Cite this article as:

Marcos, J.V., Hornero, R., Alvarez, D., del Campo, F. and Zamarron, C., 2009. Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry. *Medical Engineering & Physics*, *31*(8), pp.971-978.

DOI: https://doi.org/10.1016/j.medengphy.2009.05.010

Assessment of four statistical pattern recognition techniques to assist in obstructive sleep apnoea diagnosis from nocturnal oximetry

J. Víctor Marcos[1], Roberto Hornero[1], Daniel Álvarez[1], Félix del Campo[2], Carlos Zamarrón[3]

[1]Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, Universidad de Valladolid, Spain

[2]Servicio de Neumología, Hospital del Río Hortega, Valladolid, Spain

[3]Servicio de Neumología, Hospital Clínico Universitario, Santiago de Compostela, Spain

AUTHOR'S ADDRESS: J. Víctor Marcos

E.T.S. Ingenieros de Telecomunicación

Universidad de Valladolid

Camino del Cementerio s/n

47011 - Valladolid (Spain)

Phone: +34 983 423000, ext. 5589

Fax: +34 983 423667

E-mail: jvmarcos@gmail.es

Abstract

The aim of this study is to assess the capability of traditional statistical pattern recognition techniques to help in obstructive sleep apnoea (OSA) diagnosis. Classifiers based on quadratic (QDA) and linear (LDA) discriminant analysis, *K*-nearest neighbours (KNN) and logistic regression (LR) were evaluated. Spectral and nonlinear features from oxygen saturation (SaO_2) signals were used as inputs. A total of 187 recordings from patients suspected of suffering from OSA were available. This initial dataset was divided into training and test sets with 74 and 113 signals, respectively. Several classification algorithms were developed by applying QDA, LDA, KNN and LR with spectral features, nonlinear features and combination of both groups. The performance of each algorithm was measured on the test set by means of classification accuracy and receiver operating characteristic (ROC) analysis. QDA, LDA and LR showed better classification capability than KNN. The classifier based on LDA with spectral features provided the best diagnostic ability with an accuracy of 87.61% (91.05% sensitivity and 82.61% specificity) and an area under the ROC curve (AROC) of 0.925. Statistical pattern recognition techniques evaluated in our study could be applied as an OSA screening tool and could contribute to reduce the number of polisomnographies.

Keywords: Obstructive sleep apnoea, oxygen saturation, statistical pattern recognition, discriminant analysis, *K*-nearest neighbours, logistic regression

1. Introduction

Obstructive sleep apnoea (OSA) is characterised by repetitive pharyngeal collapse during sleep, causing intermittent cessations of breathing (apnoea) or marked reduction (hypopnoea) in airflow [1, 2]. Apnoea episodes are accompanied by hypoxaemia and usually terminate with microarousals, resulting in sleep fragmentation. These events are associated with development of severe cardiovascular and cerebrovascular diseases [1]. Epidemiological studies estimate the prevalence of OSA up to 5% of adult men in western countries [3]. Moreover, it has been reported that 83% of men and 93% of women suffering from OSA remain undiagnosed [4]. As a result, it can be considered a public health concern. Nowadays, nocturnal polysomnography (PSG) is the gold standard in OSA diagnosis. However, it presents some drawbacks. A high number of physiological signals and data are acquired in each polysomnographic test [2]. These must be separately analysed by a medical expert to obtain a final diagnosis. Furthermore, a sleep laboratory equipped with qualified technical personnel is required to perform PSG [5].

Nocturnal pulseoximetry represents an alternative to PSG in OSA diagnosis. Pulseoximetry can be performed at home and enables cost reduction. Arterial oxygen saturation (SaO_2) is monitored during sleep in a noninvasive manner. This recording provides useful information about respiratory dynamics [6]. Hypoxaemia events due to apnoeas are reflected on this signal by means of drops and posterior restorations of the saturation value. As a result, oximetry recordings from patients affected by OSA are expected to be more unstable than those from control subjects. This behaviour can be used to detect OSA [7]. Different approaches have been proposed to evaluate OSA diagnosis from SaO_2 recordings. Visual inspection of oximetry data is a straightforward technique [8]. Alternatively, computation of traditional oximetry indices from SaO_2 signals represents a basic approach to automated OSA diagnosis. These indices are the oxygen desaturation index over 2% (ODI2),

3% (ODI3) and 4% (ODI4), and the cumulative time spent below a given level of saturation, usually 90% (CT90) [7]. Their diagnostic utility has been assessed in previous studies [9-13]. Finally, advanced signal processing techniques involve a step forward in automated processing of SaO_2 signals. Both spectral and nonlinear analyses of oximetry recordings have provided promising results in OSA diagnosis [14-18].

We propose to apply traditional statistical pattern recognition techniques to assist in OSA diagnosis from oximetry data. Diagnosis of OSA can be modelled as a pattern recognition problem [19]. A patient must be classified into one of two possible groups: OSA positive or OSA negative. We used spectral and nonlinear features from SaO_2 signals to classify patients suspected of suffering from OSA. These features were applied as inputs to four statistical pattern recognition techniques: quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), *K*-nearest neighbours (KNN) and logistic regression (LR) [20-24]. These algorithms aim to determine the optimum decision boundaries in the feature space from a set of training samples [20]. They have been widely applied to classify diverse biomedical data [25-27]. Specifically, QDA, LDA, KNN and LR were used in other studies related to OSA [28-34]. Home videotape recordings [28], morphological features of the upper airway [29] and electrocardiogram (ECG) signals were analysed [30-34]. To our knowledge, this is the first research where these techniques have been used to assist in OSA diagnosis from SaO_2 signals. Other classification methods such as neural networks were applied together with oximetry recordings to perform automated OSA diagnosis. In preceding studies of our research group, the diagnostic utility of multilayer perceptron (MLP) and radial basis function (RBF) neural networks was assessed [35, 36]. Neural networks are powerful tools for classification and most of the well-known network models can be seen as a subset of statistical pattern recognition methods [20]. Nevertheless, they require an important amount of model parameters to be specified by the user, resulting in complex design and training

processes. These issues can be simplified by using traditional statistical pattern recognition techniques.

In this study, several classification algorithms based on QDA, LDA, KNN and LR were developed by using spectral and nonlinear features from SaO_2 signals. We assessed a total of 12 classifiers. The four proposed techniques were evaluated with three different inputs: spectral features, nonlinear features and combination of both. The aim of this study is to assess the ability of traditional statistical pattern recognition techniques in OSA diagnosis by only using oximetry data.

2. Subjects and signals

Our database was composed by 187 subjects suspected of suffering from OSA. Sleep studies were carried out usually from midnight to 8:00 AM in the Sleep Unit of the Hospital Clínico Universitario de Santiago de Compostela (Spain). The Review Board on Human Studies at this institution approved the protocol. Pulseoximetry and polysomnography (PSG) tests were simultaneously performed on each subject. Oximetry signals were recorded by means of a Criticare 504 oximeter (CSI, Waukeska, U.S.A.) at a sampling frequency of 0.2 Hz. The equipment used to perform PSG was a polygraph (Ultrasom Network, Nicolet, Madison, W.I., U.S.A.). Recordings obtained through PSG were electroencephalogram (EEG), electro-oculogram (EOG), chin electromiogram (EMG), airflow (three-port thermistor), ECG and measurement of chest wall movement. Recordings from PSG were analysed according to the system by Rechtschaffen and Kales [37] to obtain a medical diagnosis for each subject. Apnoea was defined as a cessation of airflow for 10 seconds or longer. Hypopnoea was defined as a reduction, without complete cessation, in airflow of at least 50%, accompanied by a decrease of more than 4% in the saturation of haemoglobin. The average apnoea-hypopnoea index (AHI) was computed for hourly periods of sleep from episodes of apnoea/hypopnoea captured in PSG. A threshold of AHI $\geq$ 10 events/h was

applied to diagnose OSA.

A total of 111 (59.4%) subjects were diagnosed as OSA positive. There were no significant differences between OSA positive and negative groups in age, BMI and recording time. The percentage of males was higher in the OSA positive group (84.7%) than in the OSA negative (69.7%). The initial population was randomly divided into training and test sets. The former was composed by 74 subjects. This set was used to optimise adjustable parameters in each classification method. The latter was composed by 113 subjects and was used to assess the performance of the proposed algorithms. Clinical and demographic statistics for the whole population as well as for training and test sets are summarised in Table 1.

INSERT TABLE 1 AROUND HERE

---

3. Methods

Traditional statistical pattern recognition techniques were applied to help in OSA diagnosis. We developed classification algorithms based on QDA, LDA, KNN and LR to process information from SaO_2 signals. They provide an effective and ease-of-use means to classify multivariate patterns. Indeed, these techniques have been commonly used to solve pattern recognition problems involving biomedical data [25-27].

Episodes of apnoea are usually associated with decrease and subsequent restoration of the oxygen saturation value. As a result, oximetry signals from subjects affected by OSA tend to present frequent fluctuations due to the repetition of apnoea events during sleep [7]. In contrast, SaO_2 signals from control subjects tend to present a constant value round 97% of saturation [9]. The unstable behaviour characteristic of oximetry recordings from OSA positive subjects can be used to detect OSA. We evaluated several algorithms to help in OSA diagnosis by means of automated analysis of SaO_2 recordings. Our algorithms were composed

of two different stages. In the first one, spectral and nonlinear analyses of SaO_2 were applied to extract features related to OSA. In the second stage, these features were used as inputs to a classifier based on statistical pattern recognition techniques.

3.1. Feature extraction

In the first stage, we used signal processing techniques to perform feature extraction from SaO_2 recordings. According to our previous research work, spectral and nonlinear analyses of oximetry signals provide useful information to diagnose OSA. Statistically significant differences were obtained between OSA positive and negative subjects by evaluating different spectral and nonlinear features [14-18]. Therefore, we propose to simultaneously apply them to classify subjects suspected of suffering from OSA.

3.1.1. Spectral analysis

Spectral analysis of SaO_2 recordings was focused in the band between 0.010 and 0.033 Hz. Our preceding studies showed that signal power associated with these frequency components is higher in subjects with OSA than in normal controls [14]. Periodicities of ventilation originate phase-lagged changes in SaO_2. The duration of apnoea events (at least 10 s) and their repetition during sleep result in fluctuations of the oximetry signal that correspond to those frequencies [14]. The nonparametric Welch's method was used to estimate the power spectral density (PSD) of SaO_2 signals [38]. The following spectral features were computed:

- Total area under the PSD (S_T). This feature provides an estimate of the signal power.
- Area enclosed in the band of interest (S_B). This measure approximates the amount of signal power contained in the band between 0.010 and 0.033 Hz.
- Peak amplitude of the PSD in this band (PA). It is the local maxima of the PSD in the band under study.

As indicated in [14], signals from positive patients tend to present a peak between 0.010 and 0.033 Hz of the PSD. Therefore, oximetry signals corresponding to these subjects

are expected to provide higher values for the selected spectral features than those obtained from control subjects. This information can be used to classify suspected OSA patients.

3.1.2. Nonlinear analysis

The occurrence of apnoea events during sleep results in unstable SaO_2 signals characterised by frequent variations. Usually, this behaviour is presented by recordings from subjects affected by OSA. According to our previous researches, nonlinear analysis of oximetry data can reflect this instability, representing a useful means to quantitatively distinguish OSA patients from normal subjects [15-18]. Thus, the following nonlinear methods were applied to SaO_2 recordings:

- Approximate entropy (ApEn) provides an estimate of the regularity of a signal [39]. High values of ApEn correspond to irregular signals. In our analysis, high irregularity is expected for oximetry recordings from OSA positive subjects [16, 18].
- Central tendency measure (CTM) quantifies the variability of a time series [40], assigning low values to signals with a high degree of chaos. Usually, this is the case of SaO_2 recordings from OSA positive subjects [15].
- Lempel-Ziv complexity (LZC) evaluates the randomness of a time series [41]. Complex signals generate high values of LZC. This is expected for recordings from subjects suffering from OSA [15].

The repetition of apnoea events during sleep is closely related to regularity, variability and complexity of SaO_2 data measured by means of ApEn, CTM and LZC [15-18]. Thus, we suggest to apply these features as inputs to classification algorithms developed in this study.

3.2. Classification

In the second stage, spectral and nonlinear features from oximetry were fed into a statistical classifier. In statistical pattern recognition, a feature or pattern vector is denoted by $\mathbf{x} = [x_1, \ldots, x_d] \in \Re^d$ and is associated to one of c categories $\{\omega_1, \ldots, \omega_c\}$. A pattern vector $\mathbf{x}$

belonging to class ω_j (j = 1, …, c) is viewed as an observation drawn randomly from the class-conditional probability function $p(\mathbf{x}|\omega_j)$ [20]. A classifier can be regarded as a function $f:\Re^d \to \{\omega_1,\ldots,\omega_c\}$, i.e., it assigns a vector $\mathbf{x}$ to one of c possible classes. The Bayes decision rule allows to minimise the probability of error in the classification task [22, 23]. It can be stated as follows:

$$\text{Decide } \omega_j \text{ for } \mathbf{x} \text{ if } p(\mathbf{x}|\omega_j)P(\omega_j) = \max_{j=1,\ldots,c} p(\mathbf{x}|\omega_j)P(\omega_j) \tag{1}$$

where $P(\omega_j)$ denotes the prior probability of class ω_j. According to the Bayes theorem [22], the following equivalency can be obtained:

$$p(\mathbf{x}|\omega_j)P(\omega_j) = p(\omega_j|\mathbf{x})p(\mathbf{x}) \tag{2}$$

where $p(\omega_j|\mathbf{x})$ is the a posteriori probability of class ω_j given $\mathbf{x}$ and $p(\mathbf{x})$ is the probability density function of the pooled data. Thus, decision rule in Eq. (1) can be equivalently expressed as:

$$\text{Decide } \omega_j \text{ for } \mathbf{x} \text{ if } p(\omega_j|\mathbf{x}) = \max_{j=1,\ldots,c} p(\omega_j|\mathbf{x}) \tag{3}$$

where $p(\mathbf{x})$ has been removed since it is constant for all classes. Given a set of training data composed of pairs ($\mathbf{x}_1$, ω_{j1}), …, ($\mathbf{x}_n$, ω_{jn}), statistical classifiers aim to establish decision boundaries in the feature space which separate patterns belonging to different classes. Two distinct classification approaches emerge: class-conditional density estimation and direct boundary construction [23]. The former corresponds to classification rule expressed in Eq. (1) and requires the estimation of $p(\mathbf{x}|\omega_j)$ and $P(\omega_j)$. Classifiers based on QDA, LDA and KNN operate in this manner. The latter is associated to expression in Eq. (3) and is represented by LR. Classifiers based on this approach directly estimate posterior probabilities $p(\omega_j|\mathbf{x})$.

Statistical classification techniques can be applied to medical decision making. We modelled OSA diagnosis as a pattern recognition task [19]. An input pattern composed of features from oximetry data must be assigned to one of two classes: OSA positive or OSA

negative. In this study, each classification technique was evaluated with three different input patterns:

- Input pattern of spectral features (Spec). Three-dimensional vector composed by spectral features S_T, S_B and PA.
- Input pattern of nonlinear features (NonLin). Three-dimensional vector defined by nonlinear features ApEn, CTM and LZC.
- Input pattern of spectral and nonlinear features (Spec-NonLin). Six-dimensional vector resulted from the combination of spectral and nonlinear features.

Thus, three different classification schemes were considered for each of the proposed statistical pattern recognition techniques. These schemes are represented in Figure 1.

INSERT FIGURE 1 AROUND HERE

3.2.1. Discriminant analysis (DA)

Classification rules based on discriminant analysis model each class-conditional density function $p(\mathbf{x}|\omega_j)$ as a multivariate normal distribution [22]:

$$p(\mathbf{x}|\omega_j) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_j|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right] \quad (4)$$

where $\boldsymbol{\mu}_j$ and $\mathbf{\Sigma}_j$ are the class ω_j mean vector and covariance matrix, respectively. Substituting this expression into Eq. (1) and taking the natural logarithm leads to the classification rule:

$$\text{Decide } \omega_j \text{ for } \mathbf{x} \text{ if } y_j(\mathbf{x}) = \max_{j=1,\ldots,c} y_j(\mathbf{x}) \quad (5)$$

where $y_j(\mathbf{x})$ is called the discriminant score for class ω_j. It is given by:

$$y_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T \mathbf{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j) - \frac{1}{2}\ln|\mathbf{\Sigma}_j| + \ln p(\omega_j) \quad (6)$$

The classification rule derived from these expressions is called QDA. It separates regions of

the feature space by establishing quadratic decision boundaries [21].

An important simplification occurs when all the class covariance matrices are presumed to be identical (homocedasticity):

$$\Sigma_j = \Sigma, \quad j = 1,...,c \tag{7}$$

In this case, Eq. (6) is expressed by:

$$y_j(\mathbf{x}) = \boldsymbol{\mu}_j^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_j^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \ln p(\omega_j) \tag{8}$$

This is referred to as LDA since it defines linear decision boundaries in the feature space [21].

Training a classifier based on QDA or LDA requires to adjust parameters $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ associated with the class-conditional densities. These parameters can be estimated from training data according to [21]:

$$\boldsymbol{\mu}_j = \frac{1}{N_j} \sum_{n=1}^{N_j} \mathbf{x}_n^j \tag{9}$$

and

$$\Sigma_j = \frac{1}{N_j} \sum_{n=1}^{N_j} (\mathbf{x}_n^j - \boldsymbol{\mu}_j)(\mathbf{x}_n^j - \boldsymbol{\mu}_j)^T \tag{10}$$

where $\mathbf{x}_n^j$ is the nth vector in the training set corresponding to class ω_j and N_j is the total number of training samples belonging to class ω_j. These represent the maximum likelihood estimates of both parameters [22]. In addition, prior probabilities $P(\omega_j)$ can be obtained as:

$$P(\omega_j) = \frac{N_j}{N} \tag{11}$$

where N is the number of samples in the training set. Moreover, the statistical distribution of the input data must satisfy some requirements in order to be modelled by QDA and LDA. Both classifiers require that all input variables have a normal distribution. Furthermore, LDA assumes homocedasticity for each input feature [21].

3.2.2. *K*-nearest neighbours (KNN)

The KNN approach estimates the density function $p(\mathbf{x})$ from the training set as [22]:

$$p(\mathbf{x}) = \frac{K/N}{V} \tag{12}$$

i.e., there are K training samples (from a total of N) contained in a spherical region of volume V centred on the point $\mathbf{x}$. Similarly, the class-conditional density $p(\mathbf{x}|\omega_j)$ is approximated by:

$$p(\mathbf{x}|\omega_j) = \frac{K_j/N_j}{V} \tag{13}$$

where K_j represents the total number of samples belonging to class ω_j that are found in a volume V centred on $\mathbf{x}$, given that N_j is the total number of training samples of class ω_j. Substituting these expressions into Eq. (2), the posterior probability $p(\omega_j|\mathbf{x})$ is computed as:

$$p(\omega_j|x) = \frac{K_j}{K} \tag{14}$$

where the approximation for prior probabilities given in Eq. (11) has been applied.

KNN algorithms classify data according to the rule expressed in Eq. (3) [22]. A metric function to compute distances and the number of neighbours must be specified by the user. For the first issue, the Euclidean distance is usually applied to identify neighbour points [20]. It is recommended to normalise input features to have zero mean and unit variance since computation of distance could be insensitive to those features with a small range of values [22]. In the case of parameter K (the size of the neighbourhood), it could be either specified by the user or adjusted by means of model selection procedures. Leave-one-out cross-validation from training data can be used to optimise the value of K [42].

3.2.3. Logistic regression (LR)

LR relates a categorical response variable (ω) with a set of input features. For dichotomous problems, input patterns are classified into one of two mutually exclusive categories, i.e., $c = 2$ and variable ω can be either ω_1 or ω_2. This case corresponds to OSA

diagnosis since subjects under study belong to one of two groups: OSA positive or negative. Therefore, the probability density for ω can be modelled by a Bernoulli distribution [24]. The probabilities associated with ω_1 and ω_2 are α and 1- α, respectively. In a classification context, the value taken by ω is conditioned to the observation of the feature vector **x**. Thus, α represents the posterior probability for class ω_1 given the input vector **x**:

$$\alpha = p(\omega_1|\mathbf{x}) \Rightarrow 1-\alpha = p(\omega_2|\mathbf{x}) \tag{15}$$

Thus, LR classifiers assign an input vector to the class with the maximum a posteriori probability value. For two-class problems, only the value of α is required to classify previously unseen data. In a LR model it is assumed that this value depends on the linear combination of several inputs. The dependence takes the following form [24]:

$$\alpha = \alpha(\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_d x_d)}} \tag{16}$$

which is called logistic function [20]. The maximum likelihood criteria is used to optimise the components of vector $\boldsymbol{\beta}$ = [β_0, …, β_d] from training data. The following expression is maximised by applying the weighted least squares (WLS) algorithm [24, 43]:

$$\prod_{\mathbf{x}_i^{(1)} \in \omega_1} \alpha(\mathbf{x}_i^{(1)}, \boldsymbol{\beta}) \prod_{\mathbf{x}_i^{(2)} \in \omega_2} \left[1-\alpha(\mathbf{x}_i^{(2)}, \boldsymbol{\beta})\right] \tag{17}$$

4. Results

We used the Matlab software to implement our algorithms. Firstly, spectral and nonlinear features were computed from signals in our database. The following specifications were applied:

- Spectral analysis. The Welch's method was used to analyse our recordings in the frequency domain. A Hanning window of 300 samples (overlapping of 50%) was applied. Fast Fourier Transforms were computed for a length of 512 samples. Finally, S_T, S_B and PA were obtained from the estimated PSD.

- Nonlinear analysis. We divided our signals into epochs of 200 samples to estimate ApEn, CTM and LZC. The measures obtained from every epoch were averaged to obtain the final value of these features [15, 18]. To perform SaO_2 analysis with ApEn, we set the run length parameter (m) and the size of the tolerance window (r) to 1 and 0.25 times the standard deviation of the original sequence, respectively [18]. In the case of CTM, a radius (ρ) equal to 0.25 was selected [15]. Finally, estimates of LZC were obtained by converting SaO_2 signals into 0–1 sequences. Each SaO_2 value was compared with the median of the samples in the corresponding epoch to transform the data [15].

Table 2 summarises the mean values of these features for the whole dataset as well as for the training and the test sets.

INSERT TABLE 2 AROUND HERE

4.1. Training

A total of 12 different classification algorithms were developed. As described before, QDA, LDA, KNN and LR were used to classify three different types of input patterns. The training set with 74 subjects was used to adjust the free parameters of each classifier:

- DA classifiers. Algorithms based on QDA and LDA were optimised by means of maximum likelihood estimates of class mean vector, covariance matrix and prior probabilities. Both the Kolmogorov-Smirnov and the Shapiro-Wilk tests were used to assess normality of distribution for each input feature, whereas homocedasticity was verified with Levene's test. Nonlinear variables met parametric test assumptions. Spectral features were transformed by means of the base 10 logarithm to fulfil the requirements of normality and homocedasticity.

- KNN classifiers. Parameter *K* of KNN algorithms was determined by leave-one-out cross-validation from data in the training set [42]. The value of *K* that provided the highest accuracy was selected.
- LR classifiers. Maximum likelihood estimates were obtained for LR classifiers by applying the WLS algorithm on data in the training set.

4.2. Testing

The classification ability of our algorithms was assessed on the test set with 113 subjects. Sensitivity, specificity and accuracy were computed. Moreover, we applied receiver operating characteristic (ROC) analysis [44]. ROC curve analysis suppresses the requirement for a threshold value by evaluating the performance of the classifier over the whole range of possible values. A plot of sensitivity versus 1 - specificity is made to obtain the ROC curve. The area under ROC curve (AROC) was used as a measure of classification performance [45, 46]. It represents the probability of correct classification for a randomly chosen pair of OSA positive and OSA negative subjects [44]. Table 3 summarises the results achieved by our classifiers.

INSERT TABLE 3 AROUND HERE

The implemented algorithms showed a high ability to discriminate SaO_2 recordings from OSA positive patients. The highest accuracy (87.61%) was achieved by the LDA classifier with spectral features from oximetry data. However, other configurations such as KNN with nonlinear features or LR with spectral features have provided similar accuracy results. An accuracy of 86.73% was obtained for both algorithms. The best value of AROC was reached by means of LR with the combination of spectral and nonlinear features. This classifier achieved an AROC of 0.930. Other algorithms based on LDA provided comparable

results for this statistic. An AROC of 0.925 was reached by the LDA classifier with spectral features. An identical AROC value was obtained by applying both spectral and nonlinear features as inputs to the LDA.

5. Discussion

We presented several classification algorithms based on traditional statistical pattern recognition techniques to help in OSA diagnosis. QDA, LDA, KNN and LR were used to process spectral and nonlinear features extracted from SaO_2 recordings. The proposed classifiers have shown to be a useful tool in screening for OSA. All our classification algorithms provided a correct classification rate higher than 84%. The best classification performance was provided by the LDA classifier with spectral features from oximetry data. An accuracy of 87.61% (91.05% sensitivity and 82.61% specificity) and an AROC of 0.925 were reached. Additionally, other configurations provided similar results. Classification of spectral features by means of LR achieved 86.73% accuracy (79.10% sensitivity and 97.83% specificity) and 0.920 AROC. LDA with nonlinear features provided an accuracy of 86.73% (88.06% sensitivity and 84.78% specificity) and an AROC of 0.912.

Minimum differences can be appreciated among our classifiers when classification accuracy is observed. In addition, we considered AROC statistics in order to assess the predictive ability of the proposed algorithms [45]. According to our results, classifiers based on KNN exhibited poor robustness. They provided significantly lower values of AROC than QDA, LDA and LR. Similarly to QDA and LDA, KNN is a parametric method to approximate class-conditional densities [20, 23]. Nevertheless, techniques based on DA represent a more suitable approach for the proposed problem. Also, LR has shown to be an efficient method to diagnose OSA from oximetry data. LR represents a regression approach to classification. In contrast to QDA, LDA and KNN, class posterior probabilities are directly estimated [23]. The LR classifier with spectral and nonlinear features achieved the highest

AROC with a value of 0.930. Similar results were obtained by QDA and LDA. However, LR avoids prior assumptions about normality and homocedasticity of the input features.

We compared the diagnostic performance of our classification algorithms with that of traditional oximetry indices. ODI2, ODI3, ODI4 and CT90 represent a classic approach for automated OSA diagnosis from SaO_2 signals [7]. These indices were evaluated on our SaO_2 signal database. For each of them, the threshold value that provided the highest accuracy on the training set was selected as optimum. Subsequently, sensitivity, specificity and accuracy were computed by applying this threshold to signals from the test set. Additionally, we computed AROC from test data. Results are summarised in Table 4.

INSERT TABLE 4 AROUND HERE

According to accuracy and AROC values, oximetry indices showed poor diagnostic capability. They provided high specificity but low sensitivity. Indices ODI2 and ODI3 achieved the best accuracy with a value of 83.19%. On the other hand, CT90 provided the highest AROC with a value of 0.774. Therefore, the algorithms proposed in our study significantly improved the diagnostic capability of classic oximetry indices.

Previously, other researchers analysed the utility of these indices in OSA diagnosis. Netzer et al. [9] reviewed several studies focused on this topic. The values of sensitivity ranged from 31 to 98% and for specificity from 41 to 100%. The study by Vázquez et al. [10] reported the best diagnostic ability by means of ODI4. A sensitivity of 98% and a specificity of 88% were reached. However, a conservative definition of OSA (AHI $\geq$ 15 events/h) was applied. Additionally, the definition of arousal differs from the criteria proposed by The Atlas Task Force [47]. As a result, the prevalence of OSA in the population under study could be affected. Magalang et al. [11] compared the diagnostic accuracy of several oximetry indices.

They were ranked according to the AROC value computed on a database with 224 subjects. OSA was defined as AHI ≥ 15 events/h. The Δ index (a measure of the variability of SaO_2) provided the best diagnostic ability with 0.881 AROC, 91% sensitivity and 59% specificity. ODI2, ODI3 and ODI4 provided slightly lower performance with AROC values of 0.868, 0.873 and 0.852, respectively. The combination of several indices increased AROC up to 0.9 (90% sensitivity and 70% specificity). Series et al. [12] reported a sensitivity of 85% and a specificity of 93% by using ODI2. Both central and obstructive apnoea syndromes were considered. Finally, Roche et al. [13] developed a LR model to compute the probability of having OSA. The cumulative time spent below 80% of saturation was used as input variable together with clinical information. This model was tested on a set of 108 subjects. An accuracy of 53% was reached, which reflects poor classification ability.

Our results suggest that the use of advanced signal processing techniques and classification algorithms increases the diagnostic capability of SaO_2 signals. In addition to traditional statistical classifiers assessed in this study, other classification techniques could be applied for modelling OSA diagnosis. Previously, we evaluated the performance of multilayer perceptron (MLP) and radial basis function (RBF) neural networks [35, 36]. ApEn, CTM and LZC computed from oximetry data were applied as network inputs. Neural networks represent a powerful tool for classification purposes. They have the ability to learn complex nonlinear input-output relationships and adapt themselves to the environment [20]. Moreover, no prior assumptions about the distributional form of the data are required [48]. The MLP classifier achieved a classification accuracy of 85.50% and an AROC of 0.900, while an accuracy of 86.10% and an AROC of 0.910 were obtained with the RBF network. Thus, traditional statistical classifiers provided better results than MLP and RBF networks. In addition, designing and training these neural classifiers are complex tasks since several user dependant parameters must be specified. Therefore, it could be stated that statistical classifiers

developed in this study represent a more efficient approach for the proposed OSA diagnosis problem.

Statistical pattern recognition techniques have been applied to the OSA diagnosis problem in preceding researches. However, other biomedical signals were analysed instead of SaO_2 recordings. DA was used to classify patients from ECG signals [30-33]. An accuracy of 100% was reached by LDA and QDA [31, 32]. However, borderline subjects were removed from the test set to compute final results. Similarly, KNN algorithms were used with ECG features. A sensitivity of 83.90% and a specificity of 88.50% were achieved [34]. In addition, classification of morphological features of the upper airway by means of KNN provided a sensitivity of 64.29% and a specificity of 86.67% [29]. On the other hand, a LR model was developed to process data from home videotape recordings [28]. It yielded 89% sensitivity and 77% specificity in classification of children with OSA. By comparison with the cited studies, our algorithms provided significant results in OSA detection. However, it should be taken into account that different datasets were used to estimate the classification performance of each algorithm.

There are some limitations in our study. A larger training set could increase the generalization capability of our classifiers. The requirement of normally distributed data for algorithms based on QDA and LDA is another limitation. Indeed, a nonlinear transformation had to be applied on spectral features to satisfy these conditions. Similarly, the need for optimizing K is a handicap for KNN algorithms.

In summary, several classification algorithms to assist in OSA detection were presented. They were based on traditional statistical pattern recognition techniques. Spectral and nonlinear features from SaO_2 signals were applied as inputs. QDA, LDA, KNN and LR were assessed in our study. From our results, QDA, LDA and LR have shown to be a robust tool to help in OSA diagnosis. They provided good values of both classification accuracy and

AROC. Classification of spectral features from oximetry by means of LDA provided an accuracy of 87.61% and an AROC of 0.925. This algorithm achieved the best performance from those implemented in our research. Indeed, more advanced approaches based on neural networks were improved by traditional statistical classification methods. Therefore, the proposed algorithms could be useful for screening of OSA patients since only SaO_2 signals are required. Thus, the demand for polisomnographic studies could be reduced.

Acknowledgements

This research has been supported by Ministerio de Ciencia e Innovación and Consejería de Sanidad de la Junta de Castilla y León under projects TEC2008-02241 and SAN673/VA03/08, respectively.

References

[1] Malhotra A, White DP. Obstructive sleep apnoea. Lancet 2002;360:237-45.

[2] Qureshi A, Ballard RD. Obstructive sleep apnea. J Allergy Clin Immunol 2003;112:643-51.

[3] Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. Am J Respir Crit Care Med 2002;165:1217-39.

[4] Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997;20:705-6.

[5] Gagnadoux F, Pelletier-Fleury N, Philippe C, Rakatonanahary D, Bernard F. Home unattended vs hospital telemonitored polysomnography in suspected obstructive sleep apnea syndrome: a randomized crossover trial. Chest 2002;121:753-8.

[6] Jubran A. Pulse oximetry. Intensive Care Med 2004;30:2017-20.

[7] Bloch KE. Getting the most out of nocturnal pulse oximetry. Chest 2003;124:1628-30.

[8] Rodríguez JM, De Lucas P, Sánchez MJ, Izquierdo JL, Peraíta R, Cubillo JM. Usefulness of the visual analysis of night oximetry as a screening method in patients with

suspected clinical obstructive sleep apnea syndrome. Arch Bronconeumol 1996;32:437-41.

[9] Netzer N, Eliasson AH, Netzer C, Kristo DA. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. Chest 2001;120:625-33.

[10] Vázquez JC, 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 apnoea. Thorax 2000;55:302-7.

[11] 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;124:1694:701.

[12] Sériès F, Kimoff RJ, Morrison D, Leblanc MH, Smilovitch M, Howlett J, Logan AG, Floras JS, Bradley TD. Prospective evaluation of nocturnal oximetry for detection of sleep-related breathing disturbances in patients with chronic heart failure. Chest 2005;127:1507-14.

[13] Roche N, Herer B, Roig C, Huchon G. Prospective testing of two models based on clinical and oximetric variables for prediction of obstructive sleep apnea. Chest 2002;121:747-52.

[14] Zamarrón C, Gude F, Barcala J, Rodríguez JR, Romero PV. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome. Chest 2003;123:1567-76.

[15] 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.

[16] Del Campo F, Hornero R, Zamarrón C, Abásolo DE, Álvarez D. Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea. Artif Intell Med 2006;37:111-8.

[17] Álvarez D, Hornero R, García M, Del Campo F, Zamarrón. 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.

[18] Hornero R, Álvarez D, Abásolo D, Del Campo F, Zamarrón C. Utility of aproximate entropy from overnight pulse oximetry data in the diagnosis of the obstructive sleep apnea syndrome. IEEE Trans Biomed Eng 2007;54:107-13.

[19] Kulikowski CA. Pattern recognition approach to medical diagnosis. IEEE Trans Syst Man Cybern 1970;6:173-8.

[20] Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 2000;22:4-37.

[21] Friedman JH. Regularized discriminant analysis. J Am Stat Assoc 1989;84:165-75.

[22] Bishop CM. 'Neural networks for pattern recognition'. Oxford University Press Inc., New York, 1995.

[23] Holmström L, Koistinen P, Laaksonen J, Oja E. Neural and statistical classifiers: taxonomy and two case studies. IEEE Trans Neural Netw 1997;8:5-17.

[24] Jobson JD. 'Applied multivariate data analysis: categorical and multivariate methods'. Springer, New York, 1991.

[25] Jekova I, Bortolan G, Christov I. Assessment and comparison of different methods for heartbeat classification. Med Eng Phys 2008;30:248-57.

[26] Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. IEEE Trans Biomed Eng 2002;49:963-74.

[27] Curran E, Sykacek P, Stokes M, Roberts SJ, Penny W, Johnsrude I, Owen AM. Cognitive tasks for driving a brain-computer interfacing system: a pilot study. IEEE Trans Neural Syst Rehabil Eng 2003;12:48-54.

[28] Sivan Y, Kornecki A, Schonfeld T. Screening obstructive sleep apnoea syndrome by home videotape recording in children. Eur Respir J 1996;9:2127-31.

[29] Caballero P, Álvarez-Sala R, García-Río F, Prados C, Hernán MA, Villamar J, Álvarez-Sala JL. CT in the evaluation of the upper airway in healthy subjects and in patients with obstructive sleep apnea syndrome. Chest 1998;113:111-6.

[30] Schrader M, Zywietz C, Von Einem V, Widiger B, Joseph G. Detection of sleep apnea in single channel ECGs from the PhysioNet database. Comput Cardiol 2000;27:263-6.

[31] 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 apnoea. IEEE Trans Biomed Eng 2003;50:686-96.

[32] de Chazal P, Penzel T, Heneghan C. Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. Physiol Meas 2004;25:967-83.

[33] Redmond SJ, Heneghan C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. IEEE Trans Biomed Eng 2006;53:485-96.

[34] Méndez MO, Ruini DD, Villantieri OP, Matteucci M, Penzel T, Cerutti S, Bianchi AM. Detection of sleep apnea from surface ECG based on features extracted by an autoregressive model. Proceedings of the 29th Annual International Conference of the IEEE EMBS 2007;6105-8.

[35] Marcos JV, Hornero R, Álvarez D, del Campo F, López M, Zamarrón C. Radial basis function classifiers to help in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. Med Biol Eng Comput 2008;46:323-32.

[36] Marcos JV, Hornero R, Álvarez D, del Campo F, Zamarrón C, López M. Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry. Comput Methods Programs Biomed 2008;92:79-89.

[37] Rechtschaffen A, Kales A. 'A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects'. Brain Information Services, Brain Research Institute, University of California, Los Angeles, 1968.

[38] Welch PD. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodogram. IEEE Trans Audio Electroacoust 1967;15:70-3.

[39] Pincus SM. Assessing serial irregularity and its implications for health. Ann N Y Acad Sci 2001;954:245-67.

[40] Cohen ME, Hudson DL, Deedwania PC. Applying continuous chaotic modelling to cardiac signals. IEEE Eng Med Biol Mag 1996;15:97-102.

[41] Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans Inf Theory 1976;22:75-81.

[42] Bazan JG, Szczuka MS, Wojna A, Wojnarski M. On the evolution of rough set exploratory system. Lect Notes Comput Sci 2004;3066:592-601.

[43] Nabney IT. 'Netlab: algorithms for pattern recognition'. Springer, Berlin, 2002.

[44] Hanley JA, McNeil BJ. The meaning and use of the area under a receiving operating characteristic (ROC) curve. Radiology 1982;143:29-36.

[45] Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 2005;17:299-310.

[46] Bradley AP. The use of the area under the ROC curve I the evaluation of machine learning algorithms. Pattern Recognit 1997;30:1145-59.

[47] The Atlas Task Force. EEG arousals: scoring rules and examples. A preliminary report from the Sleep Disorder task Force of the American Sleep Disorders Association. Sleep 1992;15:173-84.

[48] Zhang GP. Neural networks for classification: a survey. IEEE Trans Syst Man Cybern C Appl Rev 2000;30:451-62.

Tables

Table 1. Demographic and clinical statistics of all subjects, training set and test set. Data are presented as mean ± standard deviation. OSA Positive: patients with a positive diagnosis of obstructive sleep apnoea; OSA Negative: patients with a negative diagnosis of obstructive sleep apnoea; BMI: body mass index; AHI: apnoea/hypopnoea index computed as events for hourly periods.

ALL SUBJECTS			
	All	OSA Positive	OSA Negative
Subjects (n)	187	111	76
Age (years)	57.97 ± 12.84	58.30 ± 12.88	57.57 ± 12.87
Males (%)	78.61	84.68	69.74
BMI (kg/m^2)	29.54 ± 5.51	30.45 ± 4.92	28.42 ± 6.02
Recording Time (h)	8.19 ± 0.62	8.17 ± 0.75	8.22 ± 0.33
AHI (n/h)		40.07 ± 19.64	2.04 ± 2.36
TRAINING SET			
	All	OSA Positive	OSA Negative
Subjects (n)	74	44	30
Age (years)	58.25 ± 12.14	56.73 ± 13.61	59.59 ± 10.19
Males (%)	75.68	79.55	70.00
BMI (kg/m^2)	29.62 ± 5.71	30.19 ± 5.09	28.93 ± 6.40
Recording Time (h)	8.22 ± 0.41	8.20 ± 0.49	8.25 ± 0.27
AHI (n/h)		38.11 ± 18.18	2.60 ± 2.51
TEST SET			
	All	OSA Positive	OSA Negative
Subjects (n)	113	67	46
Age (years)	57.91 ± 13.39	59.37 ± 12.38	56.03 ± 14.54
Males (%)	80.53	88.06	69.57
BMI (kg/m^2)	29.49 ± 5.41	30.63 ± 4.84	28.07 ± 5.80
Recording Time (h)	8.17 ± 0.72	8.14 ± 0.88	8.20 ± 0.37
AHI (n/h)		41.36 ± 20.58	1.67 ± 2.21

Table 2. Mean values of spectral and nonlinear features used as inputs to our classifiers. Results are presented for all subjects under study, training set and test set. Data are presented as mean ± standard deviation. S_T: total area under the PSD; S_B: area of the PSD enclosed in the band between 0.010 and 0.033 Hz; PA: peak amplitude of the PSD in that band; ApEn: approximate entropy; CTM: central tendency measure; LZC: Lempel-Ziv complexity.

ALL SUBJECTS			
	All	OSA Positive	OSA Negative
S_T	15.21 ± 25.75	23.31 ± 29.54	3.37 ± 11.22
S_B	5.88 ± 11.51	9.51 ± 13.78	0.59 ± 1.53
PA	840.75 ± 1841.80	1363.05 ± 2239.52	77.93 ± 251.64
ApEn	0.80 ± 0.38	1.03 ± 0.28	0.47 ± 0.25
CTM	0.47 ± 0.28	0.30 ± 0.20	0.71 ± 0.18
LZC	0.49 ± 0.18	0.60 ± 0.13	0.34 ± 0.14
TRAINING SET			
	All	OSA Positive	OSA Negative
S_T	14.86 ± 22.69	21.17 ± 23.81	5.60 ± 17.50
S_B	4.82 ± 8.16	7.50 ± 9.55	0.88 ± 2.35
PA	718.46 ± 1621.60	1122.41 ± 1986.94	126.01 ± 393.52
ApEn	0.79 ± 0.37	1.01 ± 0.26	0.46 ± 0.23
CTM	0.48 ± 0.27	0.32 ± 0.19	0.71 ± 0.18
LZC	0.48 ± 0.17	0.58 ± 0.12	0.33 ± 0.12
TEST SET			
	All	OSA Positive	OSA Negative
S_T	15.44 ± 27.67	24.72 ± 32.86	1.92 ± 2.66
S_B	6.58 ± 13.24	10.83 ± 15.89	0.40 ± 0.48
PA	920.84 ± 1975.65	1521.07 ± 2392.28	46.58 ± 56.57
ApEn	0.81 ± 0.39	1.04 ± 0.29	0.48 ± 0.26
CTM	0.46 ± 0.29	0.29 ± 0.21	0.71 ± 0.17
LZC	0.50 ± 0.19	0.61 ± 0.13	0.34 ± 0.14

Table 3. Diagnostic results provided by the classification algorithms implemented in our study. QDA: quadratic discriminant analysis; LDA: linear discriminant analysis; KNN: *K*-nearest neighbour; LR: logistic regression; Spec: input pattern vector composed by spectral features; NonLin: input pattern vector composed by nonlinear features; Spec-NonLin: input pattern vector composed by both spectral and nonlinear features; Se: sensitivity; Sp: specificity; Ac: accuracy; AROC: area under ROC curve.

Classifier	Input Pattern	Se (%)	Sp (%)	Ac (%)	AROC
QDA					
	Spec†	89.55	78.26	84.96	0.907
	NonLin	86.57	82.61	84.96	0.905
	Spec-NonLin†	91.05	78.26	85.84	0.913
LDA					
	Spec†	91.05	82.61	87.61	0.925
	NonLin	88.06	84.78	86.73	0.912
	Spec-NonLin†	86.57	80.44	84.07	0.925
KNN					
($K = 26$)	Spec	85.07	84.78	84.96	0.737
($K = 21$)	NonLin	88.06	84.78	86.73	0.826
($K = 21$)	Spec-NonLin	88.06	84.78	86.73	0.822
LR					
	Spec	79.10	97.83	86.73	0.920
	NonLin	86.57	82.61	84.96	0.907
	Spec-NonLin	85.08	86.96	85.84	0.930

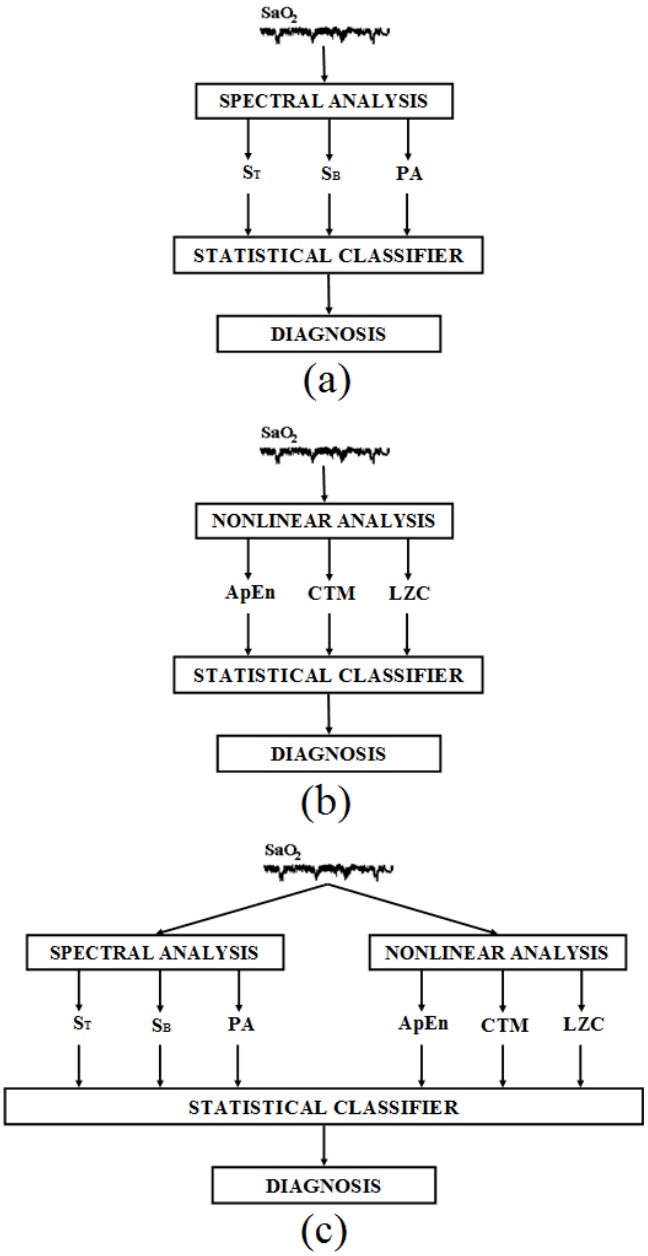
†: Transformations of spectral features by means of the base 10 logarithm.

Table 4. Diagnostic results achieved by traditional oximetry indices from our database of SaO_2 recordings. ODI4: oxygen desaturation index over 4%; ODI3: oxygen desaturation index over 3%; ODI2: oxygen desaturation index over 2%; CT90: cumulative time spent below 90% of saturation; *l*: optimum threshold computed from training data; Se: sensitivity reached on the test set; Sp: specificity reached on the test set; Ac: accuracy reached on the test set; AROC: area under ROC curve.

Index	*l*	Se (%)	Sp (%)	Ac (%)	AROC
ODI2	9	76.12	93.48	83.19	0.706
ODI3	8	76.12	93.48	83.19	0.725
ODI4	7.6	73.13	95.65	82.30	0.758
CT90	11	68.66	95.65	79.65	0.774

Figure legends

Fig. 1. The three classification schemes proposed for each of the evaluated statistical pattern recognition techniques. (a) Using spectral features as inputs to the classifier; (b) Using nonlinear features as inputs to the classifier; (c) Using both spectral and nonlinear features as inputs to the classifier.



Conflict of interest statement

There are no conflicts of interest that could inappropriately influence this research work.