Automated Prediction of the Apnea-Hypopnea Index from Nocturnal Oximetry Recordings

J. Víctor Marcos, Student Member, IEEE, Roberto Hornero, Senior Member, IEEE, Daniel Álvarez, Student Member, IEEE, Mateo Aboy, Member, IEEE, and Félix Del Campo

Abstract—Nocturnal polysomnography (PSG) is the gold-standard for sleep apnea-hypopnea syndrome (SAHS) diagnosis. It provides the value of the apnea-hypopnea index (AHI), which is used to evaluate SAHS severity. However, PSG is costly, complex, and time-consuming. We present a novel approach for automatic estimation of the AHI from nocturnal oxygen saturation (SaO₂) recordings and the results of an assessment study designed to characterize its performance. A set of 240 SaO₂ signals was available for the assessment study. The data were divided into training (96 signals) and test (144 signals) sets for model optimization and validation, respectively. Fourier-time domain and frequency-domain signals were used to quantify the effect of SAHS on SaO₂ recordings. Regression analysis was performed to estimate the functional relationship between the extracted features and the AHI. Multiple linear regression (MLR) and multilayer perceptron (MLP) neural networks were evaluated. The MLP algorithm achieved the highest performance with an intraclass correlation coefficient (ICC) of 0.91. The proposed MLP-based method could be used as an accurate and cost-effective procedure for SAHS diagnosis in the absence of PSG.

Index Terms—Apnea-hypopnea index, multiple linear regression, multilayer perceptron neural networks, oxygen saturation, regression analysis, sleep apnea-hypopnea syndrome.

I. INTRODUCTION

The sleep apnea-hypopneasynrome (SAHS) is characterized by repetitive complete (apnea) or partial (hypopnea) collapse of the upper airway during sleep [1]. Apnea events are associated to hypoxemia, heart rate variations, and arousals. Epidemiological data support the finding that SAHS may have a role in the initiation or progression of diverse respiratory, cardiovascular, and cerebrovascular diseases [1], [2]. The incidence of SAHS has been estimated at 5% of adults in western countries [2]. Furthermore, previous studies revealed that a high percentage of patients (82% of men and 93% of women) with moderate or severe SAHS might remain undiagnosed [3]. Therefore, early detection and treatment of SAHS are required in order to prevent long-term effects and end-organ damages.

Nowadays, nocturnal polysomnography (PSG) is the gold-standard for SAHS diagnosis [4]. PSG studies are performed in special sleep units and generally involve monitoring several physiological recordings such as electrocardiograms (ECG), electroencephalograms (EEG), electromyograms (EMG), electrooculograms (EOG), airflow signals, respiratory effort, and oxygen saturation (SaO₂) or oximetry [1]. Polysomnographs are usually provided with specific software to assist medical doctors in the interpretation of these signals. Although it may be used during the examination of PSG data, manual analysis performed by a sleep specialist is required for accurate identification of apnea/hypopnea episodes. The number of detected events is divided by the hours of sleep to compute the apnea-hypopnea index (AHI), which is used to assess SAHS severity [4]. However, PSG studies have drawbacks since they are costly, time-consuming, and require subjects to be overnight in a special medical facility [5]. Additionally, the demand for PSG studies is progressively growing as people and clinicians are becoming aware of SAHS whereas the available infrastructure is insufficient to support it [6]. Consequently, research focused on alternative diagnostic methods that overcome some of the limitations associated to PSG has notably increased.

New techniques for simplified SAHS detection have been commonly based on the analysis of a reduced set of data. The utility of clinical and demographic variables [7], [8] as well as ECG [9], [10] has been widely studied. In the context of this problem, SaO₂ signals recorded through nocturnal pulse oximetry are of special interest since they can be easily acquired and enable for portable monitoring [11]–[18]. Pulse oximetry is a noninvasive technique used to monitor arterial blood oxygen saturation. Oximetry recordings contain essential information about SAHS and play a crucial role to interpret PSG studies. Apneas and hypopneas are usually accompanied by marked desaturation events due to the lack of airflow. As a result, patients with SAHS tend to present unstable SaO₂ signals due to frequent drops in the saturation value. A different behavior tends to be observed in healthy patients. Their recordings reflect normal ventilation, which corresponds with a saturation value near 96% and the absence of repetitive abrupt changes in the SaO₂ profile [19], [20].

Several signal processing techniques have been proposed to extract relevant diagnostic information from SaO₂ recordings.
Preceding studies showed that spectral and nonlinear features can reflect the occurrence of apneas [11], [13], [21]. Moreover, highly accurate diagnostic models based on logistic regression [17] or neural networks [14], [15] have been built from these features. However, these algorithms only provide a categorical decision for each subject (labeled as non-SAHS or SAHS) and no information about SAHS severity is given. Automatic computation of the AHI has been addressed to obtain a more detailed characterization of the patient’s state. In addition to clinical and demographic variables [7], [8], SaO₂ signals have been frequently used for this purpose. The linear relationship between the number of desaturation events and the AHI was previously evaluated [22], [23]. Multivariate pattern analysis techniques were also applied in order to deal with several measurements from oximetry simultaneously [12], [24]. Furthermore, the combination of ECG and SaO₂ features has been assessed in a preceding work [9]. However, these previous studies did not evaluate the utility of the estimated AHI to rank SAHS severity and only correlation analysis was performed.

In the present study, we hypothesize that an accurate estimation of the AHI can be automatically derived from SaO₂ data, providing more useful diagnostic information to practitioners compared with other methods based on a two-class classification approach. We propose modeling SAHS diagnosis as a regression task. Time-domain and frequency-domain features from oximetry data were used to reflect the occurrence of apneas and hypopneas during sleep. An approximation to the functional relationship between the extracted feature pattern and the AHI was derived from multivariate regression analysis. Multiple linear regression (MLR) models and multilayer perceptron (MLP) neural networks were used for this purpose [25]. The overall objective was to evaluate the degree of severity of SAHS (no SAHS, mild-SAHS, moderate-SAHS, and severe-SAHS) from the estimated AHI. As a result, a complete description about SAHS would be provided in a fully automatic way.

II. SUBJECTS AND DATA

A group of 240 subjects suspected of suffering from SAHS were included in the study. All of them presented typical symptoms such as daytime sleepiness, snoring, or apnea events reported by the subject or a bedmate. For a decision threshold of \( AHI = 10 \text{ h}^{-1} \), a positive diagnosis of SAHS would correspond to 160 of these subjects while the remaining 80 subjects would be non-SAHS cases.

The initial population was randomly divided into training and test sets. The training set was composed of 96 subjects and was used for model optimization. Once this process is completed, a test set composed of previously unseen samples is required to objectively assess the performance of the estimator. In this assessment study, the oximetry signals from 144 subjects composed the test set. Table I summarizes clinical data of subjects in both sets.

Each subject underwent complete overnight PSG. It was carried out from midnight to 08:00 AM in the sleep unit of Hospital Universitario Río Hortega, Valladolid, Spain. The Review Board on Human Studies approved the protocol and subjects gave their consent to participate in the study. Patients were monitored using a polysomnograph (Alice 5, Respironics, Philips Healthcare, The Netherlands). ECG, EEG, chin EMG, EOG, nasal airflow, and body position were recorded and stored on a computer. Simultaneously, a Nonin PureSAT pulse oximeter (Nonin Medical Inc., USA) was used to record SaO₂ signals at a sampling frequency (fₛ) of 1 Hz. The averaging time was set to 3 s, following the recommendations from the American Academy of Sleep Medicine [26]. Oximetry recordings were saved to separate files to be offline processed. Artifacts represented by drops to zero were removed. Finally, a sleep specialist analyzed the complete set of recordings using the rules proposed by Rechtschaffen and Kales [27] and derived the AHI for each subject.

III. METHODS

The proposed method comprises two different stages. In the first one, feature extraction from SaO₂ data is carried out in order to capture the dynamical behavior of the signal. The second stage corresponds to regression analysis, which aims to provide an analytical expression for the AHI as a function of the extracted features.

A. Feature Extraction

In the feature extraction phase, information in the SaO₂ recording was summarized into a reduced set of measurements or features. They are defined in order to represent different signal properties related to the degree of SAHS severity.

Prior domain knowledge about the influence of apnea events on SaO₂ dynamics was used to define a set of 14 measurements. According to the domain used for SaO₂ analysis, the extracted features were divided into two groups: time-domain and frequency-domain features.

1) Time-Domain Analysis: Marked drops in the amplitude of oximetry signals reflect desaturation events due to apneas. Subjects with low AHI are expected to present SaO₂ tracings with minor oscillations around 96% during most of the time [19]. In contrast, a high AHI reflects the repetition of apneas, resulting in SaO₂ recordings with marked instability. Conventional statistics and nonlinear methods were used to characterize this dynamic behavior in the time domain.

Statistical analysis represents an easy-to-use tool to study SaO₂ signals. The distribution of SaO₂ values tends to reflect different properties depending on the AHI. Mean (\( μ_S \)), variance (\( σ_S^2 \)), skewness (\( γ_S \)), and kurtosis (\( δ_S \)) were computed to quantify the central tendency, the degree of dispersion, the asymmetry, and the peakedness, respectively, for variable \( s \) representing

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>96</td>
</tr>
<tr>
<td>Age (years)</td>
<td>52.35 ± 13.76</td>
</tr>
<tr>
<td>Males (%)</td>
<td>77.08</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>29.83 ± 4.17</td>
</tr>
<tr>
<td>AHI (h⁻¹)</td>
<td>24.75 ± 25.19</td>
</tr>
</tbody>
</table>
the SaO2 value. These measurements are defined as [28]
\[ \mu_S = \sum_k s_k \rho_S (s_k) \]  
(1)
\[ \sigma^2_S = \sum_k [s_k - \mu_S]^2 \rho_S (s_k) \]  
(2)
\[ \gamma_S = \left( \frac{1}{\sigma_S^4} \right) \sum_k [s_k - \mu_S]^4 \rho_S (s_k) \]  
(3)
\[ \delta_S = \left( \frac{1}{\sigma_S^6} \right) \sum_k [s_k - \mu_S]^6 \rho_S (s_k) \]  
(4)
where \( \rho_S \) denotes the probability density function of variable \( s \).

It was obtained from the relative frequency observed in the sequence of SaO2 samples \( s = \{s_1, \ldots, s_i, \ldots, s_T \} \).

On the other hand, nonlinear analysis of SaO2 signals by means of approximate entropy (ApEn) [29], central tendency measure (CTM) [30], and Lempel-Ziv complexity (LZC) [31] was performed to measure irregularity, variability, and complexity, respectively. As stated by previous studies, these properties are usually more pronounced in oximetry recordings from subjects with higher AHI [13], [21].

To compute ApEn, patterns \( s_i \) composed of \( m \) consecutive samples from the original sequence \( s \) are obtained. For a given pattern, the regularity (frequency) \( C_{m}^r (i) \) is expressed as
\[ C_{m}^r (i) = \frac{N^m (i)}{T - m + 1} \]  
(5)
where \( N^m (i) \) denotes the number of patterns \( s_j \) of length \( m \) to a distance less or equal than \( r \) from \( s_i \). The ApEn is defined as the following ratio [29]:
\[ \text{ApEn} (m, r) = \lim_{T \to \infty} \left[ \frac{1}{T - m + 1} \sum_{i=1}^{T-m} \ln C_{m}^r (i) \right. \]
\[ - \frac{1}{T - m^* + 1} \sum_{i=1}^{T-m^*} \left. \ln C_{m^*}^r (i) \right] \]  
(6)
where \( m^* = m + 1 \). Briefly, ApEn measures the logarithmic likelihood that runs of patterns that are close (within \( r \)) for \( m \) contiguous observations remain close (within the same tolerance width \( r \)) on incremental comparisons [29].

CTM is obtained from second-order difference scatter plots representing \( (s_{k+2} - s_{k+1}) \) vs. \( (s_{k+1} - s_k) \). A circular region of radius \( \rho_{CTM} \) is defined around the origin to compute CTM. The number of points that fall in this region is counted and divided by the total number of points [30]
\[ \text{CTM} = \frac{\sum_{k=1}^{T-2} \delta_k}{T - 2} \]  
(7)
where \( \delta_k = 1 \) if the \( k \)th point is inside the circle and 0 otherwise.

LZC is a nonparametric measure of complexity in a signal. It is related to the number of distinct substrings and the rate of their recurrence along a given sequence. To compute LZC, the original signal \( s \) must be transformed into a two-symbol sequence \( p = \{p_1, \ldots, p_i, \ldots, p_T \} \). Each SaO2 sample is compared with the median value of the samples to perform the transformation.

Then, the sequence \( p \) is scanned from left to right and the complexity counter \( c(T) \) is increased by one unit every time a new subsequence of consecutive characters is encountered. The value of LZC is given by [31]
\[ \text{LZC} = \frac{c(T)}{b(T)} \]  
(8)
where \( b(T) \) is a normalization factor. It is given by
\[ b(T) = \frac{T}{\log_2 T}. \]  
(9)

Oximetry recordings are generally nonstationary. Thus, each time-domain feature was computed by dividing the signal into 512-sample epochs, computing the value of the feature for each epoch and averaging over all the epochs. Several design parameters must be adjusted for the proposed nonlinear methods.

They were set to the optimum values proposed in previous studies [13], [17]. In the case of ApEn, the sequence length \( m \) was set to 1 while the optimum width of the tolerance window \( r \) was fixed at 0.25 times the standard deviation of the samples in each signal epoch [13]. To compute CTM, a radius \( \rho_{CTM} = 1 \) was selected as optimum [17]. Finally, LZC was computed by converting SaO2 samples in each epoch into a 0–1 sequence. Each sample was compared with the median value from the epoch to transform the data [17].

2) Frequency-Domain Analysis: Previous studies have shown the influence of repeated apnea events on the spectral properties of SaO2 signals. Specifically, it has been found that signal power associated with frequency components between 0.010 and 0.033 Hz tends to increase with the AHI [11]. Apneas originate phase-lagged changes in SaO2 signals. Their duration usually ranges from 30 s to 2 min, including the awakening response after the event. Patients suffering from SAHS may have several consecutive episodes of apneas or hypopneas. Thus, the repetition of these events will be produced at a rate between 30 s and 2 min, which correspond with the frequency values mentioned before. Thus, high-power components in this range denote fluctuations in oximetry recordings due to periods with repetitive apneas.

The nonparametric Welch’s method was used to compute the power spectral density (PSD) of oximetry recordings [32]. The original series \( s \) was divided into \( M \) overlapping sequences of length \( L \) by applying a window function \( v = \{v_1, \ldots, v_i, \ldots, v_L \} \). The modified periodogram was computed for each of them by using the fast Fourier transform (FFT):
\[ \hat{P}_m (f_k) = \frac{1}{LU} \left| \sum_{i=0}^{L-1} s_i v_i e^{-j2\pi ki/L} \right|^2 \]  
(10)
where
\[ U = \frac{1}{L} \sum_{i=0}^{L-1} |v_i|^2. \]  
(11)
The estimation of the PSD was obtained as the average of the periodograms
\[
PSD(f_k) \approx \hat{P}(f_k) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{P}_m(f_k).
\] (12)

A 512-sample Hanning window and 50% overlapping were applied to estimate the PSD of SaO\textsubscript{2} signals using the Welch’s method. The length of the FFT for each signal segment was set to 1024 samples.

Initially, statistical analysis was carried out in order to characterize the spectral properties of the signal. The variable representing the frequency component (\(f\)) was considered. The normalized PSD was used as its probability density function (PDF). Mean (\(\mu_F\)), variance (\(\sigma_F^2\)), skewness (\(\gamma_F\)), and kurtosis (\(\delta_F\)) were computed according to the expressions in (1), (2), (3), and (4), respectively.

In order to reflect the incidence of apnea events, three additional features were derived from the PSD function: the total power of the SaO\textsubscript{2} signal (\(S_T\)), the power in the band between 0.010, and 0.033 Hz (\(S_B\)), and the most significant frequency component in that band (PA). They are given by the following expressions:
\[
S_T = \sum_{f=0}^{f_{1/2}} PSD(f) \quad (13)
\]
\[
S_B = \sum_{f=0.010}^{0.033} PSD(f) \quad (14)
\]
\[
PA = \max_{PSD(f)}, f \in [0.010, 0.033] \ (Hz). \quad (15)
\]

Prior to regression analysis, each of the extracted features was normalized to have a zero mean and unit variance distribution in order to avoid differences between their magnitudes.

B. Regression Analysis

Regression techniques were used to estimate the function relating the AHI with the set of SaO\textsubscript{2} features. A 1-D continuous variable (\(t\)) was used to model the AHI value (target variable). The extracted features were grouped into a pattern \(x = (x_1, x_2, \ldots, x_d)\) representing the multivariate independent variable. The approximation is built from a finite training set \(D\) composed of \(N\) input–output independent pairs \(\{(x_n, t_n)\}_{n=1}^{N}\). Training samples are assumed to satisfy the following condition:
\[
t_n = h(x_n) + \varepsilon_n \quad (16)
\]
where \(x_n\) is known, \(h(\cdot)\) is the true function, and \(\varepsilon_n\) is an additive stochastic component (noise) characterized by a zero-mean Gaussian distribution [25].

Regression techniques define a mapping function \(y(x, w)\) that represents an approximation to \(h(\cdot)\), where \(w\) denotes a set of model adaptive parameters or weights. According to the maximum likelihood principle, these weights must be chosen in order to minimize the sum-of-squares error (\(E_D\)) between the actual and estimated AHI for patterns in \(D\) [25]
\[
E_D = \frac{1}{2} \sum_{n=1}^{N} \left[ y(x_n, w) - t_n \right]^2. \quad (17)
\]

As a result, the output of the model approximates the conditional average of the target data, which is known as the regression of \(t\) conditioned on \(x\) [25]
\[
y(x, w^*) = E[t | x] \quad (18)
\]
where \(w^*\) denotes the set of model parameters that minimizes the sum-of-squares error function. In this study, the performance of two regression techniques was analyzed: MLR and MLP networks.

1) Multiple Linear Regression: MLR models assume a linear expression for the regression function. Thus, the mapping implemented by the algorithm takes the form [28]
\[
y(x, w) = w_0 + w_1 x_1 + \cdots + w_d x_d = w^T x \quad (19)
\]
where \(w = (w_0, w_1, \ldots, w_d)^T\) are the adaptive parameters and \(x = (1, x_1, x_2, \ldots, x_d)\). Model optimization according to sum-of-squares error minimization yields the following solution [28]
\[
w = X^+ t \quad (20)
\]
where rows of matrix \(X\) are training patterns \(x_n\), \(X^+\) is its pseudoinverse matrix and vector \(t = (t_1, t_2, \ldots, t_N)^T\) contains the target values corresponding to the training patterns.

2) Multilayer Perceptron Networks: MLP networks are models for expressing knowledge using a connectionist paradigm inspired in the human brain. They are composed of multiple simple units or neurons known as perceptrons, which are characterized by an activation function \(g(t)\) [33]. Perceptrons are arranged in several interconnected layers. Each network connection between two of them is associated with a network adaptive parameter or weight \((w_j)\). The response of the network to the input pattern is provided by units in the final layer (output layer). The remaining network layers are referred to as hidden layers [33].

Typically, MLP networks with a single hidden layer composed of nonlinear perceptrons (i.e., with a nonlinear activation function) are implemented since they are capable of universal approximation [34]. The number of units in this layer must be determined by the designer. The configuration of the output layer depends on the specifications of the problem. The proposed regression task aims to approximate a 1-D continuous variable representing the AHI. Thus, a single output unit with a linear activation function is required [35]. Accordingly, the network output is given by
\[
y(x, w) = \sum_{j=1}^{N_h} \left[ w_j g_t \left( \sum_{i=1}^{d} w_{i,j} x_i + b_j \right) + b_0 \right] \quad (21)
\]
where \(w\) is the weight vector composed of all the adaptive parameters (weights and biases) in the network, \(N_{h}\) is the number of hidden units, \(w_j\) is the weight connecting hidden unit \(h_j\) with the output unit, \(b_0\) is the bias associated with the output unit, \(w_{i,j}\) is the weight connecting the input feature \(i\) with hidden unit \(h_j\), and \(b_j\) is its associated bias.
Weights are adjusted from samples in the training set during the training or learning process. The aim is to infer the statistical properties of the problem into the network. According to the maximum likelihood principle, weights are chosen in order to minimize the sum-of-squares error function. Second-order nonlinear optimization algorithms are used for this purpose [25].

Weight decay regularization can be applied to control network complexity and increase generalization capability. As stated by the bias-variance trade-off, networks with a large number of adaptive parameters (compared to the size of the training set) may overfit the data [25], [33]. Weight decay favors small weights (smooth mappings) by adding a penalty term to the error function $E_D$. It is equal to the sum of the squares of the network weights [25]

$$E_T = E_D + \nu \sum_{i} w_i^2 = \frac{1}{2} \sum_{n=1}^{N} [y(x_n, w) - t_n]^2 + \nu \sum_{i} w_i^2$$

(22)

where $\nu$ is known as the regularization parameter.

IV. RESULTS

A. Intraclass Correlation Coefficient

Regression methods were evaluated using the intraclass correlation coefficient (ICC), which is a measure of reliability between observers [36]. The model (2,1) defined by Shrout and Fleiss [36] for ICC was considered since it takes into account both the random and systematic errors. The ICC ranges from -1 to 1. A negative value indicates that more differences are observed within (error in the approximation) than between subjects. ICC values close to one reflect good reliability of the algorithm.

B. Design and Optimization From the Training Set

The MLR model has a unique solution given the training set $D$ [28]. Consequently, no design is required. Table II shows the coefficients $(w_1, \ldots, w_d)$ associated to each of the input features according to the MLR equation in (19). The additional bias term was $w_0 = 25.75$ h⁻¹.

On the other hand, MLP models require a thorough design to achieve high generalization performance. According to the bias-variance trade-off, both excessively simple and complex models will lead to poor generalization due to underfitting and overfitting, respectively [25]. Therefore, model selection is required in order to find the optimum network complexity. It is related to the number and magnitude of network weights. Thus, complexity is influenced by the number of hidden units $(N_H)$ and the regularization parameter $(\nu)$. The performance of several network configurations was compared by varying these parameters. A wide range of values was defined for them in order to analyze their effect on generalization ability: $N_H$ was varied from 2 to 50 units while $\nu$ values between 0.01 and 100 were evaluated. For each network configuration, the ICC was computed using leave-one-out cross-validation from data in the training set.

The evolution of ICC is shown in Fig. 1. The performance increased as $\nu$ varied from 0.01 up to 50, which indicates that configurations with small $\nu$ may be affected by overfitting. Setting $\nu$ higher than 50 resulted in lower performance due to an excessive reduction of network complexity. Thus, $\nu$ was set to 50. For this value, we observed that ICC gradually increased as more hidden nodes were added. However, there was no substantial improvement beyond a given value of $N_H$, which approximately corresponds to $N_H = 40$. Therefore, this number of hidden nodes was selected as optimum. Finally, an MLP network with the selected configuration was trained using the complete training set. The scaled conjugate gradient algorithm was used for weight optimization [37].

C. Performance Assessment on the Test Set

MLR and MLP algorithms were assessed on the test set. From ICC analysis, the MLP network (ICC = 0.91) outperformed the MLR model (ICC = 0.80). Fig. 2 depicts actual versus predicted AHI as well as Bland-Altman plots for MLR and MLP models.

Graphs were derived from AHI estimations computed for subjects in the test set. As reflected by the ICC value, graphic representation of the results shows that more accurate AHI estimations were provided by the MLP network. A smaller deviation from the target AHI (dotted line) can be observed for this model. This behavior is also reflected by Bland-Altman analysis. The mean of the differences between actual and predicted AHI is closer to zero for the MLP model. Furthermore, the scatter of the points is substantially higher for the MLR model, as indicated by the value of the endpoints for the 95% confidence interval. Additionally, the ability of these estimators to rank SAHS

<table>
<thead>
<tr>
<th>Feature</th>
<th>Coefficient</th>
<th>Feature</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_S$</td>
<td>-2.39</td>
<td>$\mu_F$</td>
<td>14.41</td>
</tr>
<tr>
<td>$\sigma_S$</td>
<td>-6.74</td>
<td>$\sigma_F$</td>
<td>0.95</td>
</tr>
<tr>
<td>$\gamma_S$</td>
<td>-0.73</td>
<td>$\gamma_F$</td>
<td>3.23</td>
</tr>
<tr>
<td>$\delta_S$</td>
<td>-5.55</td>
<td>$\delta_F$</td>
<td>0.90</td>
</tr>
<tr>
<td>$\text{ApEn}$</td>
<td>-4.54</td>
<td>$S_T$</td>
<td>13.45</td>
</tr>
<tr>
<td>$\text{CTM}$</td>
<td>-11.49</td>
<td>$S_H$</td>
<td>-27.08</td>
</tr>
<tr>
<td>$\text{LIC}$</td>
<td>6.08</td>
<td>$PA$</td>
<td>14.38</td>
</tr>
</tbody>
</table>

Fig. 1. Influence of the number of hidden nodes ($N_H$) and the regularization parameter ($\nu$) on generalization ability.

TABLE II

COEFFICIENTS OF THE MLR MODEL DERIVED FROM THE TRAINING SET
severity was evaluated. The predicted AHI was used to assign each subject to one of the following categories [1]: non-SAHS (0 h$^{-1}$ ≤ AHI < 5 h$^{-1}$), mild-SAHS (5 h$^{-1}$ ≤ AHI < 15 h$^{-1}$), moderate-SAHS (15 h$^{-1}$ ≤ AHI ≤ 30 h$^{-1}$), and severe-SAHS (AHI > 30 h$^{-1}$). The confusion matrices for MLR and MLP models are shown in Table III. The element $(i,j)$ of the matrix represents the number of times that a class $i$ subject was assigned to class $j$ [38]. Both algorithms revealed difficulties in differentiating between non-SAHS and mild-SAHS cases. The MLR algorithm correctly identified one non-SAHS subject more than the MLP network. However, it showed poor diagnostic ability to classify mild-SAHS and moderate-SAHS subjects. The results indicate that the MLP network achieved the highest overall performance.

Both regression algorithms were also assessed in a binary classification context in which non-SAHS and SAHS are the two only possible categories. Table IV summarizes the results obtained using an AHI of 5, 10, and 15 h$^{-1}$ as the decision threshold. The MLP network improved the classification capability of the MLR model for all the evaluated thresholds. The highest accuracy of both algorithms was achieved for a decision threshold of 15 h$^{-1}$, which represents a more conservative definition of SAHS. The MLP network provided a correct decision rate of 93.06% whereas the MLR model achieved 88.89%.

![Fig. 2. Predicted versus actual AHI for (a) the MLR model and (b) the MLP network. Bland-Altman plots obtained for (c) the MLR model and (d) the MLP network. Plots were derived from AHI estimations for subjects in the test set.](image)

**TABLE III**

<table>
<thead>
<tr>
<th>True</th>
<th>No SAHS</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>No SAHS</td>
<td>21</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Mild</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Moderate</td>
<td>1</td>
<td>2</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True</th>
<th>No SAHS</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>No SAHS</td>
<td>20</td>
<td>14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mild</td>
<td>8</td>
<td>18</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Moderate</td>
<td>1</td>
<td>3</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>Severe</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>45</td>
</tr>
</tbody>
</table>

**TABLE IV**

<table>
<thead>
<tr>
<th>AHI (h$^{-1}$)</th>
<th>MLR</th>
<th>MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>90.00</td>
<td>89.58</td>
</tr>
<tr>
<td>10</td>
<td>96.15</td>
<td>91.82</td>
</tr>
<tr>
<td>15</td>
<td>91.82</td>
<td>89.58</td>
</tr>
</tbody>
</table>

Se: sensitivity; Sp: specificity; Acc: accuracy.
V. DISCUSSION

We proposed a regression approach to model SAHS diagnosis. A novel method to estimate the AHI from SaO$_2$ recordings was presented. Time-domain and frequency-domain features were used to reflect the dynamic behavior of these signals. Regression analysis was performed to approximate the functional relationship between the extracted features and the AHI. We assessed the performance of MLR and MLP regression models. The MLP algorithm showed the highest capability to estimate the AHI (ICC = 0.91). It improved the performance achieved from linear analysis using MLR (ICC = 0.80).

The results show that our proposed MLP algorithm provides an accurate assessment of the AHI from SaO$_2$ data. It achieved a sensitivity of 94.87% and a specificity of 90.91% using a decision threshold of 15 h$^{-1}$. The network was able to identify subjects with moderate and severe SAHS with a high level of accuracy. A total of 45 of 48 severe-SAHS subjects and 21 out of 30 moderate-SAHS subjects were correctly diagnosed. As expected, most of the diagnostic errors corresponded to borderline patients. The algorithm assigned 14 non-SAHS subjects to the mild-SAHS group. However, the predicted AHI was smaller than 10 h$^{-1}$ for 10 of them. Labeling SAHS-positive patients as non-SAHS is the most relevant diagnostic error since the lack of appropriate treatment could lead to other health complications. Only one subject with AHI $\geq 15$ h$^{-1}$ (moderate or severe SAHS) was catalogued as non-SAHS by the proposed MLP algorithm. The corresponding AHI was 19.8 h$^{-1}$ and the minimum value of SaO$_2$ during the night was 91.7%, which indicates that apnea events did not tend to be accompanied by marked desaturations. Additionally, eight mild-SAHS patients were assigned to the non-SAHS group. The true AHI for five of these subjects was smaller than 10 h$^{-1}$. Therefore, based on the results of this assessment we conclude that our proposed MLP-based algorithm is a reliable method to assess SAHS severity.

Other methods to estimate the AHI from oximetry data have been suggested in the literature. Vázquez et al. [22] proposed the respiratory disturbance index (RDI), which is based on the detection of desaturation events over 4%. The correlation coefficient between RDI and AHI was 0.97. The utility of this index to classify subjects as non-SAHS or SAHS was evaluated. A sensitivity of 97% and a specificity of 80% were reached using AHI $\geq 10$ h$^{-1}$ to define SAHS. Magalan et al. [12] used several indices from oximetry data to compute the AHI. The oxygen desaturation index over 3% or 4% and the cumulative time spent below different levels of saturation were used as inputs to a multivariate adaptive regression splines (MARS) model. A correlation of 0.84 with the true AHI was achieved. This method provided a sensitivity of 90% and a specificity of 70% using AHI $\geq 15$ h$^{-1}$ to define SAHS. Similarly, Lin et al. [23] reported a correlation coefficient of 0.92 by counting the number of desaturations over 3%. In addition, other signals and data different to SaO$_2$ have been analyzed to approximate the AHI. Roche et al. [24] developed a MLR model from the combination of clinical and oximetry features. It achieved a low correlation value with the AHI derived from PSG (0.38). Its diagnostic accuracy was 62%. De Chazal et al. [9] proposed an algorithm to detect apnea epochs from ECG and SaO$_2$ features. An estimate of the AHI was derived from this epoch-based classification approach. A sensitivity of 95% and a specificity of 83% were reached through the estimated AHI. However, patients with mild SAHS were not considered for testing.

Neural networks have shown to be a powerful tool for regression analysis. As suggested in the present study, other researchers developed neural network-based regression algorithms for AHI estimation. Kirby et al. [8] used 23 clinical variables as inputs to a generalized regression neural network (GRNN), which provided a diagnostic accuracy of 91%. El-Solh et al. [7] used an MLP network with clinical and demographic data to determine the AHI. The sensitivity (95%) was significantly higher than the specificity (65%). Nevertheless, these studies did not assess the ability of their methods to rank SAHS severity. Similarly, other previous studies from our research group were focused on SAHS detection from SaO$_2$ analysis using neural networks. However, a classification approach was proposed. In this context, the output of the algorithm is a categorical variable that indicates the group membership (non-SAHS or SAHS) for the input feature pattern. MLP [14] and RBF [15] classifiers achieved an accuracy of 86% using nonlinear features from SaO$_2$ signals as inputs. Recently, linear classifiers based on discriminant analysis and logistic regression also achieved significant results. They provided an accuracy of 93% [18] and 90% [17], respectively.

Despite the high performance of these classification algorithms, the regression approach results in a more accurate model to characterize SAHS. Two main advantages can be derived from AHI estimation. First, the model provides information about the severity of SAHS. Second, it is insensitive to the criterion used for a positive diagnosis, i.e., the AHI value used to discriminate between SAHS-negative and SAHS-positive. Typically, different criteria between 5 h$^{-1}$ and 15 h$^{-1}$ are used and there is not a globally accepted standard [4].

The MLP algorithm achieved the highest diagnostic capability for the proposed regression problem. However, several limitations can be pointed out. Neural networks are complex models that require an exhaustive design process in comparison with MLR models. On the other hand, the study results indicate that the main source of errors of the MLP algorithm correspond to non-SAHS and mild-SAHS subjects. Therefore, increasing the number of samples in these groups would be desirable in order to obtain a more detailed description from training data. Future studies should also evaluate the effect of feature selection as a technique to remove redundant information and reduce model complexity. In addition, signal preprocessing for motion artifact reduction should be improved to achieve accurate SaO$_2$ measurements, resulting in more reliable AHI estimations. Internal average (low-pass) filtering performed by the pulse oximetry equipment is not capable of complete artifact removal. Furthermore, the influence of the averaging time is a relevant factor to be considered. An excessively high value of this parameter may result in underestimated readings of desaturation events associated with apneas, leading to an incorrect representation of SaO$_2$ dynamics [39]. Another limitation of the study is due to the behavior of SaO$_2$ signals. In some cases, apneas and hypopneas
VI. CONCLUSION

We proposed an algorithm for automatic estimation of AHI from SaO₂ based on feature extraction of time-domain and frequency-domain characteristics, combined with an MLP-based algorithm. The results of our assessment study show that the proposed MLP-based algorithm outperforms equivalent MLR-based algorithms that use the same input features. Our results indicate a high agreement between actual and predicted AHI (ICC = 0.91).

The proposed algorithm only requires nocturnal SaO₂ recordings as the input, eliminating the need for costly, inconvenient, complex, and time-consuming PSG studies for most subjects. Furthermore, the noninvasive nature of SaO₂ makes it possible to create low-cost portable devices designed for home monitoring that can be used for widespread and cost-effective first-line assessment of SAHS severity.

REFERENCES

J. Victor Marcos (S’07) was born in Valladolid, Spain, in 1981. He received the M.S. degree in telecommunication engineering and the Ph.D. degree in information technology and telecommunications, in 2006 and 2011, respectively, both from the University of Valladolid, Valladolid, Spain. He is a member of the Biomedical Engineering Group in the Department of Signal Theory of this university. His current research interests include biomedical signal processing using pattern recognition techniques.

Roberto Hornero (M’04–SM’11) was born in Plasencia, Spain, in 1972. He received the M.S. degree in telecommunication engineering and the Ph.D. degree, in 1995 and 1998, respectively, both from the University of Valladolid, Valladolid, Spain. He is a Professor in the Department of Signal Theory, University of Valladolid, Valladolid, Spain. He founded the Biomedical Engineering Group in 2004. The research interests of this group are connected with the field of nonlinear dynamics, chaotic theory, and wavelet transform with applications in biomedical signal and image processing.

Daniel Álvarez (S’07) was born in Bembibre, Spain, in 1978. He received the M.S. degree in telecommunication engineering and the Ph.D. degree, in 2005 and 2011, respectively, both from the University of Valladolid, Valladolid, Spain. He is currently a Researcher at the Biomedical Engineering Group in the Department of Signal Theory in this university. His current research interests include nonlinear analysis of biomedical signals.

Mateo Aboy (M’98) was born in Pontevedra, Spain, in 1980. He received the B.S. degree (high honors) in electrical engineering and the B.S. degree (high honors) in physics with a minor in mathematics from Portland State University (PSU), Portland, OR, in 2002, the M.S. degree (summa cum laude) in electrical and computer engineering from PSU in 2004, and the Ph.D. degree (summa cum laude) from the University of Vigo, Spain, in 2005.

He has a Bar Admission to practice patent law before the United States Patent and Trademark Office and was a Managing Partner and a Patent Agent at Aboy & Associates PC (www.aboypatentlaw.com). Since 2005, he has been with the Oregon Institute of Technology (OIT), Portland. He is currently a tenure Associate Professor and the Department Chair of the Electrical Engineering Department (EERE). His current research interests include practical applications of signal processing.

Félix del Campo was born in Valladolid, Spain, in 1953. He received the degree in medicine and the M.D. degree from the University of Valladolid, Valladolid, Spain, in 1978 and 1994, respectively.

He is currently with the Hospital Clínico Universitario del Río Hortega, Valladolid, and an Adjunct Professor in the Department of Medicine, University of Valladolid, Spain. His current research interests include respiratory sleep disorders.