are almost always those with lower conditional entropy values, and hence, the best features for classification purposes will be the least complex ones.

Information theory was originally based on the engineering of noisy communication channels, and it is closely associated to a large number of disciplines such as signal processing, artificial intelligence, complex systems, and pattern recognition, to name only a few. We are particularly interested in the latter. Pattern recognition is a discipline which is mainly oriented to the generation of algorithms or methods that can decide an action based upon certain recognized similarities (patterns) in the input data. Within signal classification, which is perhaps one of the most important subfields of pattern recognition, several discrepancy

measures have been used in problems coming from a wide variety of areas such as machine learning [1], image and speech processing [2], neural networks [3], and biomedical signal processing [4, 5]. Among them, the most commonly used is probably the Kullback-Leibler (KL) divergence [6, 7]. This divergence, also known as relative entropy, was used as a discriminative measure for selecting, from a large collection of orthonormal bases, the one attaining maximum information [1]. A more recent approach was introduced by Gupta et al. [8] who used this divergence as a discrepancy measure in the traditional k-nearest neighbor (k-NN) algorithm, yielding competitive classification performances in the context of raw electroencephalographic signal classification. Although it provides certain computational and theoretical advantages, the lack of symmetry of the KL divergence has motivated the development of several symmetric versions such as the so-called J divergence [9] and the well-known and widely used Jensen-Shannon divergence [10].

Sparse representation of signals constitutes a useful technique which has drawn wide interest in recent years due to its success in many applications such as signal and image processing [11]. This technique allows the analysis of the signals by means of only a few well-defined basic waveforms. Due to its advantages, such as robustness to noise and dimension reduction, sparse representation has acquired a large popularity in the area of biomedical signal processing. For example, this technique has been successfully applied to several problems including the estimation of the human respiratory rate [12] and electrocardiographic signal processing, both for signal enhancement and QRS complex detection, for improving heart disease analysis and diagnosis [13]. It is timely to point out however that, up to our knowledge, no applications of discrepancy measures to sparse representation for signal classification are known yet.

All reconstructive methods, such as principal component analysis (PCA), independent component analysis (ICA), and the previously mentioned sparse representations [14], produce particular types of signal representations minimizing a given cost functional which usually involves both fidelity and regularization terms. These methods have been successfully applied in a wide variety of problems such as signal denoising, missing data, and outliers. On the other hand, discriminative methods such as linear discriminant analysis (LDA) are oriented to find optimal decision boundaries to be used for classification tasks. It is well known that for signal classification, which is our main interest in this work, discriminative methods generally outperform reconstructive methods. It is mainly for this reason that several authors have recently developed supervised approaches based on sparse representation which are simultaneously reconstructive and discriminative [15, 16].

The obstructive sleep apnea-hypopnea (OSAH) syndrome [17] is one of the most common sleep disorders and more often than not it remains undiagnosed and therefore not treated. This syndrome is caused by repeated events of partial or total blockage of the upper airway during sleeping, which correspond to events of hypopnea and apnea, respectively. To evaluate the severity degree of the OSAH syndrome, medical physicians have created the so-called apnea-hypopnea index (AHI), which is defined as the average number of apnea-hypopnea events per hour of sleep. In terms of this index, OSAH is classified as normal, mild, moderate, or severe depending on whether such an index falls in the interval $[0, 5)$, $[5, 15)$, $[15, 30)$ or $[30, \infty)$, respectively. The gold standard test for OSAH diagnosis is a study called polysomnography (PSG). However, PSG is both costly and lengthy and the accessibility to this type of study is limited. Additionally, PSG studies require information coming from a variety of physiological signals such as electroencephalography (EEG), airflow and pulse oximetry ($SaO_2$). It is known however that cessations of breathing associated with apnea-hypopnea events are always accompanied by a drop in the oxygen saturation level in the $SaO_2$ signal record, although quite often such a drop is very small and almost impossible to detect by a human observer.

The main objective of this work is precisely to develop a technique based on sparse representations and the use of appropriate discriminative information that be able to accurately and efficiently detect apnea-hypopnea events by using only the $SaO_2$ signal. Several ways exists for combining discriminative information and sparse representations within the context of signal classification. We shall follow one consisting of using the discriminative information for detecting those atoms having the most frequent activations in order to provide them as input for a classifier. This approach was initially introduced in [4] where two methods using the absolute value of the activation differences of the atoms as a measure of the discriminative information for the detection of OSAH were presented. In this work, a rigorous formalization of such a measure is introduced and compared with several other discrepancy measures for classifying apnea-hypopnea events. Also, the combined effect of using different sizes of nonredundant dictionaries and different sparsity degrees is explored in detail. Results show clearly that the proposed measure is capable of adequately pointing out discriminative atoms in a full dictionary, yielding competitive accuracy rates in the detection of individual apnea-hypopnea events. Additionally, this new approach is computationally very cheap. In fact, it has proved to be at least twice faster than those associated to all other discrepancy measures.

The rest of this article is organized as follows: in Section 2, the obstructive sleep apnea-hypopnea syndrome is explained. Sparse representation of signals is introduced in Section 3. The problem of finding discriminative subdictionaries is described in Section 4 while several discriminative information measures are presented in Section 5. Section 6 contains a detailed description about the performed experiments. Results and discussions are introduced in Section 7 while conclusions are presented in Section 8.

## 2. Sleep Apnea-Hypopnea

Apnea-hypopnea events occur as a consequence of a functional-anatomic disturbance of the upper airway producing its partial or total blockage. At the end of an apnea-hypopnea event, a pronounced desaturation of the blood hemoglobin commonly occurs. These desaturations generate

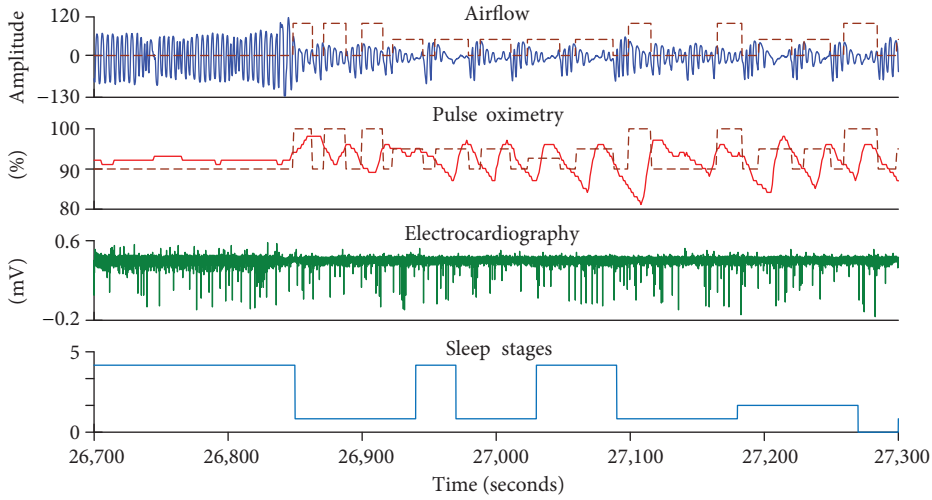


Figure 1: A portion of a few number of physiological signals coming from a full PSG. Dashed lines (brown) are apnea-hypopnea labels introduced by the medical expert.

characteristic patterns in the pulse oximetry record known as intermittent hypoxemias. The hypoxemia-reoxygenation cycles promote oxidative stress, angiogenesis, and tumor growth and favor the sympathetic activation with increment of blood pressure and systemic and vascular inflammation with endothelial dysfunction which contributes to multiorganic chronic morbidity, metabolic abnormalities, and cognitive impairment [18]. Additionally, strong correlations between neoplastic diseases and the OSAH syndrome have been described in [19]. Also, a recent study among male mice suggests that OSAH's intermittent hypoxia can be associated to fertility reduction [20]. Currently, this pathology affects more than 4% of the human population around the world [21]. Additionally, it was found that aging, male gender, snoring, and obesity are all risk factors for OSAH syndrome [22].

Although very limited in many countries, overnight polysomnography (PSG) is currently the gold standard tool for diagnosing OSAH syndrome. As previously mentioned, a full PSG consists of the simultaneous measurement of several physiological signals such as EEG, electrocardiography (ECG), respiratory effort, airflow, $SaO_2$, and electrical activity produced by skeletal muscles (EMG). Mainly due to its ease of acquisition, we are particularly interested in the $SaO_2$ signal. Figure 1 shows a typical temporal plot of just a few physiological signals coming from a full PSG. This figure also depicts a portion of an original raw airflow signal as well as the corresponding portion of the $SaO_2$ signal. The corresponding labels of apnea-hypopnea events (dashed lines) are also shown. Finally, at the bottom of this figure, the electrical activity of the heart as well as the sleep stages are shown. In a typical PSG study, after a normal period of sleep, the recorded signals are provided to medical experts who analyze the whole record and mark the apnea-hypopnea events and sleep stages, needed for the posterior evaluation the AHI index. Due to its complexity and cost, a few alternatives to PSG have been adopted. One of the most popular ones is the so-called home respiratory polygraphy (HRP) [23] which requires no neurophysiological signals. Although studies have shown that there exists a high correlation between AHI values generated by HRP and PSG studies [24], HRP still needs of several physiological signals, whose acquisition strongly affects the normal sleeping of the person. It is therefore highly desirable to develop a reliable OSAH screening system which makes use of as few as possible physiological signals. In this regard, pulse oximetry, being a cheap and noninvasive technique, has become a suitable alternative for screening purposes [25].

In this work, we shall develop a method for the detection of apnea-hypopnea events that uses only the $SaO_2$ signals. Our approach leads to a binary classification problem whose main purpose is the detection of the presence (or not) of events of apnea and hypopnea. It is timely to point out that although our method does take into consideration an appropriate fidelity term, we are by no means interested in achieving accurate signal representation.

## 3. Sparse Representations

As previously mentioned, one of the most popular reconstructive methods is based on sparse representations of the signals involved. Sparsity can be enforced by including upper bounds for the number of nonzero coefficients in the representation of the given signals in terms of atoms in a dictionary.

Formally, the problem of sparse representations of signals can be separated into two subproblems, the so-called sparse coding problem and the dictionary learning problem. We shall now proceed to describe in detail each one of these subproblems. To be more precise, let $\mathbf{x} \in \mathbb{R}^N$ be a discrete signal and let $\Phi \in \mathbb{R}^{N \times M}$ (generally with $M \geq N$) be a dictionary whose columns $\phi_j \in \mathbb{R}^N$ are atoms that we want to use for obtaining a representation of $\mathbf{x}$ of the form $\mathbf{x} = \Phi\mathbf{a}$. Here, and in the sequel, we shall refer to the vector $\mathbf{a} = [a_1\ a_2, \ldots, a_M]^T \in \mathbb{R}^M$ as a "representation" of $\mathbf{x}$. Sparsity consists essentially of obtaining a representation with as

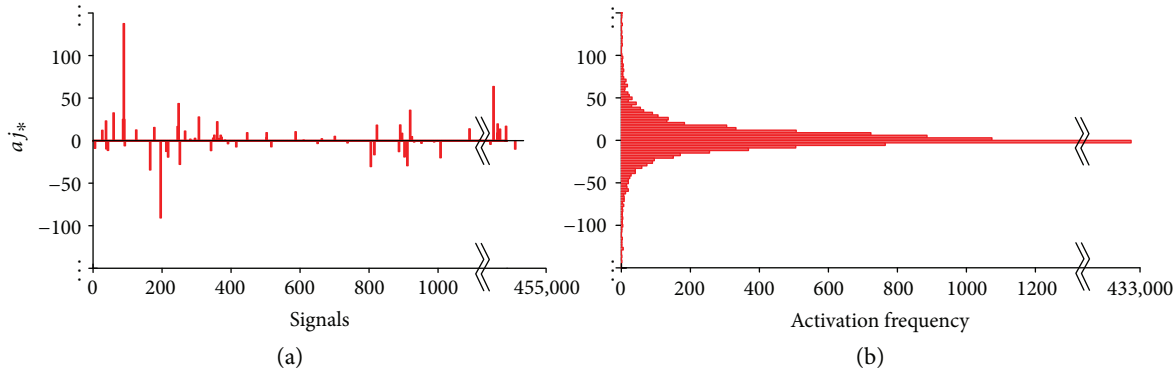


FIGURE 2: The values of the activations of a particular atom for each signal (a) and the corresponding histogram of activations (b).

few nonzero elements as possible. A way of obtaining such a representation consists of solving the following problem:

$$(P_0): \min_{\mathbf{a}} \quad \|\mathbf{a}\|_0 \quad \text{subject to} \quad \mathbf{x} = \Phi\mathbf{a}, \tag{1}$$

where $\|\mathbf{a}\|_0$ denotes the $l_0$ pseudonorm, defined as the number of nonzero elements of $\mathbf{a}$.

Several questions regarding problem $(P_0)$ immediately arise. Among them are the following: (i) does there exist an exact representation $\mathbf{x} = \Phi\mathbf{a}$?, (ii) if an exact representation exists, is it unique?, (iii) in the case of nonuniqueness, how do we find the "sparsest" representation? and (iv) how difficult is it, from the computational point of view, to solve problem $(P_0)$?. Although it is not an objective of this article to get into details about the answers to these questions, it turns out that imposing exact representation is most often a too restrictive and therefore inappropriate constrain and, on the other hand, solving $(P_0)$ is generally an NP-hard problem yielding this approach highly unsuitable for most applications. For more details, we refer the reader to ([26], § 1.8).

In order to overcome some of the difficulties which entail solving problem $(P_0)$, several relaxed versions of it have been considered. One of them consists of allowing a small representation error while imposing an upper bound on the $l_0$ pseudonorm of the representation:

$$(P_0^q): \min_{\mathbf{a}} \quad \|\mathbf{x} - \Phi\mathbf{a}\|_2 \quad \text{subject to} \quad \|\mathbf{a}\|_0 \leq q, \tag{2}$$

where $q$ is a prescribed integer parameter. This formulation takes into account the existence of possible additive noise terms; in other words, it assumes that $\mathbf{x} = \Phi\mathbf{a} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^N$ is a small energy noise term. Thus, this approach is particularly suitable in most real applications (such as biomedical signal processing) where measured signals are always contaminated by noise. Several greedy strategies have been proposed for solving problem $(P_0^q)$ [27, 28]. Among them, orthogonal matching pursuit (OMP) [28] is perhaps the most commonly used strategy. This greedy algorithm guarantees convergence to the projection of $\mathbf{x}$ into the span of the dictionary atoms, in no more than $q$ iterations. Figure 2 shows an example of the values of a particular coefficient $a_{j^*}$ associated to the atom $\phi_{j^*}$ obtained by applying the OMP algorithm for a large number (almost half a million) of segments of $SaO_2$ signals and its corresponding activation histogram.

Although preconstructed dictionaries, such as the well-known wavelet packets [29], typically lead to fast sparse coding, they are almost always restricted to certain classes of signals. It is mainly for this reason that new approaches introducing data-driven dictionary learning techniques emerged. A Dictionary Learning $(DL)$ problem consists of simultaneously finding a dictionary $\Phi$ and representations of $n$ signals $\mathbf{x}_i$, $1 \leq i \leq n$, (in terms of atoms of such a dictionary) complying with a sparsity constraint for each one of the $n$ signals, while minimizing the total representation error. The $(DL)$ problem associated to the data: $q, M, N \in \mathbb{N}$, $M \geq N$, and $n$ signals in $\mathbb{R}^N$, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, can be formally written as

$$(DL): \min_{\substack{\Phi \in \mathbb{R}^{N\times M} \\ \mathbf{a}_i \in \mathbb{R}^M,\ \|\mathbf{a}_i\|_0 \leq q,\ 1\leq i\leq n.}} \sum_{i=1}^{n} \|\mathbf{x}_i - \Phi\mathbf{a}_i\|_2 \tag{3}$$

The first data-based dictionary learning algorithms were originally developed almost three decades ago [30–32]. Some of them have their roots in probabilistic frameworks by considering the observed data as realizations of certain random variables [30, 31]. In [31] for example, the authors developed an algorithm for finding a redundant dictionary that maximizes the likelihood function of the probability distribution of the data. In that work, an analytic expression for the likelihood function was derived by approximating the posterior distribution by Gaussian functions. An iterative approach for dictionary learning, known as the "method for optimal directions" (MOD), was presented in [32]. The sparse coding stage of this method makes use of the OMP algorithm followed by a simple dictionary updating rule. A new iterative algorithm was recently proposed by Aharon et al. in [14]. This new approach, called "K singular value decompositions" (K-SVD), consists mainly of two stages: a sparse coding stage and a dictionary learning stage. The OMP algorithm is used for the sparse coding stage, which is followed by a dictionary updating step where the atoms are updated one at a time and the representation coefficients are allowed to change in order to minimize the total representation error.

## 4. Discriminative Subdictionary Construction

Although data-driven dictionary learning algorithms produce sparse representations of signals which are robust against noise and missing data, such representations turn out to be unsuitable if the final objective is signal classification. This is mainly so because those algorithms do not take into account any a priori or available information concerning class membership. In order to overcome this difficulty, some strategies which incorporate appropriate class information have been proposed [4, 16, 33]. In [33], for instance, the authors developed a discriminative dictionary learning method by efficiently integrating a single predictive linear classifier into the cost function of the K-SVD algorithm. A method incorporating a discriminative term into the cost function of the standard K-SVD algorithm is presented in [16]. This method finds an optimal dictionary which is simultaneously representative and discriminative for face recognition tasks. In this work, we make use of a simple approach for detecting discriminative atoms from a previously learned dictionary and using them to build a new subdictionary. This approach, which is originally presented in [4], consists of solving two problems, namely, (i) the above mentioned full $(DL)$ problem and (ii) a discriminative subdictionary $(DSD)$ construction problem. We shall now proceed to describe problem (iii). One way to obtain discriminative subdictionaries consists of maximizing an appropriate discriminative value functional $G(\cdot)$. Given a data matrix $\mathbf{X} \in \mathbb{R}^{N\times n}$, a class label vector $\mathbf{c} \in \mathscr{C}^n$ (where $\mathscr{C}$ is the set of all classes; in the binary case $\mathscr{C} = \{c_1, c_2\}$), a dictionary $\Phi \in \mathbb{R}^{N\times M}$ and $p \in \mathbb{N}$ (with $p < M$), the most discriminative subdictionary $\widehat{\Phi}^{\mathbf{d}} \in \mathbb{R}^{N\times p}$, according to an appropriate prescribed discriminative value functional $G_{\mathbf{X},\mathbf{c},\Phi} : \mathbb{R}^{N\times p} \to \mathbb{R}_0^+$, is defined as

$$(DSD)\colon \widehat{\Phi}^{\mathbf{d}} = \underset{\substack{\mathbf{d} \doteq [i_1 i_2 \cdots i_p] \\ i_j \in \{1,2,\ldots,M\} \\ i_j \neq i_k \forall j \neq k,}}{\arg\max} G_{\mathbf{X},\mathbf{c},\Phi}\left(\Phi^{\mathbf{d}}\right) \tag{4}$$

where for $\mathbf{d} \doteq [i_1\ i_2\ \cdots\ i_p]$, $\Phi^{\mathbf{d}}$ denotes the $N \times p$ matrix whose $j$th column is the $i_j$th column of $\Phi$. The function $G$, which must be provided, quantifies the discriminative power of each subdictionary $\Phi^{\mathbf{d}}$. Thus, large values of $G$ correspond to highly discriminative subdictionaries while small values of $G$ are associated to subdictionaries with low discriminability.

Several questions concerning problem $(DSD)$ clearly emerge. Among them are the following: (i) how do we find an appropriate discriminative value function $G$?, (ii) given the functional $G$, does problem $(DSD)$ have a solution?, (iii) if it does, is it unique?, (iv) in the case of nonuniqueness, how do we decide which subdictionary, among the optimizers, is the best for our classification purposes? and (v) how difficult is it, in terms of computational cost, to solve problem $(DSD)$?. Although this problem has not been extensively studied, is it known that solving $(DSD)$ is computationally very challenging for $p > 1$, mainly due to the combinatorial explosion problem. A way to overcome the computational complexities entailed by problem $(DSD)$ consists of defining an appropriate discriminative value functional $G$ for $p = 1$. In that way $G$ is independently evaluated at each one of the atoms (columns) of $\Phi$ and the discriminative subdictionary $\Phi^{\mathbf{d}} \in \mathbb{R}^{N\times p^*}$ is constructed by stacking side-by-side the first $p^*$ ranked columns of $\Phi$ with largest $G$ values. This simplification is based on the assumption that each atom in the dictionary is used to model specific characteristics that are not completely modeled by the other atoms. Thus, the discriminative information provided by a particular atom will be different from the information contributed by other atoms.

## 5. Discriminative Value Functions for Atom Selection

Several ways for appropriately constructing discriminative value functions $G$ exists. In this section, we present two different approaches to define such a function, namely, (i) using traditional discrepancy measures and (ii) using a new discriminative measure to which we shall refer as the "Difference of Conditional Activation Frequency" (DCAF). We shall previously need to introduce an appropriate setting and terminology regarding probability density functions (PDFs) in the context of sparse representations for signal classification.

Here, and in the sequel, we shall consider the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ as realizations of a particular random vector $\mathbf{X}$. Any sparse representation of those vectors will result in the PDFs of each coefficient $a_j$ (associated to the atom $\phi_j$) showing a very concentrated peak at zero with heavy tails (as depicted in Figure 2). In the context of binary signal classification, it is reasonable to think that if a given atom $\phi_{j^*}$ is highly discriminative, then the conditional PDFs $\pi(a_{j^*} \mid c_1)$ and $\pi(a_{j^*} \mid c_2)$ will be significantly different. Thus, if a dictionary $\Phi$ is poorly discriminative, then one should expect $\pi(a_j \mid c_1) \approx \pi(a_j \mid c_2)$ for all $j$.

Although the elements $a_j$ of the representation vector $\mathbf{a}$ are in general real numbers, for practical reasons, it is appropriate to discretize them. That can be done in the usual way by partitioning the real line $\mathbb{R}$ into intervals $I_k \doteq ((k-1/2)\Delta, (k+1/2)\Delta]$, $k \in \mathbb{Z}$, of length $\Delta$ and the associated discretized random variable $\mathscr{K}_j \doteq \sum_{k\in\mathbb{Z}} k\, \chi_{I_k}(a_j)$. The corresponding probability mass function (PMF) is $p_{\mathscr{K}_j}(k) = P(a_j \in I_k) = \int_{I_k} \pi(a_j) da_j$, $k \in \mathbb{Z}$. Figure 3 shows the estimated PMF and the corresponding conditional PMFs (given each one of the two classes), both for a nondiscriminative and a discriminative atom using $SaO_2$ signals.

We shall now proceed to define how we compute the discriminative value function $G$. Given the data matrix $\mathbf{X} \in \mathbb{R}^{N\times n}$, the corresponding class label vector $\mathbf{c} \in \mathscr{C}^n$ and a full dictionary $\Phi \in \mathbb{R}^{N\times M}$, the first step consists of obtaining the sparse matrix $\mathbf{A} \doteq [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n] \in \mathbb{R}^{M\times n}$ by applying the OMP algorithm. The $j$th row of this sparse matrix is then used for estimating the conditional PMFs $p_{\mathscr{K}_j}(\cdot|c_1)$ and $p_{\mathscr{K}_j}(\cdot|c_2)$. Finally, the value of $G$ at the atom $\phi_j$ is computed as the

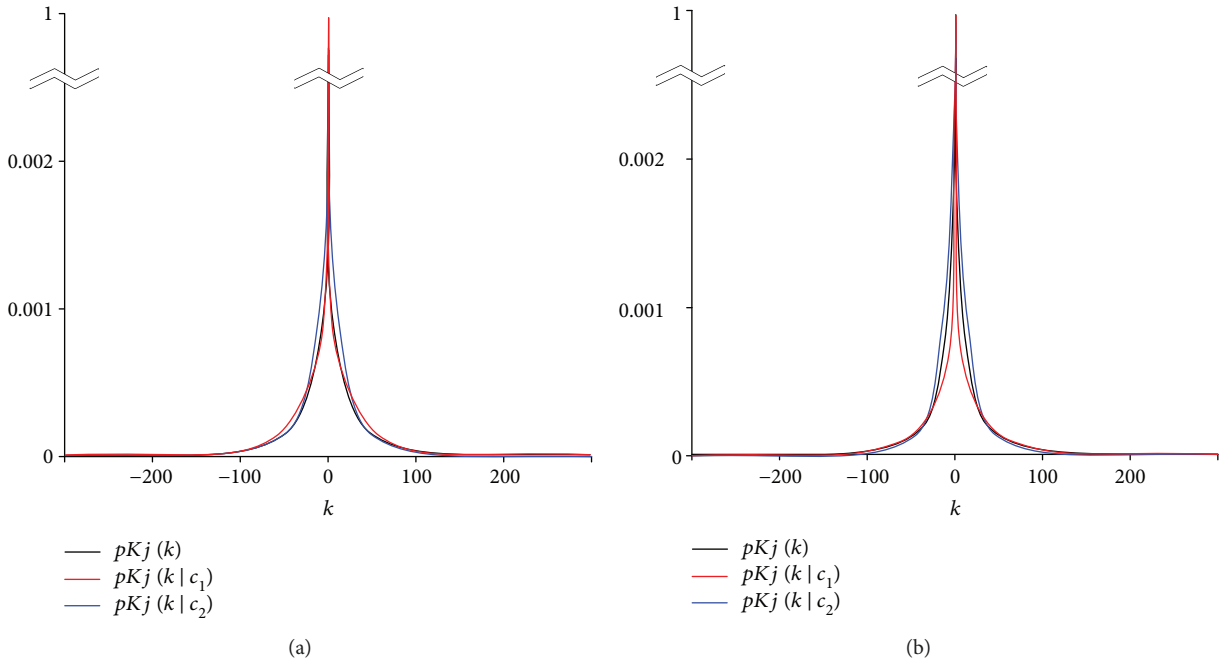


Figure 3: Estimated probability mass functions for a nondiscriminative atom $\phi_j$ (a) and a discriminative one (b).

discrepancy (as quantified by an appropriate discrepancy measure) between these two PMFs. In what follows, we introduce the discrepancy measures that we shall use in this work.

*5.1. Traditional Discrepancy Measures.* A great diversity of measures whose purpose is performing comparisons between probability distributions exists [34]. In this work, the best known and more commonly used ones are compared in terms of their performance for selecting the most discriminative atoms in a dictionary. The KL, J, and JS divergence measures were utilized, along with the Fisher score (F).

The KL divergence [7] is probably the most widely used information "distance" measure from a theoretical framework, and it was successfully applied in numerous problems for signal classification [1, 35, 36]. To compare the two conditional PMFs associated with the activation of the $j$th atom, the KL distance was used as follows:

$$\mathrm{KL}\left(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)\right) \doteq \sum_{k\in\mathbb{Z}} p_{\mathscr{K}_j}(k \mid c_1) \log\left(\frac{p_{\mathscr{K}_j}(k \mid c_1)}{p_{\mathscr{K}_j}(k \mid c_2)}\right), \tag{5}$$

assuming that $0 \log(0) \doteq 0$.

Despite the computational and theoretical properties provided by KL distance, what usually becomes a trouble in many problems of signal classification is its lack of symmetry. It can be easily seen that altering the order of the arguments in (5) can change the output value. To solve this issue, a symmetric version of the KL distance can be used such as the J divergence [9], which, even though was not initially created as a symmetric version of the KL distance, is the sum of the two possible KL distances between probability distributions. In this article, the J divergence is defined as follows:

$$\begin{aligned}\mathrm{J}\left(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)\right) \doteq{}& \mathrm{KL}\left(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)\right) \\ &+ \mathrm{KL}\left(p_{\mathscr{K}_j}(\cdot|c_2), p_{\mathscr{K}_j}(\cdot|c_1)\right).\end{aligned} \tag{6}$$

Another symmetric smoothed version of the KL distance is the JS divergence [10]. For the problem of comparing the two conditional probabilities associated to each class it is defined as

$$\begin{aligned}\mathrm{JS}\left(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)\right) \doteq{}& w_1 \mathrm{KL}\left(p_{\mathscr{K}_j}(\cdot|c_1), q_{\mathscr{K}_j}(\cdot)\right) \\ &+ w_2 \mathrm{KL}\left(p_{\mathscr{K}_j}(\cdot|c_2), q_{\mathscr{K}_j}(\cdot)\right),\end{aligned} \tag{7}$$

where $q_{\mathscr{K}_j}(\cdot) = w_1 p_{\mathscr{K}_j}(\cdot|c_1) + w_2 p_{\mathscr{K}_j}(\cdot|c_2)$ and $w_1$ and $w_2$ are the weights associated to each of the conditional PMFs, with $w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$. An interesting feature of the JS distance is the fact that different values of weights ($w_1$ and $w_2$) can be assigned to the probability distributions according to their importance. In this work, $w_1 = P(c_1)$ and $w_2 = P(c_2)$, that is, the weights are associated with the a priori probabilities of the classes. Note that computing the JS distance as defined here is the same as computing the mutual information between the class and the activations, that is, $\mathrm{JS}(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)) = \mathrm{MI}(\mathscr{K}_j, \mathscr{C})$.

Within signal classification problems, F is a measure which has been extensively used. Unlike the other measures presented here, that require estimations of the conditional PMFs, F uses just two parameters of the distributions (the means and standard deviations). This makes this measure

much less expensive computationally speaking, but implicitly assumes certain characteristics of the distribution under study (i.e., second-order characteristics). In the case of univariate binary problem at hand, the F can be defined as

$$\mathrm{F}\left(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)\right) \doteq \frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}, \tag{8}$$

where $\mu_\ell$ and $\sigma_\ell^2$ are the mean and standard deviation of $p_{\mathscr{K}_j}(\cdot|c_\ell)$ [37].

Although the abovementioned discrepancy measures provide, in a certain sense, "measures" of distance between two probability distribution functions, most of them (such as the KL divergence and those symmetric variants) are not strictly a metric. For instance, the KL divergence is a nonsymmetric discrepancy measure where the triangular inequality is not satisfied. Nevertheless, $\mathrm{KL}(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2))$ is a nonnegative measure, that is, $\mathrm{KL}(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)) \geq 0$ and $\mathrm{KL}(p_{\mathscr{K}_j}(\cdot|c_1), p_{\mathscr{K}_j}(\cdot|c_2)) = 0$ if and only if $p_{\mathscr{K}_j}(\cdot|c_1) = p_{\mathscr{K}_j}(\cdot|c_2)$.

*5.2. Difference of Conditional Activation Frequency.* In a previous work, a method called Most Discriminative Column Selection (MDCS) for the construction of a discriminative subdictionary was originally presented [4]. The sparse representations of the signals in terms of subdictionaries constructed using MDCS provided good performance in the detection of apnea-hypopnea events. In the mentioned work, the most discriminative atoms were identified by comparing the difference of conditional activation frequency.

The candidates to be considered as "most discriminative" according to [4] are those atoms with higher absolute difference between conditional activation probabilities given the class. That is, an atom is considered as highly discriminative if it is active, in proportion, more times for one of the classes. The use of this approach as a measure of discriminative power follows from the idea that one of the most expressive parameters regarding the importance of a given atom is its activation probability. Moreover, if certain atoms are active mostly for a given class, then it is assumed they represent features of importance in the description of that particular class.

Following this idea, DCAF is defined as

$$\mathrm{DCAF}\left(\eta_1^j, \eta_2^j\right) \doteq \left|\eta_1^j - \eta_2^j\right|, \tag{9}$$

where

$$\eta_\ell^j \doteq \frac{\text{number of activations of the } j\text{th atom for } c_\ell}{\text{number of } c_\ell \text{samples}}. \tag{10}$$

The measure defined in (9) is symmetric; its value is always ≥0 and is inexpensive in terms of computing (if the classes are balanced, the DCAF can be replaced just by simply counting, without the necessity of dividing with the number of samples).

It can easily be seen that the definition of $\eta_\ell^j$ in (10) is equal to the maximum likelihood estimation of the conditional probability of activation, that is,

$$p_{\mathscr{K}_j}(k \neq 0 \mid c_\ell) \approx \eta_\ell^j. \tag{11}$$

Replacing this expression in (9), we can write

$$\begin{aligned}\mathrm{DCAF}\left(\eta_1^j, \eta_2^j\right) &\approx \left|p_{\mathscr{K}_j}(k \neq 0 \mid c_1) - p_{\mathscr{K}_j}(k \neq 0 \mid c_2)\right| \\ &\approx \left|\left(1 - p_{\mathscr{K}_j}(k = 0 \mid c_1)\right) - \left(1 - p_{\mathscr{K}_j}(k = 0 \mid c_2)\right)\right| \\ &\approx \left|p_{\mathscr{K}_j}(k = 0 \mid c_2) - p_{\mathscr{K}_j}(k = 0 \mid c_1)\right|,\end{aligned} \tag{12}$$

finally expressing the DCAF in terms of the complementary conditional probabilities that the atoms will not be activated. With the exception of the F, all the measures presented in Section 5.1 can be expressed as summations, where only one of the terms is computed using the probabilities that $k = 0$. However, due to the high sparsity of the representations the terms associated with $k = 0$ are particularly important. This fact allows us to expect some correlation between the results obtained with the different discrepancy measures and the DCAF.

Figure 4 shows a representation of the conditional PMFs associated to the activations of two different atoms (left side) as well as an illustration of such functions where the peaks centered at zero ($k = 0$) were discarded (middle). It is important to note that, when excluding the zero-centered peak from the graphic, a significant reduction in the magnitude of the $y$-axis scale is produced which highlights the importance of the activation probability of sparse representations. However, the discrepancy between the distributions is not only due to the atoms activation probability, since slight differences between the probability values for all $k \neq 0$ exist (zoom-in region). Additionally, the absolute values of these differences are represented by the gray regions. It is also important to point out that these area values shown in gray ($\sum_{k\neq 0} |p_{\mathscr{K}_\ell}(k \mid c_1) - p_{\mathscr{K}_\ell}(k \mid c_2)|$) are not necessarily equal to those corresponding to the DCAF values. Nevertheless, for symmetric PMFs with high kurtosis and heavy tails (such is the case of the PMFs used in this work), the conditional and a priori distributions are similar and therefore both area values are close to each other.

## 6. Experimental Setup

This section presents the proposed system and its configuration settings, aimed at detecting patients suspected of suffering from moderate to severe OSAH syndrome. It also describes the database used for training and testing the method along with the measures selected for assessing its performance.

The main objective of our research is to explore the effect of using discrepancy measures to rank the atoms according to their discriminative power. Also, the experiments are designed to determine the effect of using dictionaries with different degrees of overcompleteness (redundant dictionaries) for the detection of apnea-hypopnea events. Additionally, the performance of the system for different sizes of subdictionaries and sparsity degrees is analyzed.

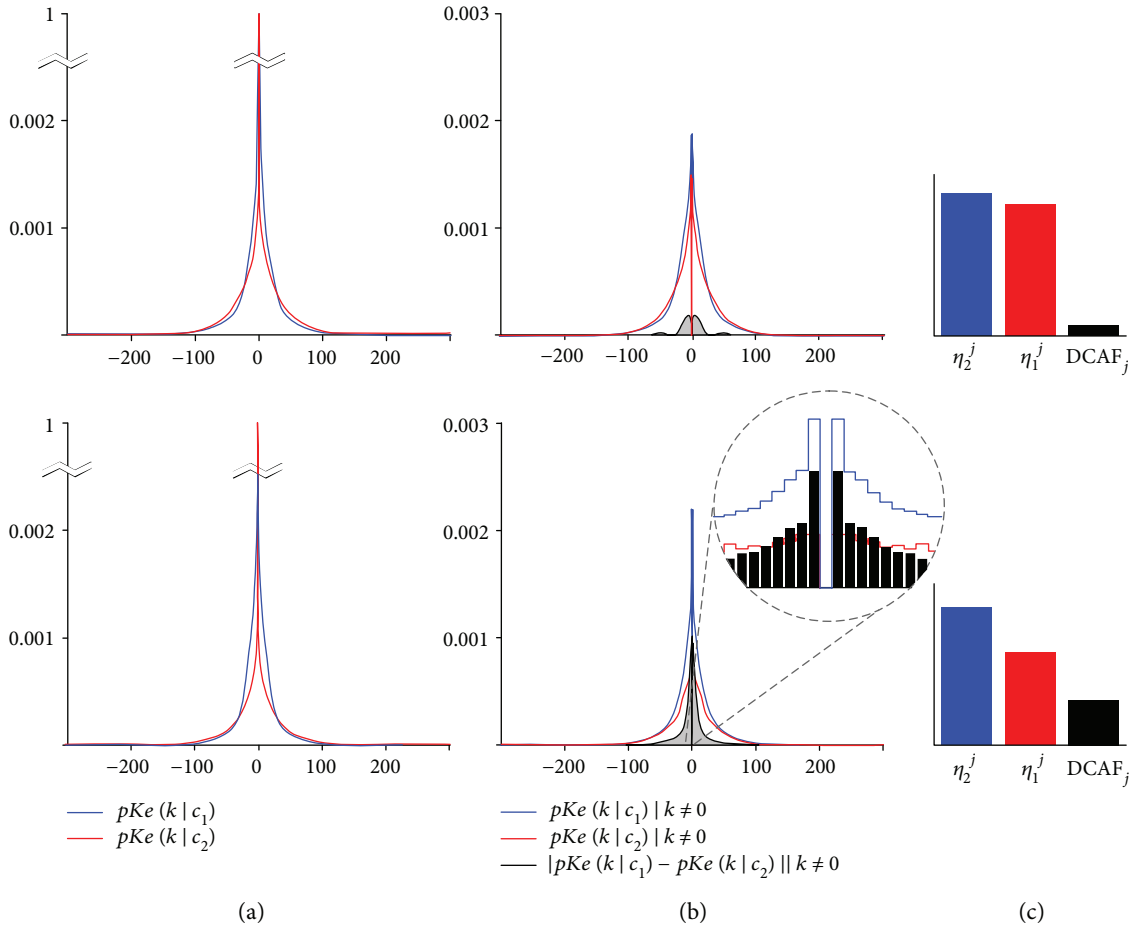


FIGURE 4: A representation of the conditional PMFs corresponding to the activations of two different atoms (a), the same functions excluding the peaks centered at zero ($k = 0$) and the absolute value of their differences (b), and a graphical interpretation of the DCAF (c). The top row corresponds to a nondiscriminative atom ($\phi_j$) while the bottom row corresponds to a discriminative one ($\phi_i$).

Figure 5 shows a simplified block diagram of the presented system. It can be observed that our system comprises a training phase (above) and a testing phase (below). To clarify the system's description, we divided it into three different stages, namely, stage I, stage II, and stage III. It can be seen that stages I and II are included into training and testing phases while stage III is only used during testing. Stage I is composed by a preprocessing block whose inputs are the raw $SaO_2$ signals, and its outputs are filtered segments of such signals, as described in Section 6.1. At the training phase, stage II receives segmented signals and finds an optimal discriminative subdictionary. During the testing phase, stage II obtains a sparse matrix in terms of the previously found subdictionary. These processes are thoroughly described in Section 6.2. Finally, the obtained sparse codes are used as input of stage III. This stage detects apnea-hypopnea events and estimates the AHI value, as described in Section 6.3.

*6.1. Database and Signal's Preprocessing.* The Sleep Heart Health Study (SHHS) dataset [38, 39] was originally designed to study correlations between sleep-disordered breathing and cardiovascular diseases. This dataset includes a large number of PSG studies, each of them containing several physiological signals such as EEG, ECG, nasal airflow and $SaO_2$. Medical expert annotations of sleep stages, arousals, and apnea-hypopnea events are also provided. In this work, only the S $aO_2$ signal (sampled at 1 Hz) and its corresponding apnea-hypopnea labels are considered for performing the experiments. In this article, the first online version of such a database (SHHS-2) is used. This version of the database contains a total of 995 freely available PSG studies (https://physionet.org/physiobank/).

The $SaO_2$ signals are mainly degraded by patient movements, baseline wander, disconnections, and the limited resolution of pulse oximeters, among other factors. When a disconnection occurs, the recording during the time interval where the sensor signal is blocked is lost. In order to overcome this inconvenience, the values of blood oxygen saturation during such an interval are linearly interpolated. To denoise the signals, a wavelet processing technique [40] is used. The denoising process is performed by zeroing the approximation coefficients at level 8, as well as the coefficients of the first three detail levels of the discrete dyadic wavelet transform with mother wavelet Daubechies 2. The signals are then synthesized using the modified wavelet

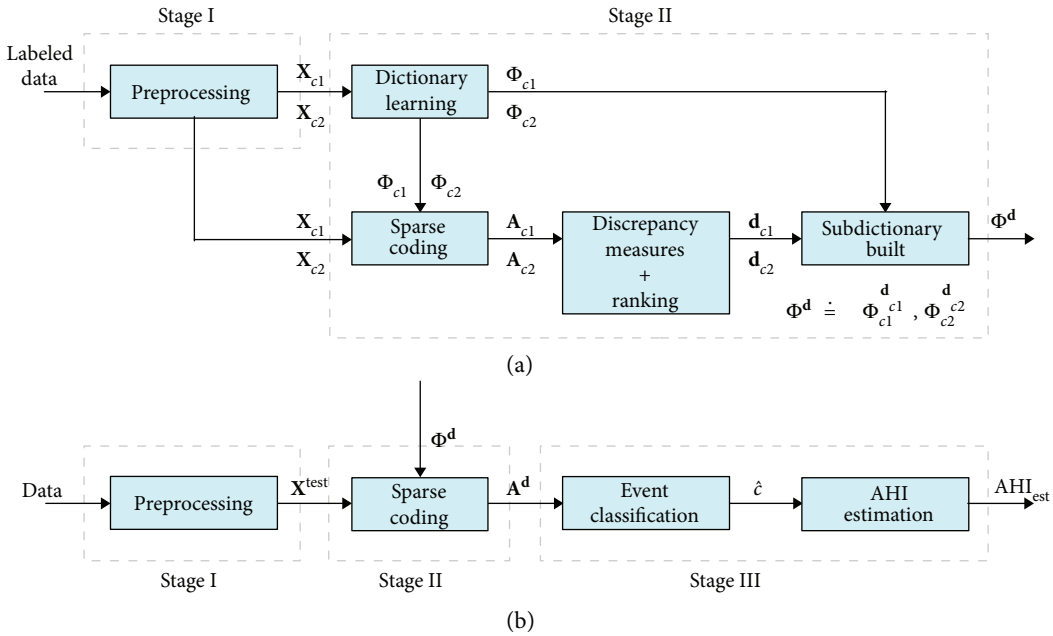


FIGURE 5: Block diagram of the proposed system during training (a) and testing (b).

coefficients by inverse discrete dyadic wavelet transform. The application of this wavelet decomposition technique has the effect of a band-pass filter where the baseline wander and both the low-frequency noise and the high-frequency noise, as well as the quantization noise are eliminated. Figure 6 shows a small fragment of the original raw $SaO_2$ signal (top) and its wavelet-filtered version (bottom). Labels of apnea-hypopnea events (dashed lines) introduced by the medical experts are also added. These labels were generated by medical experts using the airflow information and thus are not aligned to the desaturations, that is, there is a variable delay between the start time of an event and the corresponding desaturation.

The application of the sparse representation technique requires an appropriate segmentation of the signals. Segments of length $N = 128$ (corresponding to 128 seconds of the signal recording) with a 75% overlapping between two consecutive segments are taken. It is appropriate to point out that although several overlapping percentages were tested, the best system performances were yielded by a 75% overlapping. This redundancy prevents apnea-hypopnea events from being undetected. In this segmentation process, the time intervals where a disconnection occurs are discarded. The segments of pulse oximetry signals are then simultaneously arranged as column vectors $\mathbf{x}_i \in \mathbb{R}^N$ and labeled with ones ($c_1$) and minus ones ($c_2$), where a one corresponds to apnea-hypopnea events, and a minus one to the lack of it. Finally, a signal matrix $\mathbf{X}$ is built by stacking side-by-side the column vectors $\mathbf{x}_i$, that is, the signal matrix is defined as $\mathbf{X} \doteq [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]$.

As mentioned above, the entire dataset used in this work contains 995 complete studies, 41 of which were not taken into account for performing the experiments since the size of the signal vectors differs from the corresponding vector

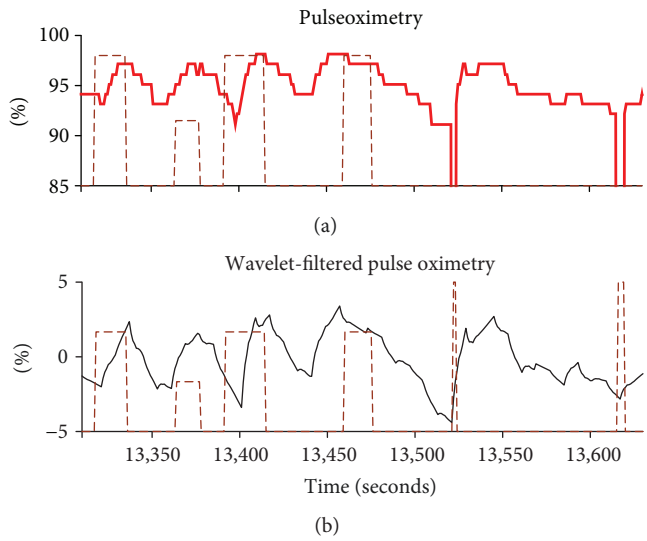


FIGURE 6: A small fragment of a pulse oximetry signal (a) and its wavelet-filtered version (b). Dashed lines represent labels of apnea-hypopnea events established by the medical expert.

of class labels. Among the remaining 954 studies, a subset of 667 (70%) studies were randomly selected and fixed for learning the dictionary and training the classifier. The remaining 287 (30%) studies were left out for the final test. The $SaO_2$ signals were filtered using wavelet filters and segmented as explained previously into column vectors of size 128. After performing the filtering and segmentation process, a signal matrix $\mathbf{X}^{\text{train}}$ of size $128 \times 455515$ is assembled by joining two previously constructed signal matrices, one for

each class, $\mathbf{X}^{\text{train}} \doteq [\mathbf{X}_{c_1}^{\text{train}}\ \mathbf{X}_{c_2}^{\text{train}}]$, which contain 183,163 and 272,352 segments, respectively. On the other hand, for each study included into the testing dataset, a testing matrix $\mathbf{X}^{\text{test}}$ is built.

*6.2. Sparse Coding and Subdictionary Construction.* In our experiments, the learning of the dictionaries is performed by using the traditional K-SVD method [14]. Optimized MATLAB codes for dictionary learning using K-SVD as well as for sparse coding using the OMP algorithm are freely available for academic and personal use at the Ron Rubinstein's personal web page (http://www.cs.technion.ac.il/~ronrubin/software.html). At the beginning, the atoms assigned to conform the initial dictionary are randomly selected from the input signal matrix for training without taking into account any information about the classes. If the signal's space dimension is fixed, which should be the effect of constructing dictionaries with different overcompleteness degree?. To answer this question, three types of dictionaries denoted by Φ1 of size $128 \times 128$, Φ2 of size $128 \times 256$, and Φ 4 of size $128 \times 512$, corresponding to redundancy factors of 1, 2, and 4, respectively, were built. First, the dictionary Φ1 was constructed by joining two subcomplete dictionaries of sizes $128 \times 64$ denoted by $\Phi 1_{c_1}$ and $\Phi 1_{c_2}$ learned using a large number of training segments (a total of 100,000 segments for each of the classes) belonging to the classes $c_1$ and $c_2$, respectively. Following the same idea, redundant dictionaries denoted by Φ2 (256 atoms) and Φ4 (512 atoms) were appropriately built. At the dictionary learning stage, the number of nonzero elements was selected and fixed as a percentage value of 12.5 of the atoms conforming the dictionary. Also, a total of 30 iterations of the K-SVD algorithm were performed.

Once the dictionary has already been trained, the sparse representation vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ corresponding to the input signals $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are obtained by applying the OMP algorithm. In such a procedure, the nearest integer number to a percentage value of 12.5 of $M$ is selected and fixed. The reason for having chosen this percentage value is because it presented the best trade-off between representativity and discriminability of the segments. Thus, sparsity values of $q = 16$, $q = 32$ and $q = 64$ are selected to represent the input signals for training in terms of the full dictionaries Φ1, Φ2 and Φ4, respectively.

Histograms are typically used to approximate data distributions. In this work, we make use of histograms of the atom's activations to approximate the PDFs. The discretization process was performed by using a Δ value of 0.5. The detection of the most discriminative atoms is obtained by maximizing the discrepancy between the conditional PMFs of the atom's activations given the classes. This objective is achieved using the proposed DCAF measure as well as those denoted by KL, J, JS, and F. The application of different discrepancy measures to the sparse vectors allows for the selection of different "discriminative atoms," which implies the construction of discriminative subdictionaries which are essentially different. The construction of subdictionaries, here denoted by $\Phi 1^{\mathbf{d}}$, $\Phi 2^{\mathbf{d}}$ and $\Phi 4^{\mathbf{d}}$, is performed by selecting atoms from Φ1, Φ2, and Φ4, respectively. Once the most discriminative atoms are detected, the subdictionary is built and consequently the feature vectors are obtained by applying the OMP algorithm. Finally, each feature vector is assigned to be the input of the ELM classifier.

*6.3. Event Detection and AHI Estimation.* Multilayer perceptron (MLP) neural networks trained for signal classification have proved to be a tool which provides quite good performances for OSAH syndrome detection [4]; however, the process of training this class of neural network becomes very costly mainly in terms of time. For this reason, in this work, we propose the use of extreme learning machine (ELM) [41] which is a type of single-hidden layer feedforward neural networks (SLFNs), instead of using MLP neural networks. Theoretically, this algorithm (ELM) results in providing a good generalization performance at extremely fast learning speed. The experimental results based on a few artificial and real benchmark function approximation and classification problems including large complex applications show that ELM can produce good generalization performance in most cases and can learn thousands times faster than conventional popular learning algorithms for feedforward neural networks [42].

Basic ELM classifier's MATLAB codes are available for download on the Guang-Bin Huang's web page (http://www.ntu.edu.sg/home/egbhuang/elm_codes.html). To train such a classifier, the main parameters to be fixed are the number of neurons in the hidden layer as well as the activation function of the neurons. In our experiments, the number of neurons in the hidden layer of the ELM corresponds to four times the feature vector dimension. Also, the well-known sigmoid activation function, which is the most common activation function in the nodes of the hidden and/or output layer, is chosen.

In order to evaluate the performance of the proposed classifier in the detection of individual apnea-hypopnea events (a local approach), or more specifically, in the identification of persons suspected of suffering from moderate to severe OSAH syndrome (a global approach), three performance measures are used. For the identification of single segments containing apnea-hypopnea events, the sensitivity ($\text{SE}_{\text{AH}}$) represents the total number of correctly classified segments of signals for which any apnea-hypopnea event occurred. Following the same idea, for the detection of individual segments of signals "not containing" any apnea-hypopnea event, the specificity ($\text{SP}_{\text{AH}}$) is defined as the total number of correctly classified segments for which any apnea-hypopnea is not present. The accuracy ($\text{AC}_{\text{AH}}$) is finally defined as follows:

$$\text{AC}_{\text{AH}} \doteq \frac{1}{n}\sum_{i=1}^{n} \delta(c_i, \widehat{c}_i), \tag{13}$$

where $n$ represents the total number of segments, $c_i$ and $\widehat{c}_i$ denote the corresponding class label of the $i$th segment and the corresponding prediction of the classifier, respectively, and $\delta(x, y)$ represents the delta function whose output is true (one) if the condition $x = y$ is satisfied and false (zero) otherwise.

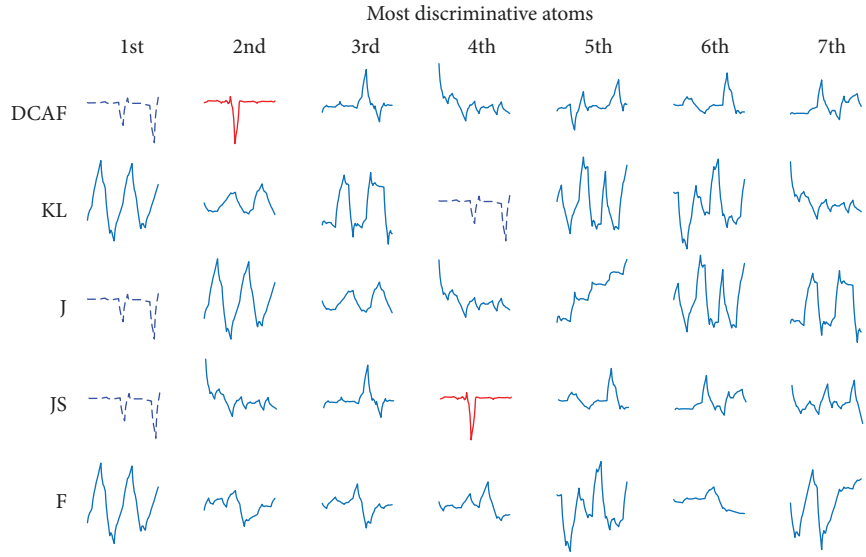


Figure 7: Waveforms corresponding to the first seven ranked atoms according to each one of the evaluated measures.

The differences in performance obtained for the event detection between each discrepancy measure were evaluated in order to test whether or not they are statistically significant. The test was performed assuming statistical independence of the classification errors for the different studies and approximating the error's binomial distribution by means of a normal distribution. This assumptions are reasonable due to the large number of $SaO_2$ signal segments available for each study (about 1100 segments per study, totaling 301,306 segments).

The estimated AHI index ($AHI_{est}$) is defined as the average number of predicted events per hour of study. This new index is used for OSAH syndrome detection. In this case, the sensitivity ($SE_{OSAH}$) is defined as the ratio of persons with OSAH syndrome for whom the final test is positive, and the specificity ($SP_{OSAH}$) is defined as the ratio of health patients for whom the final test is negative. Also, the area under the ROC curve (AUC) derived from a receiver operating characteristic (ROC) analysis [43] is used. A ROC analysis consists of computing the values of the sensitivity and specificity across all the possible detection threshold (DT) values. Then, the ROC curve is built by performing a plot of 1 – specificity versus sensitivity values. This curve has been widely used by medical physicians for evaluating diagnostic tests [44]. A comparison between two different methods can be effectively done by finding the "optimal" (in certain sense) cut-off point of the curve and evaluating their corresponding performances. Finally, the accuracy $AC_{OSAH}$ is defined as follows:

$$AC_{OSAH} \doteq \frac{1}{m}\sum_{i=1}^{m} \delta\left(AHI_{est}^{(i)} > DT, AHI^{(i)} > 15\right), \quad (14)$$

where $m$ corresponds to the total number of studies coming from the testing dataset and "DT" is the detection threshold value which adjusts overestimation of the events produced in the segmentation process. The value of DT results in the best cut-off point of the ROC curve. This point, which maximizes simultaneously sensitivity and specificity, corresponds to the minimum Euclidean distance ($d_{min}$) to the point (0,1) of the ROC curve.

## 7. Results and Discussion

In this section, results of the performed experiments are presented and discussed. This section is mainly separated into two subsections, namely, (i) the performance tuning section and (ii) the optimal system performance section.

*7.1. Performance Tuning.* This section presents results of the exploratory experiments performed to find optimal configurations of the proposed system. As explained in Section 6.2, three different full dictionaries called Φ1, Φ2, and Φ4 were learned by applying the standard K-SVD algorithm. In this process, it is expected that most dictionary atoms would capture high-frequency oscillations and normal respiration cycles in $SaO_2$ signals. It is important to point out however that typical desaturations in signals associated to apnea-hypopnea events should be encoded by some atoms. Secondly, the sparse matrices **A**1, **A**2, and **A**4 were obtained by applying the OMP algorithm. As described in Section 6.2, several measures were used to quantify the discriminative degree of individual atoms of each one of the studied dictionaries. Finally, the dictionary atoms were ranked in decreasing order of magnitude according to their discriminative power. Figure 7 shows the waveforms of the first seven ranked atoms of the dictionary Φ1 according to our measure (first row) as well as the first seven ranked atoms of such a dictionary according to all other discrepancy measures (rows from two to five). It can be seen that the most discriminative atom selected by DCAF (dashed waveform) provides information about two well-defined desaturations in the signal. It is also important to point out that this atom corresponds to the most discriminative one when using J divergence or eventually when using the JS divergence. Moreover, one can

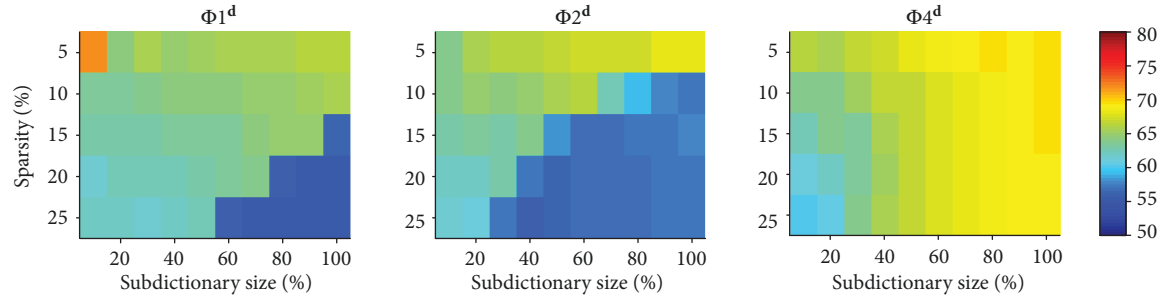


Figure 8: Averaged accuracy rates obtained by varying the percentages of the subdictionary size and the sparsity level according to a random ranking of the atoms.

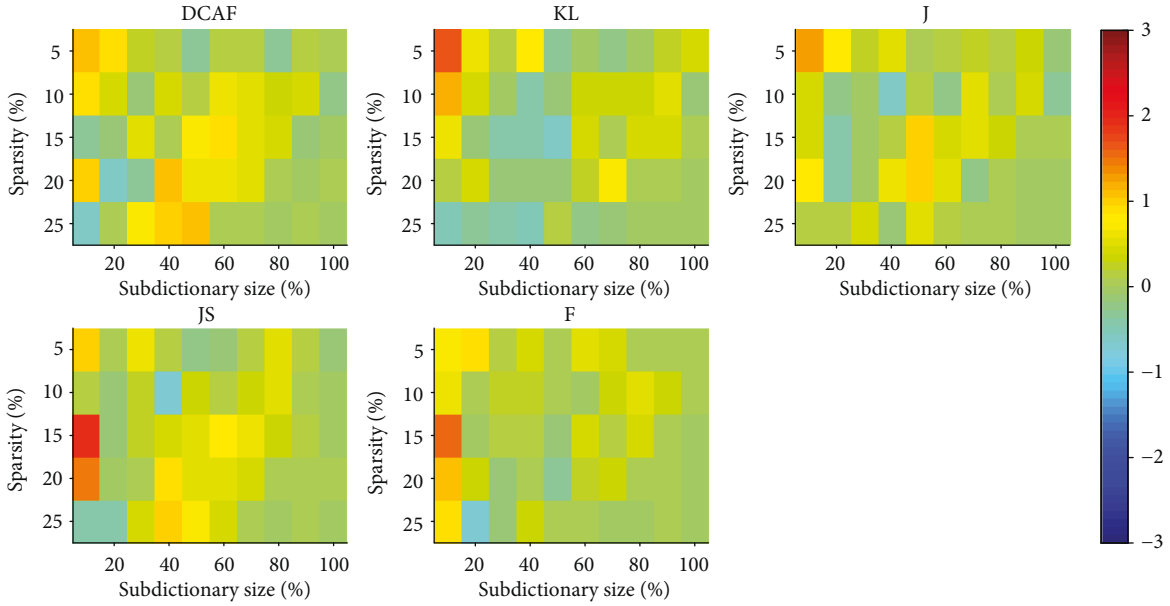


Figure 9: Five images representing differences between accuracy rates yielded by DCAF and all other discrepancy measures and random selection for Φ1.

clearly note that no highly discriminative atoms were taken when using Fisher score.

Discriminative subdictionaries called $\Phi 1^{\mathbf{d}}$, $\Phi 2^{\mathbf{d}}$, and $\Phi 4^{\mathbf{d}}$ were built by stacking side-by-side the first $p$ ranked atoms from Φ1, Φ2, and Φ4, respectively, according to their discriminative degree. It is appropriate to mention that the evaluation of several discrepancy measures leads to the construction of different discriminative subdictionaries. However, optimal values of $p$ (subdictionary size) and $q$ (sparsity level) are parameters that need to be tuned. In order to find optimal values of such hyperparameters, a grid search was performed.

The performance of our system was first tested by performing a Random Selection (RS) of the dictionary atoms. The involved results were fixed and appropriately used as reference. The random selection of the atoms was performed ten times. Additionally, for each one of the atoms' random selection, 60 iterations of the grid search were performed. Thus, the accuracy rate's variations introduced by the classifier were minimized. Figure 8 shows three images corresponding to averaged accuracy rates for each one of the evaluated dictionaries. Averaged accuracy rates (reference values) obtained by using the dictionary Φ1 for the detection of apnea-hypopnea events are shown on the left of this figure. It can be seen that sparse representations in terms of Φ1, using the smallest subdictionary size and the highest sparsity degree, result in better performance than the ones obtained by using all other configurations of Φ1 and the overcomplete dictionaries Φ2 and Φ4. In this way, two regions can be distinguished corresponding to a high-performance region and a low-performance one. The first one, which is or our interest, is yielded by simultaneously employing a small subdictionary size (10%) and a high sparsity degree (5%).

Next, DCAF and four other discrepancy measures were used for appropriately constructing discriminative subdictionaries. Then, a grid search of hyperparameters was performed by analyzing the performance that yields our system when using each one of the subdictionaries. Figure 9 shows five images corresponding to DCAF (upper left) and the other four discrepancy measures. These images represent the differences between accuracy rates obtained by using discriminative measures and the reference one (random

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R.E. Rolon, I. E. Gareis, L. Di Persia, R. Spies & H. L. Rufiner; "Complexity-based discrepancy measures applied to detection of apnea-hypopnea events"
Complexity, Vol. 2018, 2018.

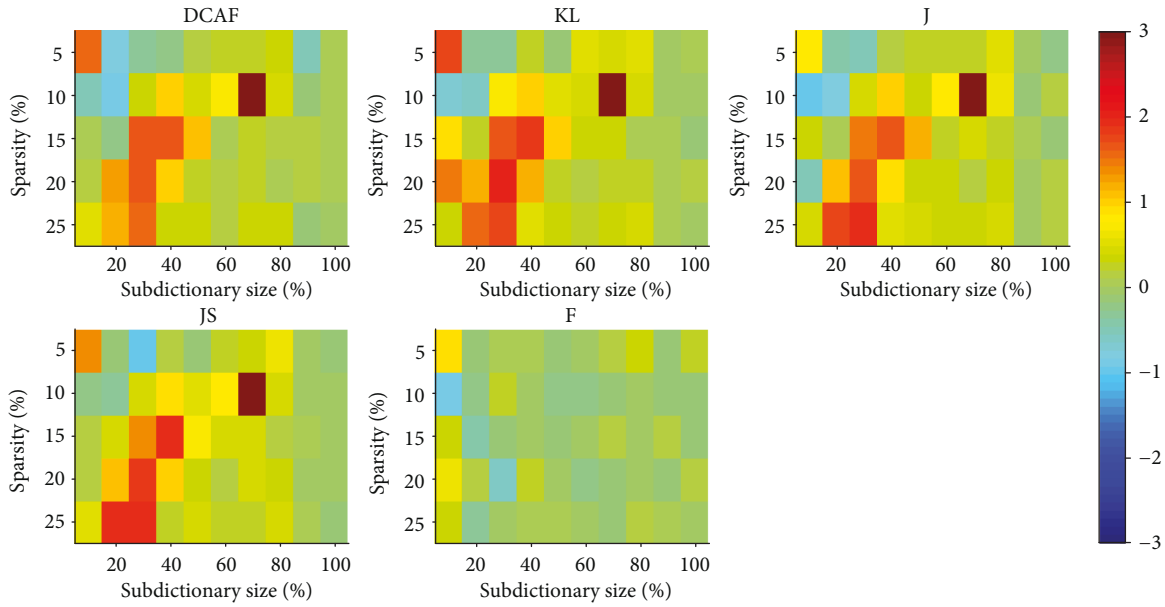


Figure 10: Five images representing differences between accuracy rates yielded by DCAF and all other discrepancy measures and random selection for Φ2.

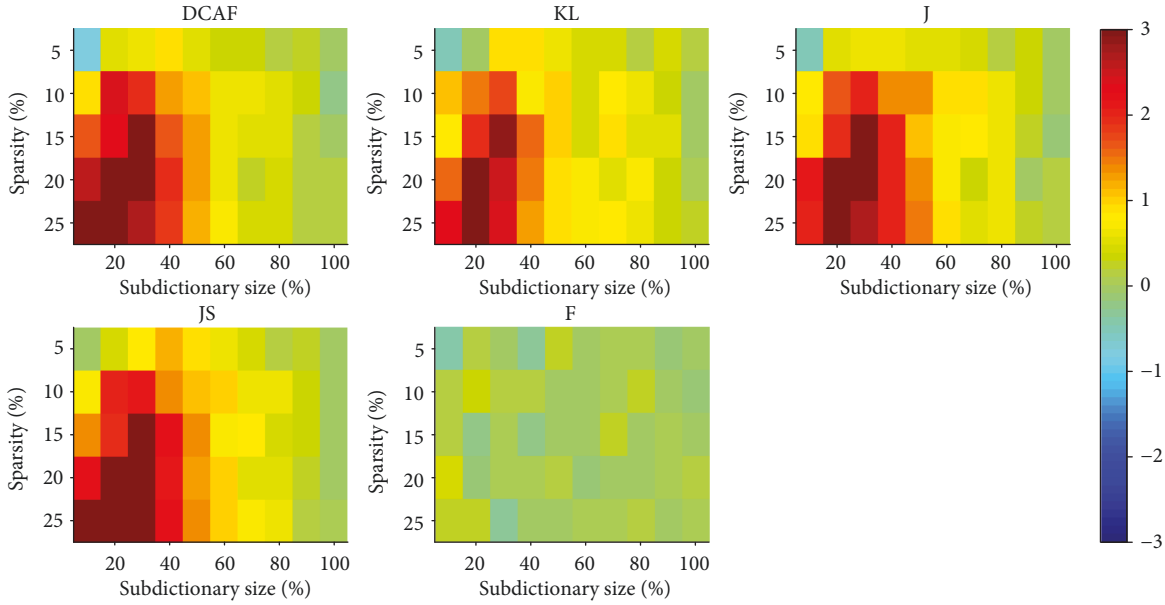


Figure 11: Five images representing differences between accuracy rates yielded by DCAF and all other discrepancy measures and random selection for Φ4.

selection) for Φ1. Also, each pixel of these images corresponds to particular percentages of subdictionary size and sparsity level. It can be observed that, independently of the discriminative measure, small percentages of subdictionary size yield good performances. It is appropriate to point out however that the effect of the dimension (subdictionary size) in the performance of the system is more important than the one induced by using discriminative measures.

Analogously, Figures 10 and 11 show five images which correspond to DCAF (upper left) and all other discrepancy measures. The images depicted in Figures 10 and 11 represent the differences between accuracy rates obtained by using these measures and the reference one for dictionaries Φ2 and Φ4, respectively.

If we compare the results shown in Figures 9–11, then it can be concluded that the proposed system presents the best performance, in terms of accuracy rate in the detection of apnea-hypopnea events, when using the full dictionary Φ1. Although similar results were obtained applying the proposed DCAF measure and those traditional ones (see Figure 9), it is important to point out that the use of discrepancy measures resulted in a significantly high

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R.E. Rolon, I. E. Gareis, L. Di Persia, R. Spies & H. L. Rufiner; "Complexity-based discrepancy measures applied to detection of apnea-hypopnea events"
Complexity, Vol. 2018, 2018.

Table 1: Averaged accuracy rates for subdictionary sizes of 10% regarding to each one of the evaluated full dictionaries.

| Measure | $\Phi 1^{\mathbf{d}}(128 \times 12)$ | | $\Phi 2^{\mathbf{d}}(128 \times 24)$ | | $\Phi 4^{\mathbf{d}}(128 \times 50)$ | |
|---|---|---|---|---|---|---|
| | Max | Avg. | Max | Avg. | Max | Avg. |
| DCAF | 72.62 | 64.68 | 65.20 | 63.15 | 65.19 | 64.21 |
| KL | 73.20 | 64.91 | 65.44 | 63.53 | 65.42 | 63.66 |
| J | 72.82 | 64.88 | 64.50 | 62.82 | 65.39 | 63.68 |
| JS | 72.55 | 64.10 | 65.02 | 63.18 | 65.87 | 64.01 |
| F | 72.23 | 65.21 | 64.57 | 63.04 | 65.64 | 62.71 |
| Full dictionary | 66.39 | 59.77 | 68.13 | 59.57 | 69.28 | 69.21 |

improvement with respect to a "random" selection of the atoms. As discussed above, the dimension reduction in the subdictionary size as well as high sparse levels yielded high accuracy rates. This is the reason for which a small subdictionary size (10%) and high sparse level (5%) were chosen to perform the final test.

System performance changes were analyzed by performing a comparison between averaged accuracy rates obtained by using discriminative subdictionaries and the ones obtained by using full dictionaries. Table 1 shows averaged accuracy percentages obtained by taken into account fixed discriminative subdictionary sizes (10%) while allowing the sparsity level to change (rows from 3 to 7). The last row of this table presents averaged accuracy percentages yielded by using full dictionaries for different sparsity levels. It can be observed that, in all of cases, discriminative subdictionaries outperform full dictionaries in the detection of apnea-hypopnea events.

The impact of sparsity degree in the performance of our system is illustrated in Table 2. These results were yielded by averaging accuracy rates obtained for a sparsity level of 5% and considering all possible subdictionary sizes (from 10% to 90%). For example, the second row shows averaged accuracy rates obtained by means of discriminative subdictionaries whose atoms were taken from $\Phi 1$, $\Phi 2$, and $\Phi 4$ by using DCAF measure.

*7.2. Optimal System Performance.* Optimal system configurations were selected and fixed to perform the final test. In the previous section, it was found that discriminative subdictionaries constructed by taken atoms from the dictionary $\Phi 1$ yield better performances than the ones constructed by selecting atoms from the dictionaries $\Phi 2$ and $\Phi 4$. Additionally, it was found that a discriminative subdictionary composed by only 12 atoms (10%) and a sparsity level of one (5%) yield in the best accuracy rate of our system.

In order to overcome the variance introduced by ELM predictors, 60 repetitions of the testing process were performed. Table 3 shows percentage values of minimum (Min), maximum (Max), average ($\mu$), and standard deviation ($\sigma$) corresponding to obtained accuracy rates in the detection of apnea-hypopnea events. Although, DCAF performs similarly to the four other discrepancy measures, its performance is achieved with a relatively low computational cost. Additionally, results show that performances obtained by using discriminative measures for constructing subdictionaries always outperform the ones yielded by making use of randomly constructed subdictionaries.

Table 2: Averaged accuracy rates by considering a sparsity level of 5% regarding to all possible subdictionary sizes.

| Measure | $\Phi 1$ | $\Phi 2$ | $\Phi 4$ |
|---|---|---|---|
| DCAF | 66.41 | 66.51 | 67.95 |
| KL | 66.49 | 66.72 | 67.98 |
| J | 66.60 | 66.56 | 67.98 |
| JS | 66.41 | 66.57 | 68.15 |
| F | 66.53 | 66.54 | 67.58 |

Table 3: Averaged accuracy rates for a subdictionary percentage of 10 for the detection of apnea-hypopnea events.

| Measure | Min | Max | $\mu$ | $\sigma$ |
|---|---|---|---|---|
| DCAF | 71.72 | 73.14 | 72.57 | 0.345 |
| Kullback-Leibler | 72.06 | 73.78 | 73.26 | 0.390 |
| Jeffrey | 71.77 | 73.31 | 72.66 | 0.319 |
| Jensen-Shannon | 71.79 | 73.11 | 72.55 | 0.295 |
| Fisher | 71.01 | 72.77 | 72.18 | 0.325 |
| Random Selection | 70.01 | 71.51 | 70.91 | 0.372 |

Table 4: A summary of the performed statistical significance tests.

| | RS | DCAF | KL | J | JS | F |
|---|---|---|---|---|---|---|
| RS | — | ✓ | ✓ | ✓ | ✓ | ✗ |
| DCAF | — | — | ✗ | ✗ | ✗ | ✗ |
| KL | — | — | — | ✗ | ✗ | ✗ |
| J | — | — | — | — | ✗ | ✗ |
| JS | — | — | — | — | — | ✗ |
| F | — | — | — | — | — | — |

Table 5: Maximum cut-off points for testing accuracy for a subdictionary percentage of 10 for the detection of apnea-hypopnea events.

| Measure | $d_{\min}$ | SE | SP | ACC | AUC |
|---|---|---|---|---|---|
| DCAF | 0.2211 | 81.88 | 87.32 | 84.60 | 0.9250 |
| Kullback-Leibler | 0.2242 | 81.46 | 87.39 | 84.43 | 0.9271 |
| Jeffrey | 0.2311 | 80.86 | 87.04 | 83.95 | 0.9283 |
| Jensen-Shannon | 0.2267 | 80.75 | 88.03 | 84.39 | 0.9244 |
| Fisher | 0.2280 | 80.66 | 87.91 | 84.29 | 0.9252 |

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R.E. Rolon, I. E. Gareis, L. Di Persia, R. Spies & H. L. Rufiner; "Complexity-based discrepancy measures applied to detection of apnea-hypopnea events"
Complexity, Vol. 2018, 2018.

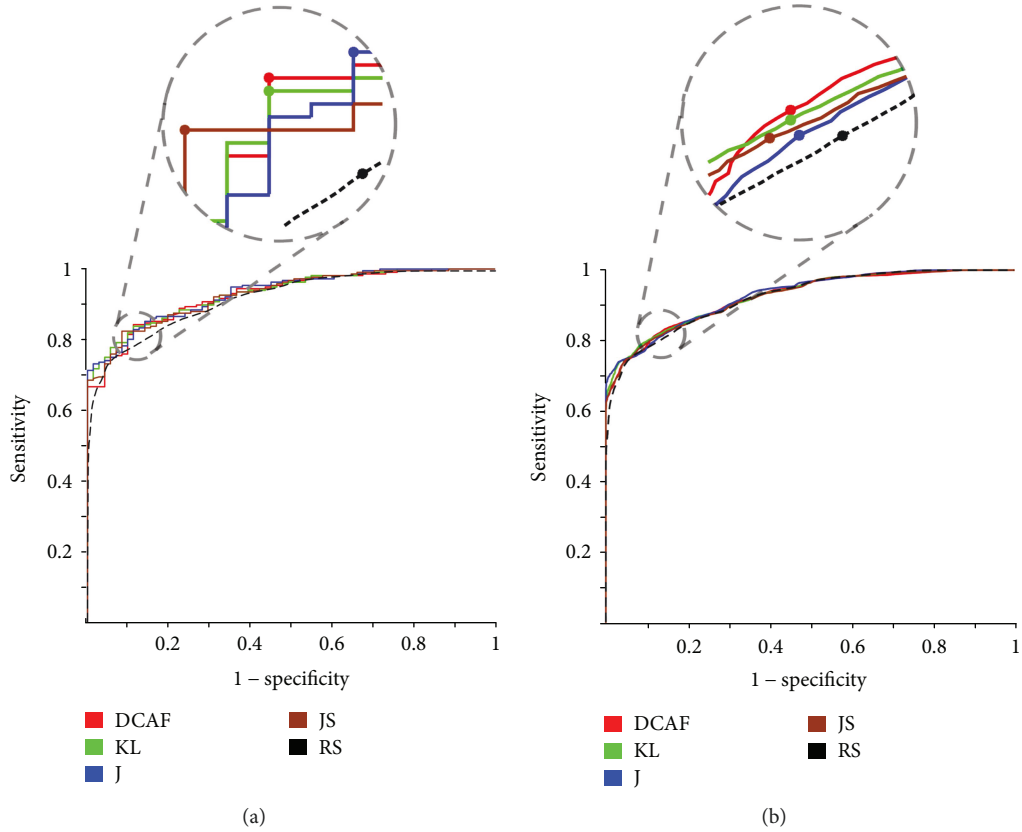


Figure 12: ROC curves corresponding to the performance measures described in Tables 5 and 6.

We have also evaluated the statistical significance of the results presented in Table 3 by computing the probability that using each one of the evaluated measures, including RS, yields in better classification performances than the others. In order to perform this test, we assumed the statistical independence of the classification errors for each study. Also, it was possible to approximate the error's binomial probability distribution by a normal distribution due to a wide availability of signals (301,306). Table 4 summarizes the results of the performed statistical significance tests by considering a $p$ value of 0.01. It can be seen that DCAF and three other discrepancy measures (KL, J, and JS divergences) are significantly different with respect to random selection. Also, no significant difference was found between F score and random selection. Additionally, it was found that DCAF does not perform significantly better than that of the KL, J, and JS divergences.

To determine the severity degree of OSAH syndrome, a ROC curve analysis was successfully performed by considering a detection AHI of 15. This index was selected in order to identify patients suspected of suffering from moderate to severe OSAH syndrome. Table 5 shows the minimum operating (cut-off) point of the ROC curves and maximum percentages of sensitivity, specificity, and accuracy as well as maximum values of area under the ROC curve for AHI diagnostic threshold values of 15 (Figure 12(a)). It can be seen that DCAF resulted in a maximum area under the ROC curve of 0.9250 and sensitivity and specificity percentages of 81.88 and 87.32, respectively. These are the maximum performance measures at which the minimum cut-off point of the ROC curve is attained. If we compare the performances attained between all of the evaluated measures, then the maximum SE and AUC value is yielded by J divergence. Also, JS divergence outperformed all the others in terms of ACC and DCAF resulted in the minimum cut-off point of the ROC curve.

Table 6: Averaged cut-off points for testing accuracy for a subdictionary percentage of 10 for the detection of apnea-hypopnea events.

| Measure | $d_{\min}$ | SE | SP | ACC | AUC |
|---|---|---|---|---|---|
| DCAF | 0.2211 | 81.88 | 87.32 | 84.60 | 0.9250 |
| Kullback-Leibler | 0.2242 | 81.46 | 87.39 | 84.43 | 0.9271 |
| Jeffrey | 0.2311 | 80.86 | 87.04 | 83.95 | 0.9283 |
| Jensen-Shannon | 0.2267 | 80.75 | 88.03 | 84.39 | 0.9244 |
| Fisher | 0.2280 | 80.66 | 87.91 | 84.29 | 0.9252 |
| Random Selection | 0.2396 | 80.85 | 85.60 | 83.23 | 0.9222 |

We additionally performed a ROC curve analysis of the averaged performances of DCAF and all the other discrepancy measures, including (RS) (Figure 12(b)). Additionally, Table 6 shows the minimum operating (cut-off) point of the averaged ROC curves as well as the maximum percentages of sensitivity, specificity, and accuracy, including the

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R.E. Rolon, I. E. Gareis, L. Di Persia, R. Spies & H. L. Rufiner; "Complexity-based discrepancy measures applied to detection of apnea-hypopnea events"
Complexity, Vol. 2018, 2018.

maximum values of AUC for the same OSAH syndrome diagnostic threshold. The results show that DCAF outperforms all the other discrepancy measures in terms of minimum optimal operating cut-off point of the ROC curve as well as in terms of sensitivity and accuracy rate. Also, KL divergence resulted in the best averaged area under the curve ROC and the maximum averaged specificity was yielded by JS divergence. A significant performance improvement was observed when using DCAF or any of the other discrepancy measures compared to random selection.

Several applications exist where it is desirable to maximize the sensitivity. For instance, if the primary purpose of the test is "screening," that is, detection of early disease in a large numbers of apparently healthy persons, then a high sensitivity is generally desired. With this in mind, if a sensitivity of 98% is chosen in the ROC curves in Figure 12, for all used measures, the method achieves a specificity close to 45%. This fact shows that the analysis of pulse oximetry signals by means of the proposed method could be potentially applied as an efficient diagnostic screening tool in clinical practice.

In a previous work [4], it was shown that the MDCS method using DCAF to select discriminative atoms in a given dictionary provides good accuracy rates in the detection of apnea-hypopnea events. In that work, a comparative analysis of the performances yielded by MDCS and other methods [45–47] has shown that MDCS outperforms all the others. It was also observed that the computational cost of MDCS is slightly higher than those required by the other three methods. On the other hand, in this work, we show that MDCS using DCAF for selecting discriminative atoms performs similarly than MDCS using several other traditional discrepancy measures. It is important to highlight that DCAF is very easy to compute and yields competitive performance rates in the detection of apnea-hypopnea events at a low computational cost.

## 8. Conclusions

Sparse representations of signals constitute a powerful technique which yields high accuracy rates in the detection of apnea-hypopnea events. In this work, the difference of conditional activation frequency (DCAF) measure was successfully used for accurately pointing out discriminative atoms in a full dictionary. Additionally, we compared the performance of the DCAF with four widely used discrepancy measures. It was found that the DCAF and three other discrepancy measures (KL, J, and JS divergences) outperform the random selection of atoms, unlike F score. Additionally, DCAF is cheaper to compute. Discriminative subdictionaries were successfully constructed by taking the best ranked atoms of full dictionaries according to their discriminative power. Results show that sparse representations of signals in terms of discriminative subdictionaries result in better performances than the ones obtained in terms of full dictionaries in the detection of apnea-hypopnea events by using only pulse oximetry signals. In this context, it was found that more sparse solutions almost always yielded in better performances. Additionally, it was observed that larger dictionary overcompleteness worsens the performance of the system. Future research lines include more analysis of the DCAF measure, the study of its properties, and an extension of such a measure to multiclass problems.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PIP 2014–2016 no. 11220130100216-CO and PIP 2012–2014 no. 114 20110100284-KA4, by the Air Force Office of Scientific Research, AFOSR/SOARD, through Grant no. FA9550-14-1-0130, by the Universidad Nacional del Litoral through projects CAI + D PIC no. 504 201501 00098 LI (2016) and PIC no. 504 201501 00036 LI (2016), and by the Asociación Gremial de Docentes of the Universidad Tecnológica Nacional (FAGDUT), Paraná section. The authors would like to thank Dr. Luis D. Larrateguy, who is a specialist in sleep-related disorders, for his valuable comments and suggestions.

## References

[1] N. Saito and R. R. Coifman, "Local discriminant bases and their applications," *Journal of Mathematical Imaging and Vision*, vol. 5, no. 4, pp. 337–358, 1995.

[2] S. Tabibian, A. Akbari, and B. Nasersharif, "Speech enhancement using a wavelet thresholding method based on symmetric Kullback–Leibler divergence," *Signal Processing*, vol. 106, pp. 184–197, 2015.

[3] M. Sánchez-Gutiérrez, E. M. Albornoz, H. L. Rufiner, and J. G. Close, "Post-training discriminative pruning for RBMs," *Soft Computing*, pp. 1–15, 2017.

[4] R. E. Rolón, L. D. Larrateguy, L. E. Di Persia, R. D. Spies, and H. L. Rufiner, "Discriminative methods based on sparse representations of pulse oximetry signals for sleep apnea–hypopnea detection," *Biomedical Signal Processing and Control*, vol. 33, pp. 358–367, 2017.

[5] V. Peterson, H. L. Rufiner, and R. D. Spies, "Generalized sparse discriminant analysis for event-related potential classification," *Biomedical Signal Processing and Control*, vol. 35, pp. 70–78, 2017.

[6] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[7] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[8] A. Gupta, S. Parameswaran, and C.-H. Lee, "Classification of electroencephalography (EEG) signals for different mental activities using Kullback Leibler (KL) divergence," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1697–1700, Taipei, Taiwan, April 2009.

[9] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.

[10] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 2006.

[11] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[12] X. Zhang and Q. Ding, "Respiratory rate estimation from the photoplethysmogram via joint sparse signal reconstruction and spectra fusion," *Biomedical Signal Processing and Control*, vol. 35, pp. 1–7, 2017.

[13] Y. Zhou, X. Hu, Z. Tang, and A. C. Ahn, "Sparse representation-based ECG signal enhancement and QRS detection," *Physiological Measurement*, vol. 37, no. 12, pp. 2093–2110, 2016.

[14] M. Aharon, M. Elad, and A. Bruckstein, "*rm*K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[15] D. S. Pham and S. Venkatesh, "Joint learning and dictionary construction for pattern recognition," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.

[16] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern*, pp. 2691–2698, San Francisco, CA, USA, June 2010.

[17] M. J. Sateia, "International classification of sleep disorders-third edition: highlights and modifications," *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014.

[18] N. A. Dewan, F. J. Nieto, and V. K. Somers, "Intermittent hypoxemia and OSA: implications for comorbidities," *Chest*, vol. 147, no. 1, pp. 266–274, 2015.

[19] W. Kukwa, E. Migacz, K. Druc, E. Grzesiuk, and A. M. Czarnecka, "Obstructive sleep apnea and cancer: effects of intermittent hypoxia?," *Future Oncology*, vol. 11, no. 24, pp. 3285–3298, 2015.

[20] M. Torres, R. Laguna-Barraza, M. Dalmases et al., "Male fertility is reduced by chronic intermittent hypoxia mimicking sleep apnea in mice," *Sleep*, vol. 37, no. 11, pp. 1757–1765, 2014.

[21] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705-706, 1997.

[22] J. Durán, S. Esnaola, R. Rubio, and A. Iztueta, "Obstructive sleep apnea–hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr," *American Journal of Respiratory and Critical Care Medicine*, vol. 163, no. 3, pp. 685–689, 2001.

[23] R. Thurnheer, K. E. Bloch, I. Laube, M. Gugger, M. Heitz, and Swiss respiratory Polygraphy registry, "Respiratory polygraphy in sleep apnoea diagnosis. Report of the Swiss respiratory polygraphy registry and systematic review of the literature," *Swiss Medical Weekly*, vol. 137, no. 5-6, pp. 97–102, 2007.

[24] E. García-Díaz, E. Quintana-Gallego, A. Ruiz et al., "Respiratory polygraphy with actigraphy in the diagnosis of sleep apnea-hypopnea syndrome," *Chest*, vol. 131, no. 3, pp. 725–732, 2007.

[25] A. Yadollahi, E. Giannouli, and Z. Moussavi, "Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals," *Medical & Biological Engineering & Computing*, vol. 48, no. 11, pp. 1087–1097, 2010.

[26] M. Elad, *Sparse and Redundant Representations*, Springer, New York, NY, USA, 2010.

[27] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[28] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[29] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," in *Wavelets and Their Applications. NATO ASI Series (Series C: Mathematical and Physical Sciences)*, J. S. Byrnes, J. L. Byrnes, K. A. Hargreaves, and K. Berry, Eds., vol. 442, pp. 363–379, Springer, Dordrecht, 1994.

[30] M. S. Lewicki and B. A. Olshausen, "Probabilistic framework for the adaptation and comparison of image codes," *Journal of the Optical Society of America A*, vol. 16, no. 7, p. 1587, 1999.

[31] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.

[32] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 5, pp. 2443–2446, Phoenix, AZ, USA, March 1999.

[33] Z. Jiang, Z. Lin, and L. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[34] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, no. 4, pp. 349–369, 1989.

[35] W. Gersch, F. Martinelli, J. Yonemoto, M. D. Low, and J. A. Mc Ewan, "Automatic classification of electroencephalograms: Kullback-Leibler nearest neighbor rules," *Science*, vol. 205, no. 4402, pp. 193–195, 1979.

[36] P. J. Moreno, P. P. Ho, and N. Vasconcelos, "A Kullback–Leibler divergence based kernel for SVM classification in multimedia applications," in *Advances in Neural Information Processing Systems*, S. Thrun, L. K. Saul, and P. B. Schölkopf, Eds., vol. 16, pp. 1385–1392, MIT Press, 2004.

[37] C. C. Aggarwal, *Data Classification: Algorithms and Applications*, CRC Press, 2014.

[38] S. F. Quan, B. V. Howard, C. Iber et al., "The Sleep Heart Health Study: design, rationale and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.

[39] B. K. Lind, J. L. Goodwin, J. G. Hill, T. Ali, S. Redline, and S. F. Quan, "Recruitment of healthy adults into a study of overnight sleep monitoring in the home: experience of the Sleep Heart Health Study," *Sleep and Breathing*, vol. 7, no. 1, pp. 13–24, 2003.

[40] F. Lestussi, L. Di Persia, and D. Milone, "Comparison of online wavelet analysis and reconstruction: with application to ECG," in *2011 5th International Conference on Bioinformatics*

*and Biomedical Engineering*, pp. 1–4, Wuhan, China, May 2011.

[41] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[42] J. Tang, C. Deng, and G. B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809–821, 2016.

[43] J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," *Investigative Radiology*, vol. 14, no. 2, pp. 109–121, 1979.

[44] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627–635, 2013.

[45] E. Chiner, J. Signes-Costa, J. M. Arriero, J. Marco, I. Fuentes, and A. Sergado, "Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?," *Thorax*, vol. 54, no. 11, pp. 968–971, 1999.

[46] J. C. Vázquez, W. H. Tsai, W. W. Flemons et al., "Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea," *Thorax*, vol. 55, no. 4, pp. 302–307, 2000.

[47] G. Schlotthauer, L. E. Di Persia, L. D. Larrateguy, and D. H. Milone, "Screening of obstructive sleep apnea with empirical mode decomposition of pulse oximetry," *Medical Engineering and Physics*, vol. 36, no. 8, pp. 1074–1080, 2014.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R.E. Rolon, I. E. Careis, L. Di Persia, R. Spies & H. L. Rufiner; "Complexity-based discrepancy measures applied to detection of apnea-hypopnea events"
Complexity, Vol. 2018, 2018.