Cite this document as:

Daniel Álvarez, Roberto Hornero, J. Víctor Marcos, Félix del Campo. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. Medical Engineering & Physics, vol. 34 (8), pp. 1049-1057, 2012.

DOI: 10.1016/j.medengphy.2011.11.009

Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis

Daniel Álvarez[1], Roberto Hornero[1], J. Víctor Marcos[1], and Félix del Campo[2]

[1]Biomedial Engineering Group, E.T.S.I. Telecomunicación, University of Valladolid, Paseo de Belén 15, 47011, Spain.

[2]Hospital Universitario Pío del Río Hortega de Valladolid, c/ Dulzaina 2, 47013, Valladolid, Spain.

AUTHOR'S ADDRESS: Daniel Álvarez

E.T.S.I. de Telecomunicación

University of Valladolid

Paseo de Belén 15

47011 - Valladolid (Spain)

Phone: +34 983 42300, ext. 4716

Fax: +34 983 423667

E-mail: dalvgon@gmail.com

Abstract

Nocturnal pulse oximetry (NPO) has demonstrated to be a powerful tool to help in obstructive sleep apnoea (OSA) detection. However, additional analysis is needed to use NPO alone as an alternative to nocturnal polysomnography (NPSG), which is the gold standard for a definitive diagnosis. In the present study, we exhaustively analysed a database of blood oxygen saturation (SpO_2) recordings (80 OSA-positive and 160 OSA-negative) to obtain further knowledge on the usefulness of NPO. Population set was randomly divided into training and test sets. A feature extraction stage was carried out: 16 features (time and frequency statistics and spectral and nonlinear features) were computed. A genetic algorithms (GAs) approach was applied in the feature selection stage. Our methodology achieved 87.5% accuracy (90.6% sensitivity and 81.3% specificity) in the test set using a logistic regression (LR) classifier with a reduced number of complementary features (3 time domain statistics, 1 frequency domain statistic, 1 conventional spectral feature and 1 nonlinear feature) automatically selected by means of GAs. Our results improved diagnostic performance achieved with conventional oximetric indexes commonly used by physicians. We concluded that GAs could be an effective and robust tool to search for essential oximetric features that could enhance NPO in the context of OSA diagnosis.

Keywords: Obstructive sleep apnoea syndrome, nocturnal pulse oximetry, feature extraction, feature selection, genetic algorithms, logistic regression.

1. Introduction

The obstructive sleep apnoea (OSA) syndrome is a sleep-related disorder characterised by frequent breathing pauses, which lead to deep oxyhaemoglobin desaturations, blood pressure and heart rate acute changes, increased sympathetic activity and cortical arousals [1]. A wide variety of significant consequences affects people suffering from OSA including hypersomnolence, neurocognitive dysfunction, metabolic deregulation or respiratory failure [2]. Moreover, OSA is frequently linked with conditions associated to the main causes of mortality in adults, such as hypertension, stroke or myocardial infarction [2,3]. It is estimated that approximately 20% of adults have at least mild OSA and 7% of adults have moderate-to-severe OSA [4]. Unlike its high prevalence, 90% of cases in men and 98% of cases in women may go undiagnosed for many years [2].

The gold standard method for a definitive OSA diagnosis is in-hospital, technician-attended nocturnal polysomnography (NPSG) [5]. However, this methodology is labor-intensive, expensive and time-consuming [5]. The main alternatives to NPSG are aimed at reducing the number of recordings to be analysed, focusing on the use of portable monitoring [5,6]. Several studies have been developed to assess automated analysis of single cardiorespiratory-related signals [7–15]. Single-lead electrocardiogram (ECG) [7,8], single-channel airflow (AF) [9–11] and blood oxygen saturation (SpO_2) from nocturnal pulse oximetry (NPO) [12–14] have been predominantly studied. Previous studies based on single-lead ECG do not use portable devices [7,8]. Single-channel AF-based studies commonly use the respiratory disturbance index (RDI) to detect OSA [9]. However, the RDI often include all other abnormal respiratory events [9] and portable devices using nasal pressure or thermal sensors are less accurate than standard pneumotachometers [6]. On the other hand, portable NPO are handier, less expensive, easy-to-use and highly reliable [12]. SpO_2 from NPO could provide relevant information to detect apnoeas, making NPO an essential tool to obtain

simple ambulatory methodologies aimed at reducing waiting lists [6,12]. However, some limitations decrease its ability as a single tool for OSA diagnosis at patient's home [12]. Regarding the diagnosis of OSA syndrome, the American Academy of Sleep Medicine (AASM) suggests that portable monitoring should not be used in patient groups with significant comorbid medical conditions, patients suspected of having other sleep disorders and for general screening of asymptomatic populations [6]. Furthermore, the AASM recommends that the use of portable monitoring should be limited to patients with a high pre-test probability of moderate-to-severe OSA based on clinical evaluation [6]. Thus, there is still a great demand on new studies that can provide additional information to improve the usefulness of SpO_2 from NPO to help in OSA diagnosis.

In the present study, feature extraction and feature selection procedures were carried out to analyse SpO_2 recordings. A large set of features was developed to obtain as much information as possible from oximetry signals. Statistical, spectral and nonlinear analyses were carried out to compose an initial feature set. Previous studies showed the usefulness of multivariate pattern analysis in OSA diagnosis [14,16]. Additionally, feature selection techniques could be very useful to derive a smaller but optimal subset for classification purposes. There are many potential benefits of variable selection after feature extraction, both computational and in prediction performance [17,18]. Previous studies on the usefulness of SpO_2 recordings in the context of OSA diagnosis commonly assessed single features or small subsets [13,19,20]. When larger feature sets are evaluated, a feature selection stage is not implemented [21–23] or suboptimal variable selection is carried out [14,16]. In this research, we hypothesised that an exhaustive analysis of the search space by means of GAs could provide further knowledge on SpO_2 dynamics. GAs provide a parameter optimisation strategy that has demonstrated to be a powerful tool for variable selection [24]. GAs were used to find the optimum feature subset with a given number of

variables. Since the goal of our study was to maximise OSA diagnostic accuracy, we used the classification performance of a predefined classifier to guide the search. A logistic regression (LR) classifier was used to investigate classification performance. Our study was aimed at enhancing diagnostic ability of NPO to improve diagnostic accuracy reached by conventional oximetric indexes. To achieve this goal, the present research focuses on assessing the usefulness of GAs to provide suitable reduced oximetric feature subsets in the context of OSA diagnosis from a wide feature space of oximetry measures: time *vs*. spectral and linear *vs*. nonlinear.

2. Subjects and signals under study

The population set consisted of 240 subjects (186 males and 54 females) derived to the Sleep Unit of the Hospital Universitario Pío del Río Hortega of Valladolid (Spain). All subjects showed high suspicion of suffering from OSA based on clinical evaluation. Complete in-hospital NPSG studies were carried out from midnight to 08:00 AM. Patients were monitored using a polysomnograph Alice 5 by Respironics (Philips Healthcare, The Netherlands). Rechtschaffen and Kales standard rules were used to study sleep architecture. The standard apnoea – hypopnoea index (AHI) was used to diagnose OSA and characterise its severity [5]. According to the AASM rules [25], apnoea was defined as a drop in the peak thermal airflow sensor greater than or equal to 90% from baseline lasting at least 10 s, whereas hypopnoea was defined as a nasal pressure signal excursion drop greater than or equal to 50% during at least 10 s, accompanied by a desaturation greater than or equal to 3% from pre-event baseline and/or the event is associated with an arousal. An AHI $\geq$ 10 events per hour (e/h) was considered as diagnostic of OSA.

A positive diagnosis of OSA was confirmed in 160 patients, with an average AHI of 36.6 $\pm$ 25.7 e/h. The remaining 80 subjects composed the OSA-negative group, with an

average AHI of 3.9 ± 2.4 e/h. The initial population set was randomly divided into two independent groups: the training set (40%) and the test set (60%). The training set was used to obtain different feature subsets and LR models from the variable selection procedure. On the other hand, the test set was used to assess the optimum models from the training stage, in order to validate our methodology. Table 1 displays demographic and clinical features for the initial, training and test population groups.

INSERT TABLE 1 ROUND HERE

The polysomnograph equipment used in the present study included a Nonin PureSAT® pulse oximeter (Nonin Medical Inc., USA), with 3 s or faster averaging interval at a minimum heart rate of 60 beats per minute or greater. Thus, the NPO equipment outperforms the recommendations of the Task Force on Respiratory Scoring of the AASM, which requires a maximum signal averaging time of ≤ 3 seconds at a heart rate of 80 beats per minute or more [6, 25]. A finger probe with a pair of red (for measuring deoxygenated haemoglobin) and infrared (for measuring oxygenated haemoglobin) light sources was used for measuring peripheral SpO_2. A commercial PureLight® sensor (Nonin Medical Inc., USA) was used, with high-performance in the presence of motion artefacts and low perfusion for adult, paediatric and neonate patients. SpO_2 was recorded at a sampling rate of 1 Hz. All SpO_2 recordings from NPSG were saved to separate files and processed offline. SpO_2 signals presented zero samples at the beginning of the acquisition process and drops to zero due to patient movements along the recording time. An automatic pre-processing stage was carried out to remove these artefacts.

3. Methods

The present study was divided into three main stages: feature extraction, feature selection and classification. In the feature extraction stage, oximetric recordings were exhaustively analysed to parameterise SpO_2 dynamics from NPO. The outcome of this step was a wide oximetric feature set, which was the input to the subsequent feature selection stage. Genetic algorithms (GAs) were evaluated for variable selection. Additionally, a logistic regression (LR) classifier was used in the classification stage. The Matlab® software version 7.6 (R2008a) was used to implement feature extraction methods and to develop the feature selection (*Genetic Algorithm and Direct Search Toolbox*^{TM}) and classification (*Statistics Toolbox*^{TM}) stages. The full-model (all variables are included) and all single-feature models were also computed. Additionally, we applied this methodology to a conventional oximetric feature set composed of oximetric indexes commonly used by physicians. Training and test groups were used to obtain the optimum model and to assess its performance on an independent population, according to the training/test paradigm in the feature selection and classification procedure.

3.1. Feature extraction stage

Oximetric recordings were parameterised by means of 16 features from four feature subsets: time domain statistics, frequency domain statistics, spectral and nonlinear features.

3.1.1. Time domain statistics

First to fourth-order statistical moments were computed as their average over the total number of equal-length non-overlapping 512-sample histograms from each overnight SpO_2 profile in the time domain as follows [26]:

$$first\ statistical\ moment \equiv E[x] = \mu = \frac{1}{N}\sum_{n=1}^{N} x_n \ , \qquad (1)$$

$$n^{th} - order\ statistica\ l\ moment \equiv E\left[(x-\mu)^n\right]. \quad (2)$$

Arithmetic mean (*M1t*), variance (*M2t*), skewness (*M3t*) and kurtosis (*M4t*) in the time domain were computed to quantify central tendency, amount of dispersion, asymmetry and peakedness, respectively [14, 22].

3.1.2. Frequency domain statistics

The power spectral density (PSD) of each oximetric recording was estimated applying the Welch's method. A 512-sample Hanning window with 50% overlap (102.70 ± 8.44 512-sample segments per recording) and 1024-points discrete Fourier transform (DFT) were used. These input parameters improved performance and statistical characteristics (bias and variance) of the PSD estimate [27]. The following statistics were computed:

i) First to fourth-order moments (*M1f–M4f*) in the frequency domain. The amplitude (W/Hz) of the PSD function at each single spectral component was used to obtain the histogram in the frequency domain [14,22].

ii) Median frequency (*MF*). *MF* is defined as the spectral component which comprises 50% of the total signal power, with higher values corresponding to signals with significant spectral components at higher frequencies [28]:

$$0.5 \sum_{f=0\,Hz}^{0.5\,f_S} PSD(f) = \sum_{f=0\,Hz}^{MF} PSD(f). \quad (3)$$

iii) Spectral entropy (*SE*). *SE* is a disorder quantifier related to the flatness of the spectrum. Higher *SE* values correspond to signals with broader spectral content [29]:

$$SE = -\sum_j p_j \ln\left(p_j\right), \quad (4)$$

where p_j is the normalised value of the PSD at each frequency component.

3.1.3. Spectral features

The frequency band 0.014 – 0.033 Hz proposed by Zamarrón *et al*. was studied [13]. The minimum and the maximum apnoea cycle length define this range, which shows a power increase due to apnoea events in SpO_2 recordings from OSA patients [13]. The following spectral features were derived:

i) Total spectral power (P_T). It is computed as the total area under the PSD [13]:

$$P_T = \sum_{f=0Hz}^{0.5 f_S} PSD(f) . \quad (5)$$

ii) Peak amplitude (*PA*) in the apnoea frequency band. It is the local maximum of the SpO_2 spectral content in the apnoea frequency range [13]:

$$PA = \max_{PSD}\{PSD(f)\},\ f(Hz) \in [0.014, 0.033] . \quad (6)$$

iii) Relative power (P_R) in the apnoea frequency band. It is the ratio of the area enclosed under the PSD in the apnoea frequency band to the total signal power [13]:

$$P_R = \frac{\sum_{f=0.014}^{0.033} PSD(f)}{\sum_{f=0}^{0.5 f_S} PSD(f)} . \quad (7)$$

3.1.4. Nonlinear features

i) Sample entropy (*SampEn*). *SampEn*(*m, r, N*) is a family of statistics defined to quantify irregularity, with larger values corresponding to more irregular data [30]:

$$SampEn(m, r, N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right], \quad (8)$$

where A^m and B^m are the average number of (m)-length and (m+1)-length segments $X_m(i)$ ($1 \le i \le N\text{-}m\text{+}1$) with $d[X_m(i),X_m(j)] \le r$ ($1 \le j \le N\text{-}m$, $j \ne i$), respectively, and

$$d\left[X_m(i),X_m(j)\right] = \max_{k=0,\ldots,m-1}\left(\left|x(i+k)-x(j+k)\right|\right). \quad (9)$$

In the present study, we used the recommended input parameters m = 1 and r = 0.25 times SD [30]. These values have demonstrated to be the optimal input parameters in the context of SpO_2 analysis from NPO [31].

ii) Central tendency measure (*CTM*). *CTM* provides a variability measure from second order difference plots, assigning larger values to signals with lower variability [32]:

$$CTM = \frac{1}{N-2}\sum_{i=1}^{N-2}\delta(d_i), \quad (10)$$

where

$$\delta(d_i) = \begin{cases} 1 \text{ if } \left[\left(x(i+2)-x(i+1)\right)^2 + \left(x(i+1)-x(i)\right)^2\right]^{1/2} < \rho \\ 0 \text{ otherwise} \end{cases}. \quad (11)$$

In the present study, we applied the recommended radius $\rho = 1$ to compute *CTM* [33]. Previous studies have shown that $\rho = 1$ is the optimal value in the context of SpO_2 analysis from NPO [33].

iii) Lempel – Ziv complexity (*LZC*). *LZC* is a non-parametric measure of complexity, with larger values corresponding to high complexity data [34]. The original signal is codified into a binary sequence using a threshold T_d. The complexity counter $c(n)$ is increased every time a new subsequence is encountered [35]:

$$LZC = \frac{c(n)}{b(n)}, \quad (12)$$

where b(n) is a normalisation parameter:

$$b(n) \equiv \frac{n}{\log_2(n)}. \quad (13)$$

In this study, the median value was used as threshold T_d. A previous study by our group showed that the median value was suitable to capture SpO_2 changes after binary codification [35].

All features in the time domain were computed dividing every SpO_2 signal into non-overlapping segments of 512 samples, i.e. a value was computed per every signal segment (51.33 ± 4.22 512-sample segments per recording) for each feature. Finally, we averaged over the total number of segments to obtain a single value per subject.

3.2. Feature selection stage

A GA is a searching process based on the laws of evolution and natural selection [36]. A population from a GA optimisation procedure comprises a group of chromosomes or candidate solutions, which are modified iteratively: a particular group of chromosomes are selected from an initial population to generate the offspring by means of predefined genetic operations. The offspring replaces chromosomes in the current population based on certain replacement strategies [36]. The optimisation process is carried out in cycles called generations. While conventional approaches just evaluate and improve a single feature subset, GAs intensively analyse the whole feature space by modifying and improving a group of subsets at the same time. In this study, we extensively assessed the suitability of GAs for feature selection in the context of OSA diagnosis from NPO. To achieve this goal, our feature selection methodology took into account the number of features, in order to explore each *k*-dimensional subspace: GAs were applied to obtain the optimal feature subset for a given number of input features, from 2 to *p*–1, where *p* is the dimension of the original

feature space. An optimal subset was defined as the group of input variables to a LR model that achieved the highest classification performance.

An individual or chromosome from the population is just a combination of a predetermined number of features from NPO [37]. A feature subset in the GA search space is codified with a finite binary sequence, where the *k*-th bit denotes the absence (0) or the presence (1) of the *k*-th feature. The classification accuracy of a LR model is used as the objective value to assess each chromosome performance and to achieve parent selection. A fitness function is used to map each objective value to a proportional predefined fitness interval. In this study, a common proportional fitness scaling function was used. Additionally, conventional roulette and tournament schemes were used as parent selection strategies. One-point crossover was applied to produce offspring: a crossover point was randomly selected and the portions of both parents beyond this point were exchanged to form the offspring [36]. The crossover operation rate is controlled by the probability term P_c, which usually assumes high values, close or equal to one [24]. In the present research, P_c values between 0.5 and 0.9 were applied [24,38]. Uniform mutation was applied to introduce variations into the offspring: multiple bits uniformly distributed over the range of the chromosome are replaced if a probability rate is passed. P_m is the probability of switching bits in the chromosome. In this study, P_m values between 0.01 and 0.09 were applied [24]. The percentage of individuals in the old population preserved after each generation (elite) varied between 0% and 25% in the present study. All implementations of GAs were computed during 100 generations [37]. These parameters are commonly used in the context of feature selection to effectively explore the searching space [24,38]. We applied GAs to study our feature set composed of 16 time, frequency, linear and nonlinear features. Additionally, we applied feature selection by means of GAs to the conventional oximetric feature set.

3.3. Conventional oximetric indexes

The following conventional indexes were computed offline from each SpO_2 recording in our database:

i) Lowest saturation during the recording time (LO_2) [19].

ii) Cumulative time spent below a saturation of 90% (*CT90*) [19].

iii) Saturation impairment time (*SIT90*), which quantified the area of the SpO_2 profile under a saturation of 90% to measure the severity of desaturations [39].

iv) The number of dips in the SpO_2 signal greater than or equal to 3% per hour of recording (*ODI3*). A desaturation event was defined as a decrease in $SpO_2 \geq 3\%$ from baseline for at least 10 s and at a rate > 0.1%/s, returning within 60 s to a 1%-interval of the initial value [21].

v) The delta index (Δ), which measured SpO_2 variability independently of the definition of desaturation. The Δ index was computed as the sum of the absolute variations between the first and the last sample of all 12-s intervals in the SpO_2 recording, normalised by the number of intervals [40].

3.4. Statistical analysis

The Kolmogorov-Smirnov test was used to assess the normal distribution of the features involved in the study. Homogeneity of variances was assessed by means of the Levene's test. Normality and homoscedasticity could not be verified for all the variables under study. Statistical differences were evaluated by means of the non-parametric Mann-Whitney U test and LR was used to assess classification performance.

Training and testing were carried out using two independent population groups. The training set was used to perform the feature selection process, where a number of LR models were computed: 16 one-feature LR models from our original feature set from oximetry, 5

one-feature LR modes from the conventional oximetric indexes set, 2 LR full-models composed of all features from both feature sets, and 17 models from the GAs feature selection procedure applied to our original feature space (14 optimum models composed of 2 to 15 variables) and to the conventional oximetric indexes space (3 optimum models composed of 2 to 4 variables). There were two stopping criteria for the LR algorithm: changes in the coefficients of the model were lower than a predefined tolerance or the maximum number of iterations was reached. In the present study, the algorithm always stopped before reaching the maximum number of iterations. Thus, every LR model was optimised using the same training set from our population under study. The diagnostic accuracy in the training set was used to select the most significant LR models from oximetry. Every model was subsequently assessed in the test set according to the training/test paradigm in the feature selection and classification procedure. Sensitivity (OSA-positive patients correctly classified), specificity (OSA-negative subjects rightly classified) and accuracy (the total percentage of subjects correctly classified) were computed to quantify classification performance.

4. Results

4.1. Training Set

Figs. 1 (a) and (b) display SpO_2 profiles for a common OSA-negative subject and a representative OSA-positive patient in the time and frequency domains, respectively. Additionally, Figs. 1 (c) – (r) display feature values of every non-overlapping 512-sample segment for both signals. These plots show how apnoea events affect the instantaneous and the average feature values of OSA-positive patients. The three main desaturations periods in the SpO_2 recording of the OSA-positive patient caused significant changes in *M1t, M2t, M1f, M2f, P_T, PA* and *CTM*, which achieved instantaneous values that significantly differ from

their average. In addition, *M4t* also showed significant changes corresponding to constant periods from the SpO_2 profile of the OSA-negative subject. Average values (mean ± SD) of all 16 oximetry features for the OSA-positive and the OSA-negative groups in the training set are displayed in Table 2. On average, SpO_2 recordings from OSA-negative subjects in the time domain had significantly lower variance and higher kurtosis (peakedness) than OSA-positive patients. In the frequency domain, OSA-positive patients had significantly higher mean and variance and lower skewness (symmetry) and kurtosis than OSA-negative subjects. Similarly, OSA-positive patients presented significantly higher *MF* and *SE*. *PA*, P_T and P_R from OSA-positive patients were also significantly higher than conventional spectral measures from the OSA-negative group. Finally, OSA-positive patients had significantly lower *CTM* (higher variability) than non-OSA subjects.

Table 3 summarises the results of conventional oximetric indexes in the training set. Average (mean ± SD) cumulative desaturation time, severity, number of desaturations and variability were significantly higher in the OSA-positive group. The average minimum saturation during the recording time was significantly lower in the OSA-positive group.

INSERT TABLES 2 AND 3 AND FIGURE 1 ROUND HERE

Tables 4 to 7 summarise classification performance in the training set of all LR models involved in the study. Tables 4 and 5 show diagnostic assessment of every one-feature LR model composed of each single feature from our original feature set and from the conventional indexes feature set, respectively. LR models composed of *PA* and *CTM* achieved the highest accuracies (84.4% and 83.3%, respectively) in our feature set from oximetry, whereas *ODI3* and Δ reached the highest performance (88.5% accuracy) in the conventional oximetric indexes set. GAs provided a superset of optimum feature subsets, one

for each predefined number of features. Tables 6 and 7 summarise the performance assessment of optimum LR models from the GAs feature selection procedure in the training set. Additionally, classification results of the LR full-models from our oximetric feature set (composed of 16 features) and from the conventional oximetric indexes set (composed of 5 features) are provided. A significant increase in the overall diagnostic accuracy can be observed for LR models from the feature selection process based on GAs in both feature sets from oximetry. Additionally, the performance achieved by both LR full-models was also improved. Optimum subsets with 6 to 14 features from our original feature set achieved the highest diagnostic accuracy (95.8%) in the training set, with sensitivities and specificities ranging from 96.9% to 98.4% and 90.6% to 93.8%, respectively. Fig. 2 (a) shows the accuracy of the best individual and the average accuracy of the population *vs.* the number of iterations in a realisation of the GA, using $P_m = 0.9$, $P_c = 0.09$ and Elite = 4. Fig. 2 (b) shows the number of variables selected at each iteration.

INSERT TABLES 4 TO 7 AND FIGURE 2 ROUND HERE

4.2. Test Set

Subjects in the test group were used to assess classification performance of optimum LR models and feature subsets from NPO selected in the training set. Table 8 show classification statistics of optimum LR models with 6 to 14 features previously selected. The LR model with 6 features from GAs achieved the highest performance in the test group, with 90.6% sensitivity, 81.3% specificity and 87.5% accuracy. We can observe that the simplest LR model (the model with the lowest number of features in the test group) provided the highest accuracy. Additionally, all feature subsets from our original oximetric feature space

were included in the overall optimum feature subset: time domain statistics (*M1t*, *M3t*, *M4t*), frequency domain statistics (*MF*), conventional spectral measures (*PA*) and nonlinear measures (*CTM*).

INSERT TABLE 8 ROUND HERE

---

5. Discussion

In the present research, we analysed a database of SpO_2 signals from NPO. The methodology was divided into three main stages: feature extraction, variable selection and classification. In the feature extraction stage, SpO_2 signals were parameterised in time and frequency domains using linear and nonlinear measures. In the variable selection stage, an exhaustive analysis of the original feature set was carried out. A LR classifier was used to assess classification ability. The same methodology was applied to a set of conventional oximetric indexes commonly used by physicians. Independent training and test sets were randomly obtained from the database to develop and validate the methodology. Our results supported the suitability of GAs for feature selection in the context of OSA diagnosis. All *k*-feature LR models from GAs achieved large accuracy in the training set (close or higher than 90.0%) for both oximetric feature spaces under study, significantly improving the performance reached by every one-feature LR model. Moreover, the GAs feature selection procedure provided suitable reduced subsets that improve the performance achieved by the LR full-model, both for our original feature set and for the conventional indexes feature set. A reduced subset of 6 complementary measures (*M1t*, *M3t*, *M4t*, *MF*, *PA* and *CTM*) provided the highest classification ability in the test set (87.5% accuracy). LR models with larger number of features had lower generalisation ability, showing a larger performance

decrement in the test set. Oximetric features from our optimum model together account for the information from different complementary analyses (linear and nonlinear) and domains (time and frequency).

Diagnostic ability of NPO has been improved during the last years by means of automated analysis of SpO_2 recordings. Conventional oximetric indexes based on depth, time and number of desaturations, have been widely applied. Average and minimum SpO_2, CT indexes, SIT indexes and ODIs provide useful information on changes in SpO_2, reaching sensitivities and specificities ranging from 32% – 98% and 57% – 97%, respectively [12,39]. However, different limitations have been documented [21]: CT indexes do not achieve high diagnostic accuracy, there is not a standard definition for oxygen desaturation, correlation with AHI is not high and their sensitivity and specificity greatly vary among studies. On the other hand, the Δ index has been traditionally used to quantify SpO_2 variability independently of the definition of desaturation, achieving sensitivities and specificities ranging from 40% to 91% and from 59% to 88%, respectively [21,40,41]. In the present study, we assessed the diagnostic performance of our methodology and conventional indexes using the same database to properly drawn conclusions. Our results have shown that slight improvement was obtained when automatically selected indexes from GAs were added to a LR model. Additionally, our optimum LR model from GAs improved the diagnostic accuracy of common oximetric indexes in the training set (95.8% *vs*. 91.7%).

Some features in our original feature set have been previously assessed within the OSA diagnosis context. In the work by Zamarrón *et al*. [13], the limits of the apnoea frequency band 0.014 – 0.033 (Hz) and the usefulness of conventional measures based on the peak amplitude and relative power were studied. Similarly, in the studies by Hornero *et al*. [31] and Álvarez *et al*. [33,35], the optimum input parameters of nonlinear measures of irregularity, variability and complexity were determined in the context of OSA diagnosis.

However, the diagnostic performance of these methods was not assessed on an independent database. In the present study, we applied our previous knowledge on spectral and nonlinear analysis of SpO_2 recordings to apply these methods on an independent population set from a different Sleep Unit. Thus, the present study allowed us to further assess the strength of these measures in the context of OSA diagnosis. Our current results showed some differences with previous studies. Regarding conventional spectral measures, we obtained more balanced sensitivity *vs*. specificity pairs, although the overall accuracy is preserved. The *CTM* also achieved a similar performance. On the other hand, the accuracy of *SampEn* and *LZC* decreased significantly, from 80 – 84% to 57%. In our opinion, the decrease on the overall performance of both methods could be due to the higher sampling frequency of our new oximetry equipment (1 Hz) compared to the sampling frequency used to record the SpO_2 signal in previous studies (0.2 Hz) [13,31, 33,35]. Nevertheless, our optimum model from GAs improved the diagnostic accuracy of every single feature in the training set. These results suggested that features from our proposed time statistics, frequency statistics, spectral and nonlinear feature sets provided complementary information on SpO_2 dynamics, which together improved NPO ability in the context of OSA diagnosis.

Regarding training results, we can observe that diagnostic accuracy increases with added features, from 91.7% (2 features) to 95.8% (6 to 14 features). Then, the performance decreased to 93.8% accuracy when 15 features are used. This could be due to the course of dimensionality: the performance of a classifier is degraded when there are too many parameters relative to the size of the training data [17]. The subsequent testing process is aimed at further assessing these LR models, in order to avoid bias and over-fitting. LR models with the highest performance in the training set (9 models from GAs) were selected according to the training/test paradigm for feature selection and classification. The LR model with the smallest number of features (6 independent variables) achieved the highest accuracy

in the test set. In our point of view, large accuracies are reached in the training set increasing the model complexity, i.e. adding new features. However, the greater complexity (and accuracy), less generalisation power. Our results suggested that the less complex LR model can explain essential information about SpO_2 dynamics linked with apnoeas resulting in a more stable model, which reached high diagnostic performance on independent datasets.

Other researches attempted to improve OSA diagnosis by means of multivariate analysis. Using CTs and Δ indexes, 88.0% sensitivity and 70.0% specificity were reached from stepwise linear regression [41]. Multivariate adaptive regression splines were applied to ODIs, CTs and Δ indexes, obtaining 90% sensitivity and 70% specificity [21]. A sensitivity of 82% and a specificity of 84% were obtained using LR and spectral features from the high frequency range [23]. Higher performance (91.1% sensitivity and 82.6% specificity) were obtained applying linear discriminant analysis to P_T, PA and P_R spectral features [42]. An accuracy of 100% was reported applying quadratic discriminant analysis to a wide set of time and spectral measures from RR-interval series when borderline patients were removed from the study [8]. Recent studies in our group applied dimensionality reduction and stepwise feature selection procedures before classification [14,16]. PCA was applied to a small set of 3 spectral and 3 nonlinear features [16]. First-to-fifth principal components were selected and 93.0% accuracy was reached on the test set. However, a significantly unbalanced sensitivity *vs*. specificity pair was obtained (97.0% sensitivity and 79.3% specificity). Forward stepwise LR was also previously applied to a wide feature set from NPO, reaching 89.7% accuracy (92.0% sensitivity and 85.4% specificity) using cross-validation in a single population set [14]. Additionally, the study by Khandoker *et al*. [43] applied a similar scheme to assist in OSA diagnosis from ECG recordings: wavelet analysis, hill-climbing and support vector machines are applied for feature extraction, selection and classification, respectively. An accuracy of 92.85% was reached in the test set (100% in the

training set) using 4 features automatically selected from an initial set composed of 28 features. A population set composed of 125 subjects (83 for training and 42 for testing) were used to develop and further assess the methodology. In the present study, we used a new and larger population database to prospectively assess variable selection techniques. To our knowledge, this is the first study where an exhaustive search of a wide feature space from oximetry was made by means of GAs to help in OSA diagnosis.

We should take into account some limitations regarding the general application of our methodology. Desaturations in the overnight SpO_2 profile could not be exclusively due to apnoea events typical of OSA, which could influence our results. According to the AASM rules, our findings should only be applied to patients with prior symptoms of sleep apnoea and without additional significant respiratory or sleep-related breathing disorders. On the other hand, a control group composed of normal subjects without suspicion of sleep-related breathing disorders could provide significant information about the consistence of our optimum model. Moreover, further work is required to test the performance of our methodology from ambulatory portable monitoring at patient's home. Additionally, the initial feature set could be increased, in order to include new features with additional and complementary information. Specifically, it would be very interesting to assess the influence of conventional indexes (CTs, SITs, ODIs and Δ indexes) on the performance of a LR model from oximetry. Moreover, additional feature selection and dimensionality reduction techniques could be assessed in the context of OSA diagnosis from oximetry, such as variable ranking, hill-climbing, factor analysis, subspace clustering or simulated annealing.

6. Conclusions

In summary, we found that feature selection procedures could provide essential information from NPO in the context of OSA diagnosis. An optimal feature subset were

found by means of GAs from our proposed oximetric feature set composed of time statistics, frequency statistics, spectral and nonlinear features. *M1t*, *M3t*, *M4t*, *MF*, *PA* and *CTM* were automatically selected, which achieved 90.6% sensitivity, 81.3% specificity and 87.5% accuracy on an independent test set. Thus, GAs could be a useful tool to exhaustively investigate an oximetric feature space in order to find feature subsets with high diagnostic power. Our methodology outperforms the diagnostic ability of oximetric indexes commonly found in the state-of-the-art of OSA diagnosis. We conclude that complementary feature extraction plus exhaustive variable selection methodologies could obtain essential information from SpO_2 recordings to improve NPO usefulness in the diagnostic assessment of OSA syndrome.

Acknowledgements

This work has been partially supported by Ministerio de Ciencia e Innovación and FEDER grant TEC 2008-02241 and the grant project from the Consejería de Sanidad de la Junta de Castilla y León GRS 337/A/09. D. Álvarez was in receipt of a PIRTU grant from the Consejería de Educación de la Junta de Castilla y León and the European Social Fund (ESF).

References

[1] Young T, Skatrud J, Peppard PE. Risk Factors for obstructive sleep apnea in adults. JAMA 2004;291:2013–2016.

[2] Patil SP, Schneider H, Schwartz AR, Smith PL. Adult obstructive sleep apnea. Pathophysiology and diagnosis. Chest 2007;132:325–337.

[3] Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea. Am J Resp Crit Care 2002;165:1217–1239.

[4] Lopez-Jimenez F, Sert FH, Gami A, Somers VK. Obstructive sleep apnea. Implications for cardiac and vascular disease. Chest 2008;133:793–804.

[5] Flemons WW, Littner MR, Rowlet JA, Gay P, Anderson WM, Hudgel DW, McEvoy RD, Loube DI. Home diagnosis of sleep apnea: A systematic review of the literature. Chest 2003;124:1543–1579.

[6] Collop NA, Anderson WMcD, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, Hudhel D, Sateia M, Schwab R. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Journal of Clinical Sleep Med 2007;3:737–747.

[7] Penzel T, McNames J, de Chazal P, Raymond B, Murray A, Moody G. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. Med Biol Eng Comput 2002;40:402–407.

[8] De Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnea. IEEE Trans Biomed Eng 2003;50:686–696.

[9] Salisbury JI, Sun Y. Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique. Med Eng Phys 2007;29:336 – 343.

[10] Nakano H, Tanigawa T, Furukawa T, Nishima N. Automatic detection of sleep-disordered breathing from a single-channel airflow record. Eur Respir J 2007;29:728 –736.

[11] Caseiro P, Fonseca-Pinto R, Andrade A. Screening of obstructive sleep apnea using Hilbert-Huang decomposition of oronasal airway pressure recordings. Med Eng Phys 2010;32:561–568.

[12] Netzer N, Eliasson AH, Netzer C, Kristo DA. Overnight pulse oximetry for sleep-disordered breathing in adults. Chest 2001;120:625–633.

[13] Zamarrón C, Romero PV, Rodríguez JR, Gude F. Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. Clin Sci 1999;97:467–473.

[14] Álvarez D, Hornero R, Marcos JV, del Campo F. Multivariate Analysis of Blood Oxygen Saturation Recordings in Obstructive Sleep Apnea Diagnosis. IEEE Trans Biomed Eng 2010;57:2816–2824.

[15] Foo JYA. Pulse transit time in paediatric respiratory sleep studies. Med Eng Phys 2007;29:17–25.

[16] Marcos JV, Hornero R, Álvarez D, del Campo F, Aboy M. Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis. Medical & Biological Engineering & Computing 2010;48:895–902.

[17] Bishop CM. 'Neural networks for pattern recognition'. Oxford: Oxford University Press, 1995.

[18] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research 2003;3:1157–1182.

[19] Chaudhary B, Dasti S, Park Y, Brown T, Davis H, Akhtar B. Hour-to-hour variability of oxygen saturation in sleep apnea. Chest 1998;113:719–722.

[20] Vazquez J, Tsai WH, Flemons WW, Masuda A, Brant R, Hajduk E, Whitelaw WA, Remmers JE. Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnea. Thorax 2000;55:302–307.

[21] Magalang UJ, Dmochowski J, Veeramachaneni S, Draw A, Mador MJ, El-Solh A, Grant BJB. Prediction of the apnea-hypopnea index from overnight pulse oximetry. Chest 2003;124:1694–1701.

[22] Marcos JV, Hornero R, Álvarez D, Nabney IT, del Campo F, Zamarrón C. Classification of oximetry signals using Bayesian neural networks to assist in the detection of the obstructive sleep apnoea syndrome. Physiol Meas 2010;31:375–394.

[23] Chung-Ching H, Chung-Chieh Y. Smoothed periodogram of oxyhemoglobin saturation by pulse oximetry in sleep apnea syndrome. Chest 2007;131:750–757.

[24] Siedlecki W, Sklansky J. A note on genetic algorithms for large scale feature selection. Pattern Recognition Letters 1989;10:335–347.

[25] Iber C, Ancoli-Israel S, Chesson AL, Quan SF for the American Academy of Sleep Medicine. 'The AASM Manual for the scoring of sleep and associated events. Rules, terminology and technical specifications'. 1ST Ed, American Academy of Sleep Medicine. Westchester, Illinois 2007.

[26] Jobson JD. 'Applied multivariate data analysis. Volume I: Regression and experimental design'. Eds Springer-Verlag. New York 1991.

[27] Carter GC. Coherence and time delay estimation. Proceedings of the IEEE 1987;75:236–255.

[28] Poza J, Hornero R, Abásolo D, Fernández A, García M. Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease. Med Eng Phys 2007;29:1073–1083.

[29] Inouye T, Shinosaki K, Sakamoto H, Toi S, Ukai S, Iyama A, Katsuda Y, Hirano M. Quantification of EEG irregularity by use of the entropy of the power spectrum. Electroencephalogr Clin Neurophysiol 1991;79:204–210.

[30] Richman JS, Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278:H2039–H2049.

[31] Hornero R, Álvarez D, Abásolo D, del Campo F, Zamarrón C. Utility of approximate entropy from overnight pulse oximetry data in the diagnosis of obstructive sleep apnea syndrome. IEEE T Bio-Med Eng 2007;54:107–113.

[32] Cohen ME, Hudson DL, Deedwania PC. Applying continuous chaotic modeling to cardiac signals analysis. IEEE Eng Med Biol 1996;15:97–102.

[33] Álvarez D, Hornero R, García M, del Campo F, Zamarrón C. Improving diagnostic ability of blood oxygen saturation from overnight pulse oximetry in obstructive sleep apnea detection by means of central tendency measure. Artif Intell Med 2007;41:13–24.

[34] Lempel A, Ziv J. On the complexity of finite sequences. IEEE T Inform Theory 1976;IT–22:75–81.

[35] Álvarez D, Hornero R, Abásolo D, del Campo F, Zamarrón C. Nonlinear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection. Physiol Meas 2006;27:399–412.

[36] Tang KS, Man KF, Kwong S, He Q. Genetic algorithms and their applications. IEEE Signal Processing Magazine 1996;13:22–37.

[37] Yom-Tov E, Inbar GF. Feature selection for the classification of movements from single movement-related potentials. IEEE Trans Neural Systems and Rehab Eng 2002;10:170–177.

[38] Yang J, Honavar V. Feature subset selection using a genetic algorithm. IEEE Intelligent systems and their applications 1998;13:44–49.

[39] Chesson AL, Anderson WMcD, Walls RC, Bairnsfather RE. Comparison of two methods of quantitative assessment of hypopxemia in patients with sleep disorders. Sleep Med 2001;2:37–45.

[40] Lévy P, Pepin JL, Deschaux-Blanc C, Paramelle B, Brambilla C. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. Chest 1996;109:395–399.

[41] Olson LG, Ambrogetti A, Gyulay SG. Prediction of sleep-disordered breathing by unattended overnight oximetry. J Sleep Res 1999;8:51–55.

[42] Marcos JV, Hornero R, Álvarez D, del Campo F, and Zamarón C. Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry. Med Eng Phys 2009;31:971–978.

[43] Khandoker AH, Palaniswami M, Karmakar C. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings. IEEE Trans Biomed Eng 2009; 13:37–48.

Table 1

Summary of demographic and clinical features for patient groups.

Features	All subjects	OSA-negative	OSA-positive
Subjects (n)	240	80	160
Age (years)	52.3 ± 13.7	47.2 ± 12.2	54.8 ± 13.8
Males (%)	77.5	65.0	83.8
BMI (kg/m^2)	29.8 ± 4.4	27.8 ± 3.7	30.8 ± 4.3
Records (h)	7.3 ± 0.6	7.3 ± 0.3	7.2 ± 0.6
AHI (e/h)		3.9 ± 2.4	36.6 ± 25.7
Features	Training set	OSA-negative	OSA-positive
Subjects (n)	96	32	64
Age (years)	52.4 ± 13.8	47.3 ± 10.6	54.9 ± 14.5
Males (%)	77.1	62.5	84.4
BMI (kg/m^2)	29.8 ± 4.2	28.3 ± 4.4	30.6 ± 3.9
Records (h)	7.3 ± 0.3	7.3 ± 0.3	7.2 ±0.4
AHI (e/h)		4.2 ± 2.2	35.0 ± 25.2
Features	Test set	OSA-negative	OSA-positive
Subjects (n)	144	48	96
Age (years)	52.2 ± 13.7	47.2 ± 13.2	54.7 ± 13.4
Males (%)	77.8	66.7	83.3
BMI (kg/m^2)	29.8 ± 4.5	27.5 ± 3.3	31.0 ± 4.7
Records (h)	7.3 ± 0.7	7.3 ± 0.3	7.2 ± 0.8
AHI (e/h)		3.7 ± 2.5	37.7 ± 26.2

Data are presented as mean ± SD, number (n) or percentage (%). kg/m^2: kilogram per square meter; e/h: events per hour; BMI: body mass index; AHI: apnoea-hypopnoea index.

Table 2

Average value of each feature in the training set.

	OSA-negative	OSA-positive	*p*-value
M1t	94.39 ± 2.14	93.41 ± 2.33	$p > 0.001$
M2t	0.67 ± 0.19	1.57 ± 1.30	$p << 0.001$
M3t	0.13 ± 0.53	-0.12 ± 0.32	$p > 0.001$
M4t	4.66 ± 1.09	3.75 ± 0.72	$p < 0.001$
M1f	2.51 ± 1.54	14.32 ± 33.40	$p < 0.001$
M2f	660.61 ± 852.86	31076.31 ± 184328.55	$p < 0.001$
M3f	13.57 ± 0.97	10.46 ± 3.10	$p < 0.001$
M4f	206.16 ± 22.75	140.34 ± 65.64	$p < 0.001$
MF	0.002 ± 0.001	0.009 ± 0.007	$p < 0.001$
SE	0.45 ± 0.06	0.54 ± 0.07	$p < 0.001$
P_T	1.13 ± 0.69	6.77 ± 15.91	$p < 0.001$
PA	15.55 ± 8.26	269.05 ± 737.41	$p << 0.001$
P_R	0.16 ± 0.06	0.32 ± 0.14	$p < 0.001$
SampEn	0.31 ± 0.05	0.34 ± 0.06	$p > 0.001$
CTM	0.997 ± 0.003	0.950 ± 0.095	$p << 0.001$
LZC	0.24 ± 0.03	0.26 ± 0.04	$p > 0.001$

Data are presented as mean ± SD.

Table 3

Average value of each conventional oximetric index in the training set.

	OSA-negative	OSA-positive	*p*-value
LO2 (%SpO_2)	87.90 ± 3.95	80.25 ± 9.37	$p < 0.001$
CT90 (%t)	7.21 ± 22.30	10.09 ± 17.35	$p < 0.001$
SIT90 (%SpO_2·min/h)	0.14 ± 0.46	0.43 ± 1.22	$p < 0.001$
ODI3 (n/h)	1.24 ± 1.14	17.62 ± 21.88	$p << 0.001$
Δ (%SpO_2/s)	0.04 ± 0.01	0.13 ± 0.14	$p << 0.001$

Data are presented as mean ± SD. Units: percentage of saturation (%SpO_2); percentage of time (%t); percentage of saturation per minute per hour of sleep (%SpO_2·min/h); number of events per hour of sleep; percentage of saturation per second (%SpO_2/s).

Table 4

Diagnostic assessment of each single feature in the training set.

	TP	TN	FP	FN	Se	Sp	Ac
M1t	42	22	10	22	65.6	68.8	66.7
M2t	48	28	4	16	75.0	87.5	79.2
M3t	43	20	12	21	67.2	62.5	65.6
M4t	46	25	7	18	71.9	78.1	74.0
M1f	48	25	7	16	75.0	78.1	76.0
M2f	44	21	11	20	68.8	65.6	67.7
M3f	44	28	4	20	68.8	87.5	75.0
M4f	44	27	5	20	68.8	84.4	74.0
MF	46	29	3	18	71.9	90.6	78.1
SE	50	26	6	14	78.1	81.3	79.2
P_T	47	26	6	17	73.4	81.3	76.0
PA	51	30	2	13	79.7	93.8	84.4
P_R	49	28	4	15	76.6	87.5	80.2
SampEn	45	17	15	19	70.3	53.1	64.6
CTM	51	29	3	13	79.7	90.6	83.3
LZC	36	22	10	28	56.3	68.8	60.4

TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; Se: sensitivity (%); Sp: specificity (%); Ac: accuracy (%).

Table 5

Diagnostic assessment of each conventional oximetric index in the training set.

	TP	TN	FP	FN	Se	Sp	Ac
LO2	45	25	7	19	70.3	78.1	72.9
CT90	43	26	6	21	67.2	81.3	71.9
SIT90	46	26	6	18	71.9	81.3	75.0
ODI3	57	28	4	7	89.1	87.5	88.5
Δ	56	29	3	8	87.5	90.1	88.5

TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; Se: sensitivity (%); Sp: specificity (%); Ac: accuracy (%).

Table 6

Diagnostic assessment of LR models from GAs and the LR full-model in the training set.

n	Features in the LR model	TP	TN	FP	FN	Se	Sp	Ac
2	*PR - CTM*	58	30	2	6	90.6	93.8	91.7
3	*M4t - SE - PT*	62	28	4	2	96.9	87.5	93.8
4	*MF - PT - PR - CTM*	60	30	2	4	93.8	93.8	93.8
5	*M4t - M4f - MF - SE - PT*	63	28	4	1	98.4	87.5	94.8
6	*M1t - M3t - M4t - MF - PA- CTM*	63	29	3	1	98.4	90.6	95.8
7	*M3t - M4t - M1f - M2f - M4f - MF - CTM*	63	29	3	1	98.4	90.6	95.8
8	*M3t - M4t - M2f - M4f - SE - PT - PA - CTM*	63	29	3	1	98.4	90.6	95.8
9	*M3t - M4t - M2f - M3f - SE - PT - PA - SampEn - CTM*	63	29	3	1	98.4	90.6	95.8
10	*M3t - M4t - M1f - MF - SE - PT - PA - PR - SampEn - CTM*	63	29	3	1	98.4	90.6	95.8
11	*M3t - M4t - M1f - M4f - MF - SE - PT - PR - SampEn - CTM - LZC*	63	29	3	1	98.4	90.6	95.8
12	*M1t - M3t - M4t - M1f - M3f - MF - SE - PT - PA - PR - SampEn - LZC*	62	30	2	2	96.9	93.8	95.8
13	*M1t - M2t - M3t - M4t - M2f - M3f - MF - SE - PT - PR - SampEn - CTM - LZC*	63	29	3	1	98.4	90.6	95.8
14	*M1t - M2t - M3t - M4t - M2f - M3f - MF - SE - PT - PA - PR - SampEn - CTM - LZC*	63	29	3	1	98.4	90.6	95.8
15	*M1t - M2t - M3t - M4t - M1f - M3f - M4f - MF - SE - PT - PA - PR - SampEn - CTM - LZC*	59	31	1	5	92.2	96.9	93.8
16	All features	57	32	0	7	89.1	100.0	92.7

n: number of features; TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; Se: sensitivity (%); Sp: specificity (%); Ac: accuracy (%).

Table 7

Diagnostic assessment of LR models from GAs applied to conventional indexes and the LR full-model in the training set.

n	Features in the LR models from GAs	TP	TN	FP	FN	Se	Sp	Ac
2	*ODI3*, *Δ*	55	31	1	9	85.9	96.9	89.6
3	*CT90*, *ODI3*, *Δ*	55	32	0	9	85.9	100.0	90.6
4	*CT90*, *SIT90*, *ODI3*, *Δ*	56	32	0	8	87.5	100.0	91.7
5	All features	57	30	2	7	89.1	93.8	90.6

n: number of features; TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; Se: sensitivity (%); Sp: specificity (%); Ac: accuracy (%).

Table 8

Diagnostic assessment of optimum LR models from GAs in the test set.

n	Features in the LR model	TP	TN	FP	FN	Se	Sp	Ac
6	*M1t - M3t - M4t - MF - PA- CTM*	87	39	9	9	90.6	81.3	87.5
7	*M3t - M4t - M1f - M2f - M4f - MF - CTM*	87	38	10	9	90.6	79.2	86.8
8	*M3t - M4t - M2f - M4f - SE - P_T - PA - CTM*	87	37	11	9	90.6	77.1	86.1
9	*M3t - M4t - M2f - M3f - SE - P_T - PA - SampEn - CTM*	88	37	11	8	91.7	77.1	86.8
10	*M3t - M4t - M1f - MF - SE - P_T - PA - P_R - SampEn - CTM*	87	38	10	9	90.6	79.2	86.8
11	*M3t - M4t - M1f - M4f - MF - SE - P_T - P_R - SampEn - CTM - LZC*	86	36	12	10	89.6	75.0	84.7
12	*M1t - M3t - M4t - M1f - M3f - MF - SE - P_T - PA - P_R - SampEn - LZC*	84	40	8	12	87.5	83.3	86.1
13	*M1t - M2t - M3t - M4t - M2f - M3f - MF - SE - P_T - P_R - SampEn - CTM - LZC*	88	37	11	8	91.7	77.1	86.8
14	*M1t - M2t - M3t - M4t - M2f - M3f - MF - SE - P_T - PA - P_R - SampEn - CTM - LZC*	88	37	11	8	91.7	77.1	86.8

n: number of features; TP: true positives; TN: true negatives; FP: false positives; FN: false negatives; Se: sensitivity (%); Sp: specificity (%); Ac: accuracy (%).

Figures

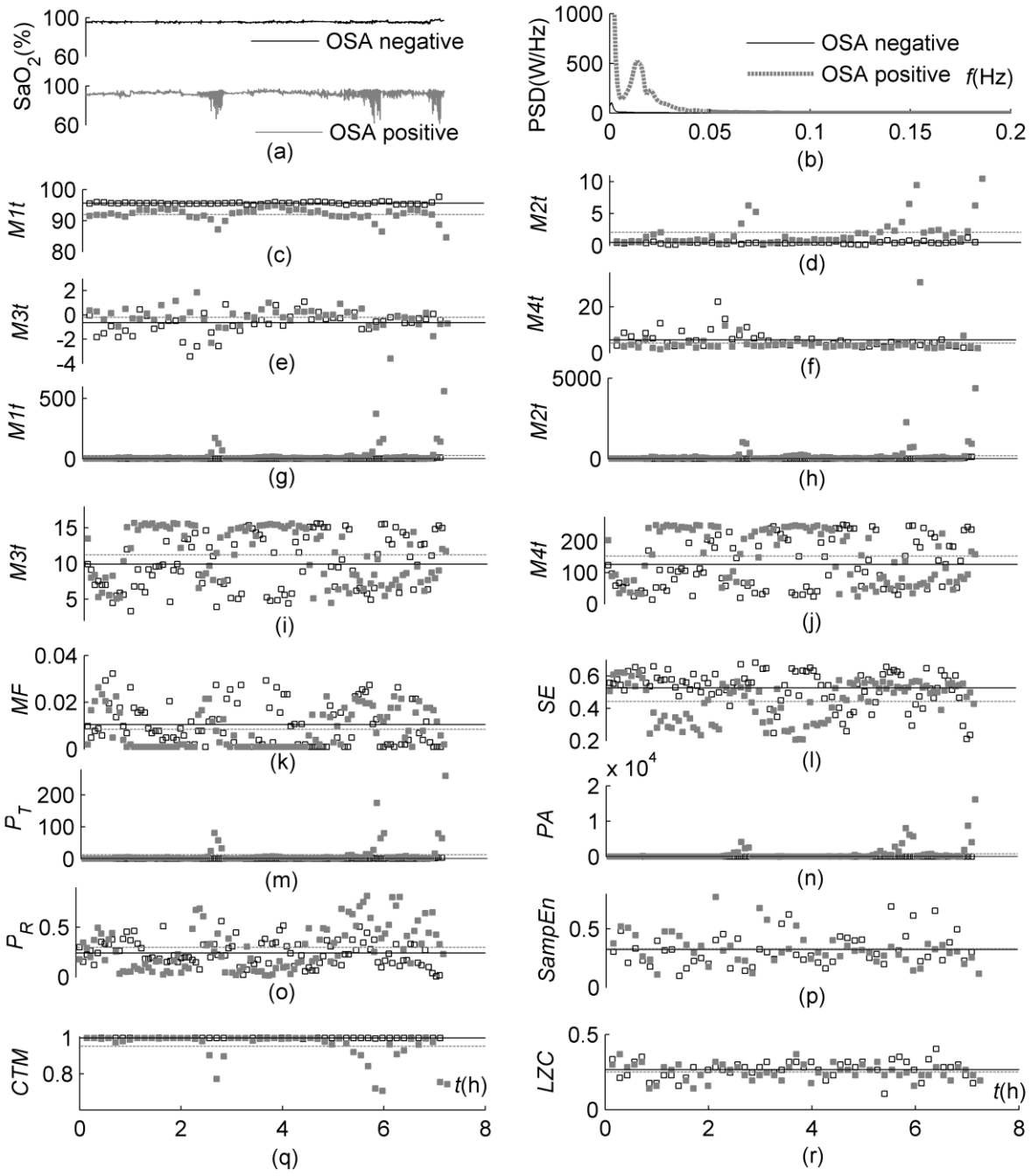

Fig. 1. Recordings from a common OSA-negative subject and a representative OSA-positive patient. SpO_2 profiles (a) in the time domain and (b) in the frequency domain. Evolution during the recording time of (c) mean in the time domain, (d) variance in the time domain, (e) skewness in the time domain and (f) kurtosis in the time domain, (g) mean in the frequency domain, (h) variance in the frequency domain, (i) skewness in the frequency domain and (j) kurtosis in the frequency domain, (k) median frequency, (l) spectral entropy, (m) signal power, (n) peak amplitude and (o) relative power in the OSA frequency range, (p)

sample entropy, (q) central tendency measure and (r) Lempel and Ziv complexity. The black solid line and the grey dashed line represent the average feature value for the OSA-negative and the OSA-positive subjects, respectively. Black squares and grey solid squares represent feature values for the OSA-negative and the OSA-positive subjects, respectively.

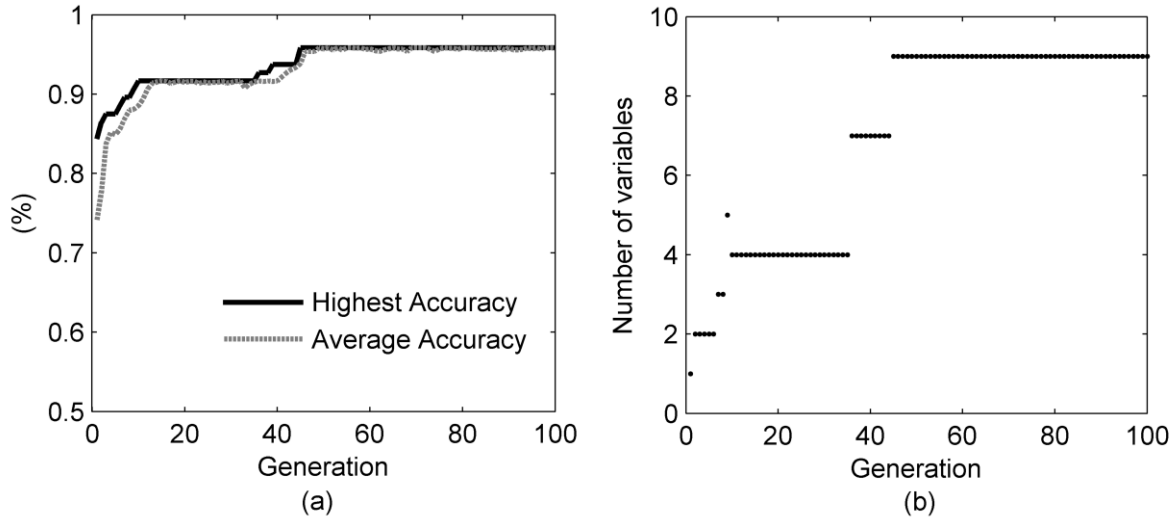

Fig. 2. An example of feature selection by means of GAs in the training set. (a) Optimal and average accuracies from the population and (b) number of selected variables at each generation throughout a realisation of the GAs procedure.