Automatic scoring of apnea and hypopnea events using blood oxygen saturation signals

R.E. Rolon[a,d], I.E. Gareis[a], L.D. Larrateguy[b], L.E. Di Persia[a], R.D. Spies[c], H.L. Rufiner[a]

[a] *Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i)-UNL-CONICET, Santa Fe, Argentina*
[b] *Centro de Medicina Respiratoria de Paraná, Entre Ríos, Argentina*
[c] *Instituto de Matemática Aplicada del Litoral, IMAL-UNL-CONICET, Santa Fe, Argentina*
[d] *Facultad Regional Paraná, Universidad Tecnológica Nacional, Entre Ríos, Argentina*

Abstract

The obstructive sleep apnea-hypopnea (OSAH) syndrome is a common and frequently undiagnosed sleep disorder. It is characterized by repeated events of partial (hypopnea) or total (apnea) obstruction of the upper airway while sleeping. To quantify the severity of the pathology, the Apnea Hypopnea Index (AHI) is used. This index is defined as the average number of apnea and hypopnea events per hour of sleep. Discriminating between these two types of events is a very challenging task and in fact most traditional methods fail to do it. A reliable recognition of such events would not only allow for an accurate estimation of the AHI index, but it would also provide useful information regarding the severity of the pathology, which is very important for clinical purposes. In this work we use a method for structured dictionary learning, which is found to be suitable for automatically differentiating between apnea and hypopnea using as a unique input blood oxygen saturation signals. The method is tested for both classification of segments and OSAH screening on the Sleep Heart Health Study database. For OSAH screening, a receiver operating characteristic curve analysis shows an average area under the curve of 0.934 and diagnostic sensitivity and specificity of 89.10% and 86.70%, respectively. These results represent important improvements with respect to all state-of-the-art procedures which where used for comparison purposes. They also provide a solid support for our conclusion that the method can be used for screening OSAH syndrome more reliably and conveniently, using only a pulse oximeter.

Keywords: Pulse oximetry, Apnea-hypopnea events, Sleep disorders screening, Structured dictionary learning, Discriminant measures, Multiclass classification problems.

1. Introduction

Pulse oximetry, being a cheap and non-invasive technique, has become a promising supporting tool for the diagnosis of sleep disorders [1, 2, 3]. Sleep disorders comprise several types of medical conditions. The most common one of them is the Obstructive Sleep Apnea Hypopnea (OSAH) syndrome, which is caused by frequent breathing pauses due to partial (hypopnea) or total (apnea) blockage of the upper airway during sleeping, which lead to several physiological changes such as blood oxygen desaturation [4, 5]. To establish the severity of this pathology, the apnea-hypopnea index (AHI) is commonly used. This index is defined as the number of apnea-hypopnea events per hour of sleep or record according to whether it refers to a complete study or a simplified one, respectively (more on this later). Most screening methods do not discriminate between apnea and hypopnea events since it is not strictly required for computing the AHI index [2]. However, a reliable recognition of individual apnea and hypopnea events would not only allow for an accurate estimation of the AHI index, but it would also supply valuable information regarding the severity of the OSAH syndrome, which is very important for clinical and decision-making purposes [6]. Nevertheless, automatically detecting and differentiating between those two events is a very challenging task, specially when the problem is addressed using a unique signal as input, such as the pulse oximetry (SaO_2).

Achieving a good AHI estimation using recordings of just a few signals is a difficult problem that requires of precise ad-hoc evaluation tools for the clinical screening of OSAH syndrome [7]. In the past decade much interest in the development of portable devices using at most two sensors for OSAH screening has been observed (e.g. [8, 9, 10, 11]). In particular, the authors in [9] present a detailed review of existing methods that use only pulse oximetry signals for automatically classifying patients having OSAH syndrome. It is important to highlight however that all methods mentioned in that review address only the detection of the pathology and do not recognize nor classify small segments of oximetry signals as normal breathing, apnea or hypopnea events. In that way, up to our knowledge, the problem of individually classifying abnormal respiratory events using only SaO_2 signals in a multiclass scenario has never been explored before. Therefore, properly identifying hypopneas which were not detected by other approaches may add value in the diagnosis and treatment of the patients.

There are methods for binary classification (existence or nonexistence of abnormal respiratory events) of SaO_2 signals from which the AHI index can be estimated [2, 3, 12, 13]. In particular, the articles [12] and [13] make use of the so called Oxygen

Desaturation Index (ODI) defined as the number of times that the SaO_2 signal falls below a prescribed percentage of signal saturation regarding a baseline level per hour of study. It is timely to point out however that although the concept of "baseline level" is somewhat intuitive, there is yet no consensus about its formal definition, and different authors have adopted different ones [12, 13]. In [12], for instance, the baseline level was defined as the desaturation mean of the previous minute, while a completely different approach was followed in [13] where it was computed using a moving time average. In [2], the authors present a method for detecting blood oxygen desaturations using specific waves (or modes) coming from empirical mode decompositions of SaO_2 signals. In that work, the desaturations are identified by making use of a few thresholds and a set of simple rules which lead to the detection of the sleep apnea-hypopnea syndrome. Finally, in [3], we introduced a different approach based on sparse representations of SaO_2 signals. In that work, the AHI index is directly estimated without computing the ODI index, as the average number of abnormal respiratory events per hour of study.

All previously mentioned approaches are unable to distinguish between apnea and hypopnea events, which is very important for having a deeper understanding of the underlying pathology and for providing better treatments [14]. Moreover, some of those approaches require of appropriate estimates of the baseline level, and poor approximations of it result in errors in the quantification of the desaturations of the SaO_2 signals. Hence, it becomes highly desirable to come up with an automatic multiclass classification method for detecting and distinguishing between normal breathing, apnea and hypopnea events in SaO_2 signals. Although some previous articles have tackled this issue, to the best of our knowledge, this is first time that the problem is addressed using only SaO_2 signals, which constitutes the main contribution of this work [15, 16].

The organization of this article is as follows. In Section 2, a brief description about abnormal respiratory events during sleep is presented. Dictionary learning methods for sparse representation are introduced in Section 3. Section 4 contains details on all designed experiments. Results and discussions are introduced in Section 5 while conclusions are finally presented in Section 6.

2. Sleep apnea

It is well known that getting enough sleep is extremely important for maintaining both mental and physical health. However, good sleeping very often becomes affected by the presence of sleep-related breathing disorders. Poor sleep quality causes excessive daytime sleepiness affecting the productivity and efficiency of people, including their ability to think clearly, react quickly and memorize efficiently, triggering bad decisions and highly increasing the risk of having domestic, work and traffic accidents [17].

Polysomnography (PSG) is the reference study for diagnosing OSAH syndrome. This study requires of specially conditioned sleep units as well as the simultaneous recording of several biomedical signals. However the accessibility to PSG is very limited mainly because PSG units are not commonly available and because the studies are both lengthy and costly, making the process of obtaining good quality signals extremely complicated. In addition, a PSG study requires the attention of specialized technicians to ensure continuous time visualization and recording of all the signals being acquired. A complete PSG study consists of the simultaneous measuring of a minimum of seven physiological signals such as electroencephalography (EEG), electrooculography (EOG), electromiography (EMG), electrocardiography (ECG), airflow and SaO_2. It is important to point out however that the continuous acquisition of these signals highly affects the quality of sleep, making it even more difficult to achieve an accurate diagnosis. Because all those difficulties, new screening approaches are always been developed. An ideal screening method can be considered as one that, on one hand leads to precise results, and on the other hand it uses as few signals as possible without degrading the quality of sleep [7].

For the reasons described above, portable systems for assessing OSAH syndrome, that can be used outside sleep units, have been developed. In this sense other evaluation procedures exist, such as home PSG, home Respiratory Poligraphy (RP) and other simplified procedures, to name a few. Although home PSG has the advantage of not requiring of any trained personnel, it still needs the acquisition of at least seven respiratory and sleep signals, just like a standard PSG. Nowadays, there is a home PSG which allows the classification of the different sleep stages by using a single-channel EEG. However, this procedure still requires of several signals whose appropriate acquisition affects the quality of sleep. On the other hand, home RP studies allow for the evaluation of cardiorespiratory variables without taking into account EEG, EOG and EMG signals and therefore they are unable to detect wakefulness and to determine sleep stages [18]. Hence, even though home RP is simpler than both standard PSG and home PSG, it still needs the continuous measurement of several physiological signals, whose acquisition affects sleep quality. Finally, simplified procedures make use of only one or two cardiorespiratory variables, such as airflow, respiratory movements, heart rate, tracheal sound and SaO_2. In particular, the SaO_2 signal has become a reasonable alternative for OSAH syndrome screening and it is the one that will be used in this article [1, 2, 3].

The severity of OSAH syndrome is classified as normal, mild, moderate or severe

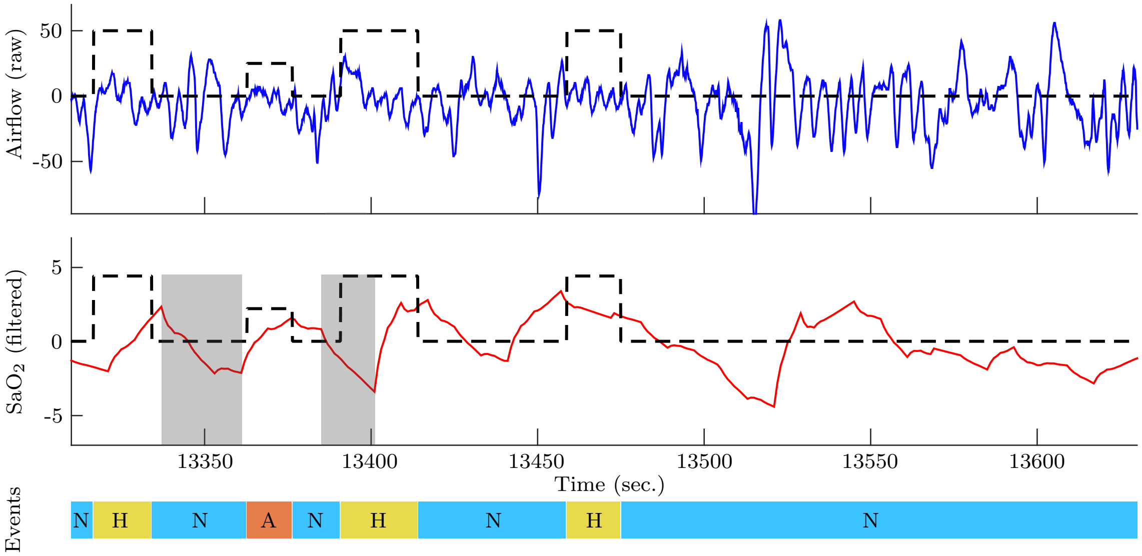

Figure 1: A small portion of an airflow signal (top), a wavelet filtered SaO_2 signal (middle) and labels of normal breathing and abnormal respiratory events (apnea and hypopnea) that occur during sleeping (bottom). Black dashed lines: apnea event (lower) and hypopnea event (higher). Data obtained from [19].

depending on whether the AHI values fall within the intervals $[0, 5)$, $[5, 15)$, $[15, 30)$, or $[30, \infty)$, respectively. It is known that towards the end of each apnea or hypopnea event, a desaturation of the hemoglobin occurs. It is therefore reasonable to think that these deasaturations contain valuable information related the particular events of apnea and hypopnea, which are very often impossible to be recognized and distinguished by the human eye. The top and middle waveforms in Figure 1 show a six-minutes portion of a typical airflow signal and the corresponding filtered SaO_2 signal, respectively (see Section 4.1) [3]. Black-dashed lines represent the beginning and end of an event (apnea for the lower and hypopnea for the higher dash lines). Also, the labels N (normal breathing), A (apnea) and H (hypopnea) are shown at the bottom. It is important to mention that these labels were introduced by medical experts, after a detailed analysis of all the signals acquired during the PSG study. By observing both the airflow and the SaO_2 signals, it can be seen that the time frame between the reduction (or stopping) of airflow and the beginning of oxygen desaturation levels is very variable. The SaO_2 signal at the middle of Figure 1 shows two gray-highlighted portions on the left, corresponding to the time intervals where desaturations produced by a hypopnea event (left) and an apnea event (right) occur. As it can be observed, the minimum saturation values and the general morphology

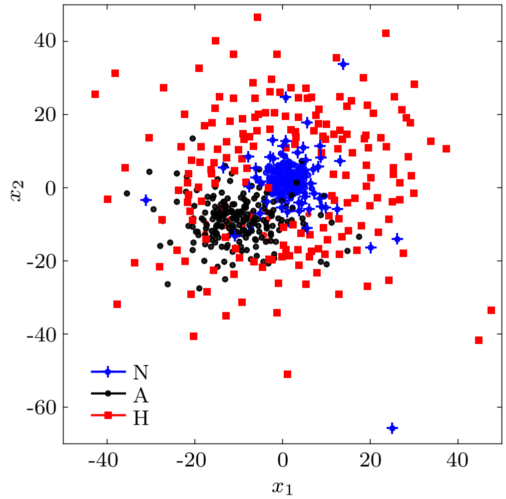

Figure 2: A representation of the class distribution after applying a mapping denoted by *Sammon mapping*, in its two most relevant attributes obtained from SaO_2 signals (estimated taking into account 200 examples for each class). Data obtained from [19].

of the signal on those two intervals are very similar. Hence, it becomes evident that automatic recognition of single apnea and hypopnea events from only SaO_2 signals is a very challenging classification problem. To further visualize the difficulty of this classification problem, a technique for dimensionality reduction called "Sammon Mapping" was applied to low-dimensional samples of SaO_2 signals [20]. Figure 2 shows projections to two-dimensional attributes of signals for the classes N, H and A. It can be observed that the distribution of the different classes in the attributes space highly overlap each other. Although the distributions representing both classes normal breathing and apnea events seems to be fairly separated, the distribution of hypopnea events presents a very high dispersion leading to a great degree of overlap with them.

3. Dictionary Learning for Sparse Representation

3.1. Basic methods

The representation of signals based on a dictionary consists of finding appropriate linear combinations of atoms in the prescribed dictionary to represent a given set of signals. This representation problem can be divided in two sub-problems: an inference problem and a learning problem. We proceed to describe each one of them. For that, let $\mathbf{x} \in \mathbb{R}^N$ be an input signal and let $\Phi \in \mathbb{R}^{N \times M}$ (usually $M \geq N$) be a dictionary whose columns $\boldsymbol{\phi}_j \in \mathbb{R}^N$, $j = 1, 2, \cdots, M$, are atoms that we want to use for representing $\mathbf{x}$ in the form $\mathbf{x} \cong \Phi\mathbf{a} = \sum_{j=1}^{M} a_j \phi_j$. Here, and in the sequel, we shall refer to the vector $\mathbf{a} = [a_1\ a_2\ \cdots\ a_M]^T \in \mathbb{R}^M$ as a "representation" of $\mathbf{x}$.

The inference problem essentially consists of finding the optimal (in a certain sense) representation $\mathbf{a}$ of the given signal $\mathbf{x}$. A sparse solution of this problem is a representation $\mathbf{a}$ with just a few non-zero components. If in a given representation a certain coefficient is non-zero, then we shall refer to it as an "active" component.

A way of obtaining a sparse representation of the signal $\mathbf{x}$ based on the dictionary Φ consists of solving the following problem:

$$(P_0) \quad \mathbf{a}^* \doteq \underset{\mathbf{a}\in\mathbb{R}^M}{\operatorname{argmin}}\, ||\mathbf{a}||_0, \quad \text{subject to } \mathbf{x} = \Phi\mathbf{a},$$

where $||\mathbf{a}||_0$ denotes the l_0 pseudo-norm, defined as the number of non-zero elements of $\mathbf{a}$.

Solving (P_0) is generally an NP hard problem yielding this approach highly unsuitable for most applications [21, §1.8]. This is so because in (P_0) we are imposing an exact representation which, in most practical cases, is neither strictly necessary nor desired. To overcome the computational burden which entails 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 pseudo-norm, i.e. solve:

$$(P_0^q) \quad \mathbf{a}^* \doteq \underset{\mathbf{a}\in\mathbb{R}^M}{\operatorname{argmin}}\, ||\mathbf{x} - \Phi\mathbf{a}||_2, \quad \text{subject to } ||\mathbf{a}||_0 \leq q,$$

where q is a prescribed integer parameter. Several approaches for solving problem (P_0^q) were proposed [22, 23, 24]. The one most widely used is Orthogonal Matching Pursuit (OMP) which consists of approximating the solution in a greedy way providing a good trade-off between computational cost and representation error [25]. Additionally, the method ensures convergence to the projection of $\mathbf{x}$ into the span of the dictionary atoms [24].

The dictionary Φ can be constructed either using a pre-specified group of atoms (such as those obtained through the Wavelet Packet decomposition) or by means of data-driven learning approaches. The dictionary learning problem associated to the data q, M, $N \in \mathbb{N}$, $M \geq N$ and a collection of n signals in $\mathbb{R}^N$, $\mathbf{x}_1, \cdots, \mathbf{x}_n$, can be formally written as:

$$\text{(DL)} \quad [\Phi^*, \mathbf{a}_1^*, \cdots, \mathbf{a}_n^*] \doteq \underset{\substack{\Phi\in\mathbb{R}^{N\times M} \\ \mathbf{a}_i\in\mathbb{R}^M, ||\mathbf{a}_i||_0\leq q, 1\leq i\leq n.}}{\operatorname{argmin}} \sum_{i=1}^{n} ||\mathbf{x}_i - \Phi\mathbf{a}_i||_2^2$$

A solution of this problem yields on one hand a dictionary Φ and, on the other hand, representations $\mathbf{a}_i$ for all the signals $\mathbf{x}_1, \cdots, \mathbf{x}_n$ (in terms of such a dictionary) complying with the imposed sparsity constraint. Although several methods for solving

(DL) exist, the most widely used is an iterative algorithm called K Singular Value Decomposition (KSVD) [26]. This approach consists of two steps: an inference step and a dictionary learning step. The OMP algorithm (for example) is used for obtaining the representation coefficients, which is then followed by a dictionary learning step where the atoms are updated one-at-a-time and the representation coefficients are adjusted in order to minimize the total representation error.

3.2. Discriminant dictionaries

As mention above, a dictionary Φ can be constructed using data-driven learning methods aimed exclusively to minimize the total representation error. However, a dictionary learned in this way quite often produces representations of signals which turn out to be unsatisfactory if the final objective is pattern recognition. This is so because, as it is well known, a good representation does not necessarily guarantee good classification performance. A way to overcome this flaw consists of incorporating available prior information about class membership of the signals into the objective function in (DL) [27, 28]. In [27], for example, a discriminant version of the standard KSVD method applied to face recognition was presented. In that work, the authors included a discriminant term into the objective function of the standard KSVD algorithm. Results have shown that such a modification constitutes an appropriate way to learn dictionaries simultaneously complying with both desired properties: low reconstruction error and high recognition rates. In [28], a sparse-constrained optimization problem combining the objective function of the classification and the representation error of both labeled and unlabeled data, was formulated.

With the objective of improving classification performance, new approaches based on the design of structured dictionaries were recently proposed [29, 30, 31, 32]. A structured dictionary can be thought of as a collection of class-specific sub-dictionaries which are designed to capture discriminant properties of each class as well as common features among all classes in the data. In this direction, an initial approach consists of learning one dictionary for each class, then classify by minimizing the representation error among all classes [33]. Recently, a method called “Most Discriminative Columns Selection” (MDCS), which was shown to be capable of efficiently building structured dictionaries in a binary classification scheme, was developed [3]. Figure 3 shows a schematic representation of the MDCS procedure for a three-class classification problem. In this case the classes are identified as N, A and H. The dictionary Φ is learned in an unsupervised way using all training signals for solving problem (DL). After that, the representation matrices $\mathbf{A}_{\mathrm{N}}$, $\mathbf{A}_{\mathrm{A}}$ and $\mathbf{A}_{\mathrm{H}}$ whose columns are the corresponding representation vectors, are computed using the three separate sets of labeled signals $\mathbf{X}^*_{\mathrm{N}}$, $\mathbf{X}^*_{\mathrm{A}}$ and $\mathbf{X}^*_{\mathrm{H}}$, respectively. Next, the atoms of Φ are ranked

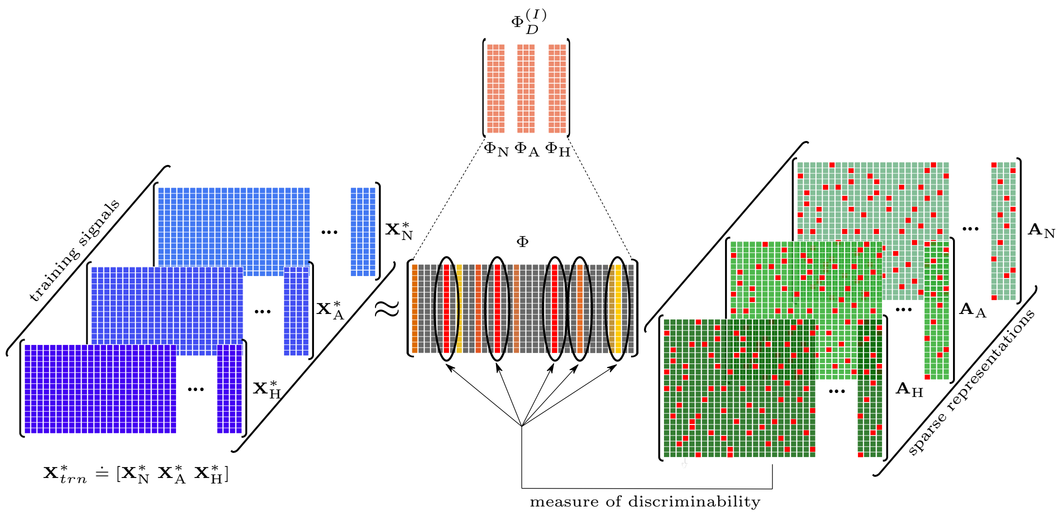

Figure 3: A schematic representation of the learning process of discriminant structured dictionaries using the MDCS method.

according to a prescribed measure of discriminability in terms of their role in the sparse representation of the signals for each class (see [34], Section 3.2). Following this ranking procedure, and given a prescribed positive integer I (more on this later), the best I atoms for each class are selected and used for building new class-specific sub-dictionaries Φ_N, Φ_A and Φ_H for classes N, A and H, respectively. The structured dictionary, which we denote by $\Phi_D^{(I)}$, is finally constructed by stacking side-by-side all sub-dictionaries, i.e. $\Phi_D^{(I)} = [\Phi_\mathrm{N}\ \Phi_\mathrm{A}\ \Phi_\mathrm{H}]$. The parameter I is used to restrict the size of the final dictionary, in the sense that $\Phi_D^{(I)}$ will end up having exactly $I \times k$ columns, where k is the number of classes. This restriction intends to improve the generalization capabilities reducing the size of the final feature vectors, what in turn, reduces the computing time required for classification.

Along MDCS, a method for discriminant features selection called "Most Discriminative Atoms Selection" (MDAS) was proposed [3]. The main difference between both MDCS and MDAS is that in the later no new structured dictionary $\Phi_D^{(I)}$ is built. Instead the original dictionary Φ is preserved and the ranking of the atoms is used only to select the components to be used for classification. It is important to point out that although both MDCS and MDAS were originally proposed for dealing only with binary classification problems, their extension to multiclass problems is straight forward. In what follows, we shall denote by MDCS-BC, MDCS-MC, MDAS-BC and MDAS-MC the binary and multiclass versions of MDCS and MDAS, respectively.

Following on the idea behind MDCS, DAS-KSVD can be thought of as its extension to multiclass classification problems. Unlike MDCS, instead of selecting all

Algorithm 1 DAS-KSVD method

procedure DAS-KSVD($\mathbf{X}_{trn}, q, r_f, I, \mathbf{c}, t$)
  $p_0(i) = 1/n$, for all i
  for $l \leftarrow 0, I-1$ **do**
    $[\mathbf{X}_{lrn}, p_{l+1}] \leftarrow$ SampleData($\mathbf{X}_{trn}, t, p_l, l$)
    $\Phi \leftarrow$ KSVD($\mathbf{X}_{lrn}, r_f, q$)
    $\mathbf{A}_{lrn} \leftarrow$ OMP($\mathbf{X}_{lrn}, \Phi, q$)
    $m_{\alpha^*,\beta^*} \leftarrow$ DiscMeasure($\mathbf{A}_{lrn}, \mathbf{c}, q$)
    $\Phi_d \leftarrow$ GetAtoms($\Phi, m_{\alpha^*,\beta^*}$)
    $\Phi_D^{(i)} \leftarrow$ SaveAtoms(Φ_d)
  end for
  return $\Phi_D^{(I)}$
end procedure

I discriminant atoms simultaneously for both classes, DAS-KSVD iterates the process of choosing only one discriminant atom for each one of the classes at each step. Moreover, DAS-KSVD incorporates a re-sampling technique and a signal degradation stage that jointly promote diversity in the generation of discriminant atoms at each iteration. More precisely, if a certain set of signals is used for learning a dictionary at a particular iteration, then the re-sampling technique forces such signals to be less likely to be chosen than the remaining ones in the following iterations. On the other hand, the signal degradation step is meant to increase robustness and it consists of adding an additive zero-mean Gaussian noise (whose magnitude increases proportionally with the iteration step) to all signals used for learning the dictionary. Hence, by promoting diversity in this way, one expects that the resulting learned atoms will be capable of highlighting different intrinsic properties of the whole training data. For more details on this, we refer the reader to [34]. The steps for constructing the dictionary with DAS-KSVD are summarized in Algorithm 1.

Figure 4 shows a schematic representation of one iteration of DAS-KSVD for a three-class classification problem. Observe that before using a method for solving (DL), a re-sampling technique is applied. Then, the dictionary Φ is learned in an unsupervised way using all learning signals $\mathbf{X}_{lrn}$. After that, the representation matrices $\mathbf{A}_{\mathrm{N}}$, $\mathbf{A}_{\mathrm{A}}$ and $\mathbf{A}_{\mathrm{H}}$ whose columns are the corresponding representation vectors, are computed using the three separate sets of learning signals $\hat{\mathbf{X}}_{\mathrm{N}}$, $\hat{\mathbf{X}}_{\mathrm{A}}$ and $\hat{\mathbf{X}}_{\mathrm{H}}$, respectively. Next, the atoms of Φ are ranked according to an appropriately defined multiclass measure of discriminability (details about this measure can be found in [34], Section 3.2). After this ranking procedure, only one atom for each

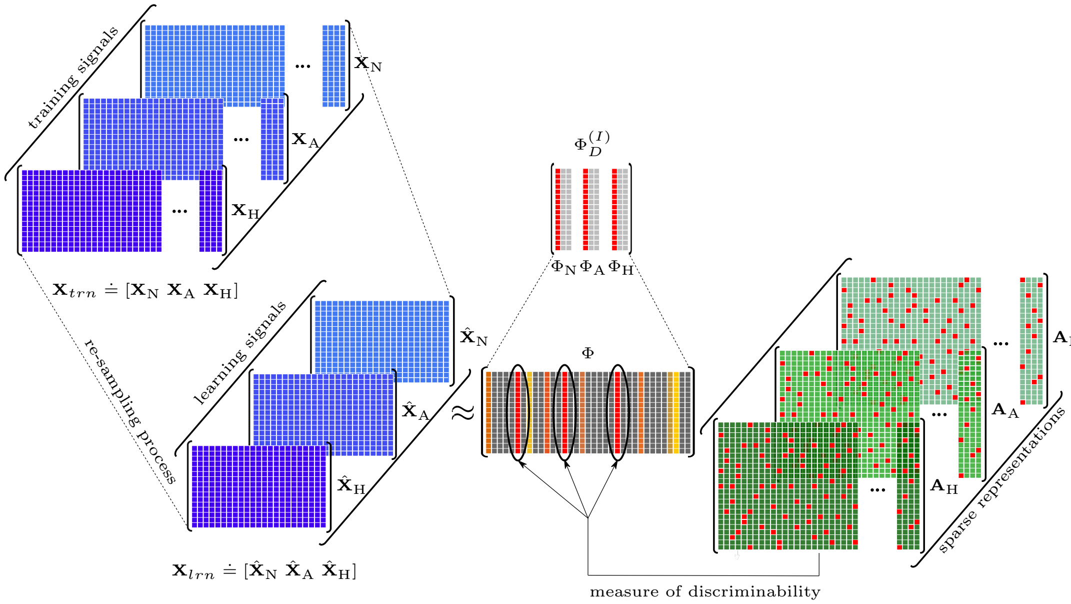

Figure 4: A schematic representation of one iteration of the learning process of discriminant structured dictionaries using the DAS-KSVD method.

class is selected and used for building new class-specific sub-dictionaries Φ_N, Φ_A and Φ_H for classes N, A and H, respectively. The structured dictionary, which is denoted by $\Phi_D^{(I)}$, is finally constructed by stacking side-by-side all sub-dictionaries, i.e. $\Phi_D^{(I)} = [\Phi_\mathrm{N}\ \Phi_\mathrm{A}\ \Phi_\mathrm{H}]$.

4. Experimental setup

The main objective of this article is the comparison of the overall classification performances in the context of OSAH syndrome screening of MDCS, MDAS (both in their binary and multiclass versions) and DAS-KSVD. To achieve that objective, two experiments were carried out. The first one was designed with the final goal of classifying the segments of SaO_2 signals in one and only one of the three classes: normal breathing (N), apnea (A) or hypopnea (H). The second experiment was designed to detect the existence or non-existence of the pathology. The whole experimental setup is described below.

4.1. Database and signal pre-processing

The Sleep Heart Health Study (SHHS) database was originally designed to explore possible correlations between sleep related breathing disorders and cardiovascular diseases [19, 35]. This database consists of several complete PSG studies, each one of them containing a group of physiological signals such as EEG, ECG, nasal airflow and SaO_2. In addition, annotations of sleep stages, arousals and events of apnea and hypopnea are provided. The criteria that medical experts adopted for identifying apnea and hypopnea events were the following [5]. An apnea event is a complete (or almost complete) blockage of the upper airflow for at least ten seconds, usually associated with a desaturation in the SaO_2 signal or an arousal. A hypopnea event is a reduction in airflow by less than a 70% of the baseline level, associated with a desaturation in the SaO_2 signal or an arousal.

In this article we make use of the first online version of the database called "Sleep Heart Health Study" (SHHS-2)[1]. This database consists of 995 complete PSG studies, 41 of which were discarded due to labeling flaws [3]. With the remaining 954 studies, we performed k-fold cross validation, with $k = 10$. For OSAH syndrome detection, all performance measures (more on it later) were calculated individually (per study) and then averaged for the reported results.

Mainly due to patient movements, baseline wander and undesired disconnections (among many other factors), the original raw SaO_2 signals require of an appropriate pre-conditioning process. For that, linear interpolation and wavelet filters, as those used in a previous work [3], were applied. Figure 1 shows a small portion of a SaO_2 signal (top) and its wavelet-filtered version (middle). Here, it is important to point out that the wavelet filtering process produces no effective signal loss. For more details on applications of such a filtering procedure to real data, we refer the reader to [36].

Signals are segmented into vectors $\mathbf{x}_i \in \mathbb{R}^N$ of length $N = 128$ (corresponding to 128 seconds of the signal recording) with a 50% overlapping between two consecutive segments. In this process, segments containing artifacts or disconnections are discarded. Then, for each fold in the cross validation, a matrix $\mathbf{X}_{trn} \in \mathbb{R}^{128\times n_{trn}}$ is constructed by stacking side-by-side n_N, n_A and n_H vectors belonging to the classes N, A and H, respectively. Clearly, $n_{trn} = n_N + n_A + n_H$. Similarly, another matrix $\mathbf{X}_{tst} \in \mathbb{R}^{128\times n_{tst}}$ is built using the vectors associated to the testing set.

4.2. Dictionary learning settings

For DAS-KSVD, all experiments were performed setting $I = 22$ (i.e. 22 iterations). Thus, the final structured dictionary consists of 66 atoms (assuming $k = 3$). For each one of the classes used to learn the full dictionary (by means of KSVD), the number of samples was set to $t = 9000$. Also, several trials were performed

[1] `https://physionet.org/physiobank/`

in order to obtain adequate values for both parameters τ_1 and τ_2. In particular, it was found that values of $\tau_1 = 0.5$ and $\tau_2 = 0.1$ are suitable for this application. In addition, $\tau_2 = 0.1$ resulted in the best trade-off between signal degradation and the number of iterations. Finally, for each and every fold in the cross validation, the average value of the optimal pair of parameters (α^*, β^*) was found to be in a circle of radius of 0.1 centered at $(0.7, 0.1)$. All parameters of the KSVD method such as the sparsity constrain q and the redundancy factor of the dictionary r_f, were set equal to those used in a previous work [34]. Finally, for both MDCS-MC and MDAS-MC, all parameters were set as for DAS-KSVD. It is important to mention, however, that these two methods make use of a different input data matrix $\mathbf{X}^*_{trn}$ which is composed of a balanced set of randomly selected segments from $\mathbf{X}_{trn}$. Since n_L segments were chosen for each class, the final size of $\mathbf{X}^*_{trn}$ was $128 \times 3n_L$ where n_L is the number of segments chosen from each class.

4.3. Classification of segments and OSAH screening

In order to classify segments of SaO_2 signals into the three different classes, a feed-forward Multilayer Perceptron (MLP) neural network was used. In particular the experiments were run using three layers (input, hidden and output). Naturally, input and output layer sizes were set to 150 and 3 corresponding to $I \times k$ and k, respectively. Several preliminary trials were performed by varying the number of neurons in the hidden layer between 100 to 1000 with a step of 100 neurons (with *tansig* activation function) in order to determine an appropriate size. The results indicated that no significant improvement is obtained with sizes above 500. To train this network, conjugate gradient descent was used. For classification purposes, both Mean Squared Error (MSE) and Cross-Entropy cost functions were used, obtaining slightly better results with the latter. Thus, final experiments only use Cross-Entropy cost function.

To carry out the first experiment, two balanced sets of 21000 and 4500 samples were randomly selected from $\mathbf{X}_{trn}$ and used for training and validation purposes, respectively. Also, an additional balanced set of 4500 samples was randomly chosen from $\mathbf{X}_{tst}$ and used for testing purposes. Then, sparse representations of these new data sets in terms of the previously learned dictionary were found and used as input of the classifier.

For the detection of OSAH syndrome, it is well known that in a typical PSG study, the recorded signals are provided to medical experts who identify and label apnea and hypopnea events, which are later used for computing the AHI index. In a similar way, in our analysis, each testing study was appropriately filtered and segmented in order to classify its segments as N, A and H, by means of the previously described process. Then, an estimated AHI (AHI_{est}) was computed by counting the total number of segments classified as A or H and dividing it by the duration of the study, in hours. This new index was used for OSAH syndrome detection. Finally, each study was considered as pathological if the obtained AHI_{est} was greater than a certain prescribed detection threshold [37].

4.4. Performance measures

To analyze and quantify the ability of the MLP to classify segments of SaO_2 signals in a multiclass scenario, a confusion matrix was constructed. The confusion matrix is a very useful tool for reporting results in multiclass classification problems because it gives a full overview of all relations between the classifier predictions and the known (true) labels. Rows and columns of such a matrix refer to known and predicted class labels of the dataset, respectively, while its diagonal and off-diagonal elements correspond to observations that are correctly and incorrectly classified, respectively. This information summarizes the types of errors that occur during training, validation and testing. Based on the confusion matrix, the overall accuracy as well as other three widely used class-specific measures (sensitivity (Se), specificity (Sp) and precision (Pr)) were extracted. In this article, the confusion matrix is normalized by dividing each one of the elements in its rows by the total number of testing samples that belong to each class.

To assess the ability of the proposed system in detecting patients suspected of suffering from moderate to severe OSAH syndrome, i.e. persons having an AHI index greater than 15, a Receiver Operating Characteristics (ROC) analysis was performed [38]. The optimal cut-off point (associated to a prescribed detection threshold) of the ROC curve is the one that simultaneously maximizes sensitivity and specificity. Also, the accuracy (Acc) and the area under the ROC curve (AUC) were computed.

5. Results and discussions

In this section we present the findings yielded by the experiments described above: classification of segments and detection of OSAH syndrome. In order to gain understanding and dive deeper into the problem of discriminating between apnea and hypopnea events, a preliminary qualitative study was carried out. For that, a structured dictionary for representing SaO_2 signals was learned by DAS-KSVD, following all procedures described in Section 4. As a result, DAS-KSVD yielded a structured dictionary $\Phi_D^{(I)} = [\Phi_{\mathrm{N}}\ \Phi_{\mathrm{A}}\ \Phi_{\mathrm{H}}]$ of size 128×150. Figure 5 shows the waveforms of some representative atoms corresponding to each one of the dictionaries Φ_{N} (upper), Φ_{A} (middle) and Φ_{H} (bottom). Several remarks are in order. First, it can be seen

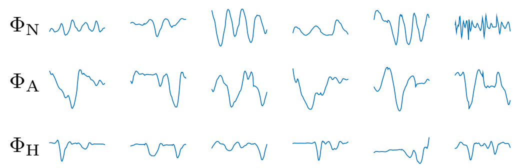

Figure 5: Typical atoms corresponding to Φ_{N} (top), Φ_{A} (middle) and Φ_{H} (bottom).

that each one of these dictionaries is composed of atoms capturing different types of class-related information. For instance, most atoms in Φ_{N} present quite regular waveforms associated to normal inhalation-exhalation changes in the oxygen saturation. On the other hand, atoms of Φ_{A}, representing apnea events, present local abrupt desaturations with clear sawtooth patterns. In pulse oximetry this is a typical behavior associated to the absence of respiratory airflow for a relatively long period of time. Finally, atoms of Φ_{H}, associated to hypopnea events, show essentially two desaturations, a large and a small one. We strongly believe that it is precisely this type of desaturation pattern in the atoms what allows for the identification of the hypopnea events.

5.1. Classification of segments

Features generated by DAS-KSVD were used to assess the ability of the MLP in classifying segments of SaO_2 signals. Table 1 shows the average normalized confusion matrix constructed using all testing samples of each fold in the cross validation (left) and a summary of all class-specific performance measures extracted from such a matrix (right). The elements in the diagonal of Table 1 (left) represent the normalized true positive rates. As it can be seen, the algorithm achieved true positive rates of 85.12%, 63.42% and 22.78% for the classes N, A and H, respectively, resulting in an overall accuracy of 57.11%. Note that if we were to limit our analysis only to the classes N and A (i.e. without tacking into account the third row and the third column of the confusion matrix), then the inter-class confusions would be relatively small. From the analysis of all these results several remarks can be drawn. First, DAS-KSVD constitutes a reasonable approach for classifying normal (breathing) and apnea events in pulse oximetry. Second, the results fall short of being good for detecting hypopnea events. In fact, more than half of them are misclassified as belonging to class N and more than one fourth are misclassified as belonging to class A. This last remark, however, is consistent with the results obtained using the *Sammon mapping* (see Section 2 and Figure 2) where we saw that the projections of class A and N segments into the first two most important attributes of the mapping are clearly well separated, while the projections of class H segments overlap the other two classes and present a wide variance.

Table 1: Average normalized multiclass confusion matrix obtained using DAS-KSVD for the classification of segments (left) and the corresponding performance measures (right).

Known \ Predicted	N	A	H
N	85.12	5.41	9.47
A	22.07	63.42	14.51
H	51.53	25.69	22.78

Class	Se (%)	Sp (%)	Pr (%)
N	85.12	63.20	53.63
A	63.42	84.45	67.09
H	22.78	88.01	48.72

In order to gain insight into the reasons why DAS-KSVD outperforms all other evaluated approaches for OSAH syndrome detection (see next section), we compared its performance with that of MDCS-BC in classifying segments of SaO_2 signals as containing an event or not (i.e. without tacking into account whether it is an apnea or a hypopnea). It is important to mention that MDCS-BC was chosen because it achieved the best performance among all previously developed methods. In order to analyze the performance of DAS-KSVD in the binary classification problem, we unified labels of segments belonging to the classes A and H which led to a new (and unique) class denoted by A+H. Table 2 shows a summary of the performance of DAS-KSVD and MDCS-BC using all testing samples. It is important to point out that, in this case, the target class is A+H. As it can be observed, although both methods yielded similar sensibility percentages, DAS-KSVD reached a significantly better specificity and precision percentages than MDCS-BC. In other words, DAS-KSVD has become more specific having fewer false positives than the other one. This clearly indicates that in the classification process, segments that were misclassified as N, are now correctly classified as H.

5.2. OSAH screening

In this article, besides analyzing the ability of DAS-KSVD to classify segments of SaO_2 signals into the classes N, A and H, we make use of these predictions to detect the presence of the pathology (according to a prescribed AHI diagnostic threshold). In that sense, a comparison between DAS-KSVD with many other state-of-the-art methods in the diagnosis of moderate to severe OSAH syndrome (AHI > 15) was performed. Table 3 shows a comparative summary of the results achieved by DAS-KSVD, MDCS-BC, MDCS-MC, MDAS-BC and MDAS-MC, and by the approaches

Table 2: Average performance measures for the classification of A+H events using both DAS-KSVD and MDCS-BC.

Method	Se (%)	Sp (%)	Pr (%)
DAS-KSVD	63.20	85.12	80.94
MDCS-BC	62.71	80.37	76.15

introduced by Chiner *et al.* [12], Vázquez *et al.* [13] and Schlotthauer *et al.* [2]. It is important to point out that all reported results are the mean value and the standard deviation in the cross validation.

Table 3: Average performance measures for moderate to severe OSAH screening using different methods.

Method	AUC	Se(%)	Sp(%)	Acc(%)
DAS-KSVD	$\mathbf{0.934 \pm 0.01}$	$\mathbf{89.10 \pm 2.14}$	$\mathbf{86.70 \pm 2.93}$	$\mathbf{87.90 \pm 1.65}$
MDCS-MC	0.924 ± 0.01	87.15 ± 3.13	88.23 ± 3.45	86.07 ± 6.37
MDAS-MC	0.891 ± 0.04	82.36 ± 9.07	86.09 ± 4.63	84.22 ± 5.07
MDCS-BC [3]	0.922 ± 0.02	87.89 ± 3.89	84.86 ± 3.84	86.38 ± 3.20
MDAS-BC [3]	0.878 ± 0.04	80.60 ± 9.07	81.83 ± 4.63	81.22 ± 5.07
Schlotthauer *et al.* [2]	0.921 ± 0.03	85.70 ± 5.68	86.00 ± 5.68	85.85 ± 3.79
Vázquez *et al.* [13]	0.909 ± 0.03	83.54 ± 6.72	88.10 ± 4.47	85.82 ± 2.76
Chiner *et al.* [12]	0.767 ± 0.04	65.57 ± 3.08	80.10 ± 5.56	72.84 ± 4.41

As can be observed in Table 3, DAS-KSVD outperforms all other evaluated approaches in their two versions: binary and multiclass. In particular, It was found that applying DAS-KSVD, the classifier yielded an average AUC value of 0.934 and sensitivity, specificity and accuracy of 89.10%, 86.70% and 87.90%, respectively. Also, the method leading to the second largest performances is the multiclass version of MDCS (MDCS-MC). When applying such a method, the classifier achieved an average AUC value of 0.924 and sensitivity, specificity and accuracy of 87.15%, 88.23% and 86.07%, respectively. In addition, if we compare the performance of DAS-KSVD with MDCS-MC, then it can be concluded that DAS-KSVD significantly enhances the overall performance achieved by MDCS-MC (assuming a p-value of 0.05).

It is also important to point out that, in most cases, multiclass classification methods outperform the binary ones in the detection of the pathology. For instance, the application of both MDCS-MC and MDAS-MC resulted in better performances than the ones achieved by their respective binary versions. More precisely, MDAS-MC obtained an average AUC value of 0.891 representing an improvement of 1.48% regarding MDAS-BC, which achieved an average AUC value of 0.878. Similarly, MDCS-MC yielded an improvement of 0.22% with respect to MDCS-BC. On the other hand, it becomes appropriate to mention that although MDAS-MC shows improvements regarding MDAS-BC, its overall performance remains still below that of MDCS-BC and Schlotthauer *et al.*

A more comprehensive analysis of Table 3 indicates that, although most discriminant methods achieved good results, DAS-KSVD outperforms all of them. The application of this method results in an average area under the ROC curve of 0.934 as well as sensitivity, specificity and accuracy of 89.10%, 86.70% and 87.90%, respectively. According to the original labels and taking into account a detection threshold of 15, each fold in the cross validation (95 studies) contains in average 73 and 22 pathological and normal (or healthy) patients, respectively. A 89.10% sensitivity indicates that of the 73 pathological cases, 65 were correctly detected (true positive) while 8 were false positive. On the other hand, an 86.70% specificity indicates that of the 22 healthy cases, 19 were appropriately identified (true negative) while only 3 were false negative. It is timely to note that for the 3 cases that DAS-KSVD yielded an AHI higher than 15, most events identified by the medical expert were precisely hypopneas and most of them were not associated with noticeable desaturations in the SaO_2 signal. This fact indicates that the final scoring process was carried out following the AASM criteria. Hence, this issue may be one of the causes that led to the misclassification of hypopneas, since its distribution highly overlaps with the one corresponding to normal breathing. Finally, if we look at the SaO_2 signal, there are a lot of cases where it becomes difficult to distinguish between normal breathing and hypopnea event.

In Table 4 we present an account of the computational costs associated to the implementation of the different methods. The programs were conducted using Matlab on a Lenovo V330-15IKB personal computer running Ubuntu 18.04 (64 bits) with an Intel Core i3 Processor 7th Generation @2.3GHz and 8GB of main memory. As it can be seen, the CPU times required by DAS-KSVD, MDCS and MDAS range between two and eight times those required by the other three methods. We emphasize, however that these computing times remain very low. In fact it takes about two seconds to analyze ten hours of data corresponding to a complete study.

The higher computational cost mentioned above, is highly compensated by better performances. In fact, since the previous methods do not include training from data

Table 4: Average computational costs (time used for computation for a single study during testing) associated to each one of the evaluated methods.

Method	Computational time (seconds)
DAS-KSVD	1.56 ± 0.057
MDCS-MC	1.40 ± 0.032
MDAS-MC	1.55 ± 0.016
MDCS-BC [3]	1.42 ± 0.008
MDAS-BC [3]	1.60 ± 0.013
Schlotthauer *et al.* [2]	0.75 ± 0.019
Vázquez *et al.* [13]	0.41 ± 0.007
Chiner *et al.* [12]	0.19 ± 0.019

and are based upon predefined rules, their performances are always lower (see bottom part of Table 3). On the other hand, although the methods based on learning from the data (middle and upper part of Table 3) are more difficult to implement and computationally more costly, they yield better performances and have the ability of adapting to new data and to changes in data recording conditions. Additionally the methods presented by our group allow distinguishing between apnea and hypopnea events.

6. Conclusions

In this article, with the objective of OSAH syndrome screening, we applied a previously developed method called DAS-KSVD to classify segments of SaO_2 signals into normal breathing and abnormal respiratory events in a multiclass scenario. It was found that the combined discriminant measure, which is used by DAS-KSVD in the process of building the structured dictionary, is capable of efficiently selecting the most discriminant atoms for each one of the classes. In addition, DAS-KSVD yielded a structured dictionary composed by three sub-dictionaries each one associated to a particular class. We evaluated DAS-KSVD in two different but related applications, namely, classification of abnormal respiratory events and detection of moderate to severe OSAH syndrome. Although it is a very challenging task, the proposed method has demonstrated to be efficient for automatically discriminating between apnea and hypopnea events in a multiclass scheme. To detect the presence or absence of events, DAS-KSVD resulted more specific than the most competitive binary-based approach (MDCS-BC). This improvement is due to the ability of DAS-KSVD in separating between (apnea or hypopnea) events and normal breathing. In a similar way, the application of DAS-KSVD led to the best reported performance in OSAH syndrome screening using a well known and publicly available database. This fact constitutes a strong evidence that our approach can be helpful in the development of new intelligent technologies for portable OSAH syndrome screening devices.

In the near future we plan to explore the use of Deep Learning tools to further enhance adaptation robustness and classification performance.

Acknowledgment

The authors would like to acknowledge the financial support of Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, of the Air Force Office of Scientific Research, AFOSR /SOARD, through Grant FA9550-14-1-0130, of the Universidad Nacional del Litoral through projects CAI+D 50120110100519 and CAI+D 5012011010 0525 and of the Universidad Tecnológica Nacional through projects TEUTIPA0004711TC and ICUTIPA0007803TC.

7. References

References

[1] A. Yadollahi, E. Giannouli, Z. Moussavi, Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals, Medical & Biological Engineering & Computing 48 (2010) 1087–1097.

[2] G. Schlotthauer, L. E. Di Persia, L. D. Larrateguy, D. H. Milone, Screening of obstructive sleep apnea with empirical mode decomposition of pulse oximetry, Medical Engineering & Physics 36 (2014) 1074–1080.

[3] R. Rolon, L. Larrateguy, L. D. Persia, R. Spies, H. Rufiner, Discriminative methods based on sparse representations of pulse oximetry signals for sleep apnea–hypopnea detection, Biomedical Signal Processing and Control 33 (2017) 358–367.

[4] J. Hedner, L. Grote, M. Bonsignore, W. McNicholas, P. Lavie, G. Parati, P. Sliwinski, F. Barbé, W. De Backer, P. Escourrou, I. Fietze, J. Kvamme, C. Lombardi, O. Marrone, J. Masa, J. Montserrat, T. Penzel, M. Pretl, R. Riha, D. Rodenstein, T. Saaresranta, R. Schulz, R. Tkacova, G. Varoneckas, A. Vitols, H. Vrints, J. Zielinski, The european sleep apnoea database (esada): report

from 22 european sleep laboratories, European Respiratory Journal 38 (2011) 635–642.

[5] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, et al., Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events, Journal of clinical sleep medicine 8 (2012) 597–619.

[6] S. Khadadah, P. Lachapelle, S. Pamidi, A. E. Olha, A. Benedetti, R. Kimoff, Does Scoring Of Autonomic Hypopneas Improve Clinical Decision Making In Obstructive Sleep Apnea?, American Thoracic Society, pp. A2606–A2606.

[7] C. Gamaldo, L. Buenaver, O. Chernyshev, S. Derose, R. Mehra, K. Vana, H. K. Walia, V. Gonzalez, I. Gurubhagavatula, Evaluation of clinical tools to screen and assess for obstructive sleep apnea, Journal of Clinical Sleep Medicine 14 (2018) 1239–1244.

[8] M. Uddin, C. Chow, S. Su, Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review, Physiological measurement 39 (2018) 03TR01.

[9] F. del Campo, A. Crespo, A. Cerezo-Hernández, G. C. Gutiérrez-Tobal, R. Hornero, D. Álvarez, Oximetry use in obstructive sleep apnea, Expert review of respiratory medicine 12 (2018) 665–681.

[10] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, T. Penzel, A review of obstructive sleep apnea detection approaches, IEEE journal of biomedical and health informatics 23 (2019) 825–837.

[11] P. I. Terrill, A review of approaches for analysing obstructive sleep apnoea-related patterns in pulse oximetry data, Respirology (2019).

[12] E. Chiner, J. Signes-Costa, J. M. Arriero, J. Marco, I. Fuentes, A. Sergado, Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies?, Thorax 54 (1999) 968–971.

[13] J.-C. Vázquez, W. H. Tsai, W. W. Flemons, A. Masuda, R. Brant, E. Hajduk, W. A. Whitelaw, J. E. Remmers, Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea, Thorax 55 (2000) 302–307.

[14] D. Köhler, B. Schönhofer, How important is the differentiation between apnea and hypopnea?, Respiration 64 (1997) 15–21.

[15] B. L. Koley, D. Dey, Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary svm classifiers, Measurement 46 (2013) 2082–2092.

[16] M. Halevi, E. Dafna, A. Tarasiuk, Y. Zigel, Can we discriminate between apnea and hypopnea using audio signals?, in: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 3211–3214.

[17] G. Medic, M. Wille, M. E. Hemels, Short-and long-term health consequences of sleep disruption, Nature and science of sleep 9 (2017) 151–161.

[18] E. García-Díaz, E. Quintana-Gallego, A. Ruiz, C. Carmona-Bernal, A. Sánchez-Armengol, G. Botebol-Benhamou, F. Capote, Respiratory polygraphy with actigraphy in the diagnosis of sleep apnea-hypopnea syndrome, Chest 131 (2007) 725–732.

[19] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, P. W. Wahl, The Sleep Heart Health Study: design, rationale, and methods, Sleep 20 (1997) 1077–1085.

[20] L. Van Der Maaten, E. Postma, J. Van den Herik, Dimensionality reduction: a comparative, J Mach Learn Res 10 (2009) 13.

[21] M. Elad, Sparse and Redundant Representations, Springer-Verlag New York, 2010.

[22] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM review 43 (2001) 129–159.

[23] S. G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (1993) 3397–3415.

[24] J. Tropp, A. Gilbert, Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, IEEE Transactions on Information Theory 53 (2007) 4655–4666.

[25] S. K. Sahoo, A. Makur, Signal recovery from random measurements via extended orthogonal matching pursuit, IEEE Transactions on Signal Processing 63 (2015) 2572–2581.

[26] M. Aharon, M. Elad, A. Bruckstein, KSVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE Transactions on Signal Processing 54 (2006) 4311–4322.

[27] Q. Zhang, B. Li, Discriminative K-SVD for dictionary learning in face recognition, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2691–2698.

[28] D. S. Pham, S. Venkatesh, Joint learning and dictionary construction for pattern recognition, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

[29] Z. Jiang, Z. Lin, L. Davis, Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2013) 2651–2664.

[30] N. Rao, F. Porikli, A clustering approach to optimize online dictionary learning, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1293–1296.

[31] X. Chen, J. Li, D. Zou, Q. Zhao, Learn Sparse Dictionaries for Edit Propagation, IEEE Transactions on Image Processing 25 (2016) 1688–1698.

[32] Z. Ataee, H. Mohseni, Structured dictionary learning using mixed-norms and group-sparsity constraint, The Visual Computer (2019) 1–14.

[33] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 210–227.

[34] R. E. Rolon, L. E. D. Persia, R. D. Spies, H. L. Rufiner, A multi-class structured dictionary learning method using discriminant atom selection, arXiv preprint arXiv:1812.01389 (2018) 1–18.

[35] B. K. Lind, J. L. Goodwin, J. G. Hill, T. Ali, S. Redline, 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 7 (2003) 13–24.

[36] F. Lestussi, L. Di Persia, D. Milone, Comparison of on-line wavelet analysis and reconstruction: with application to ecg, in: 5th International Conference on Bioinformatics and Biomedical Engineering, IEEE, pp. 1–4.

[37] R. Rolon, I. E. Gareis, L. E. Di Persia, R. D. Spies, H. L. Rufiner, Complexity-based discrepancy measures applied to detection of apnea-hypopnea events, Complexity 2018 (2018).

[38] J. A. Swets, ROC analysis applied to the evaluation of medical imaging techniques, Investigative Radiology 14 (1979) 109–121.