Analysis and Classification of Oximetry Recordings to Predict Obstructive Sleep Apnea Severity in Children

Gonzalo C. Gutiérrez-Tobal, Student Member, IEEE, Leila Kheirandish-Gozal, Daniel Álvarez, Member, IEEE, Andrea Crespo, Mona F. Philby, Meelad Mohammadi, Félix del Campo, David Gozal, and Roberto Hornero, Senior Member, IEEE

Abstract—Current study is focused around the potential use of oximetry to determine the obstructive sleep apnea-hypopnea syndrome (OSAHS) severity in children. Single-channel SpO2 recordings from 176 children were divided into three severity groups according to the apnea-hypopnea index (AHI): AHI≤1 events per hour (e/h), 1≤AHI<5 e/h, and AHI ≥5 e/h. Spectral analysis was conducted to define and characterize a frequency band of interest in SpO2. Then we combined the spectral data with the 3% oxygen desaturation index (ODI3) by means of a multi-layer perceptron (MLP) neural network, in order to classify children into one of the three OSAHS severity groups. Following our MLP multiclass approach, a diagnostic protocol with capability to reduce the need of polysonomography tests by 46% could be derived. Moreover, our proposal can be also evaluated, in a binary classification task for two common AHI diagnostic cutoffs (AHI = 1 e/h and AHI= 5 e/h). High diagnostic ability was reached in both cases (84.7% and 85.8% accuracy, respectively) outperforming the clinical variable ODI3 as well as other measures reported in recent studies. These results suggest that the information contained in SpO2 could be helpful in pediatric OSAHS severity detection.

I. INTRODUCTION

Pediatric obstructive sleep apnea-hypopnea syndrome (OSAHS) has emerged as a frequent and concerning medical condition in the past 2-3 decades. It is characterized by an abnormal breathing pattern during sleep that includes the recurrence of apneas (complete airflow cessation) and hypopneas (airflow limitation), caused by total or partial upper airway obstruction, respectively [1]. Inadequate gas exchange characterized by repetitive hypoxia, hypercapnia, and accompanied by arousal episodes during the night has been suggested as the cause for serious comorbidities related to central nervous system and cardiovascular and metabolic system [2]. Consequently, several daytime symptoms related to OSAHS, such as cognitive and behavioral irregularities as well as atypical growth are frequently present and reported by parents [3, 4]. Furthermore, the prevalence of OSAHS in children is high, with studies reporting up to 5.7% among general pediatric population [4].

The “gold standard” approach to diagnose OSAHS in children is overnight polysomnography (PSG) [4]. However, PSG has several limitations since it is both complex and costly due to the high number of physiological signals that need to be recorded [5]. Additionally, all these signals need offline inspection in order to derive the apnea-hypopnea index (AHI), which is used to establish whether OSAHS is present and its severity. Consequently, PSG is also time-consuming [6]. Moreover, children often do not tolerate well the equipment involved in PSG [7].

These limitations have led to the search for simpler diagnostic alternatives. Thus, one common approach has been the study of the diagnostic ability of reduced sets of signals derived from those involved in PSG, such as electrocardiography (ECG), photoplethysmography (PPG), airflow (AF), or blood oxygen saturation (SpO2) [8-11]. Particularly, frequency and time domain analyses of ECG-derived signals showed utility in pediatric OSAHS diagnosis [8]. Moreover, the analysis of pulse transit time variability from PPG was successfully used to classify time segments into apneic or non-apneic, as well as children into normal subjects and OSAHS patients [9]. Additionally, a recent study from our research group reported high diagnostic ability when combining the oxygen desaturation index (ODI) from SpO2 with spectral information from AF [10]. Finally, spectral, nonlinear, and statistical features from SpO2 and pulse rate variability (PRV) recordings were obtained and successfully combined to establish OSAHS in children [11].

In this study, however, we focus on the use of the information contained in a single-channel SpO2 to help in OSAHS severity detection. Our hypothesis is that exclusive utilization of data from the SpO2 channel could simplify the OSAHS diagnosis and severity assessment in children. Hence, the main objective of this preliminary work is to evaluate the diagnostic ability of the information contained in the SpO2 signal. Specifically, we analyze the spectrum of SpO2 recordings from children, divided into three groups according to their corresponding AHI. There exists a lack of consistency in the literature as to the optimal AHI cutoff to determine OSAHS in children, with most of the studies applying 1, 3, or 5 events per hour (e/h) [4]. We chose AHI < 1 e/h as the most restrictive cutoff to discard OSAHS and AHI ≥ 5 e/h to define a group with the highest OSAHS severity. Additionally, we formed another group with those patients in the range 1 ≤ AHI < 5 e/h, which is recognized as the most challenging concerning the decision to implement treatment, usually consisting of surgical removal of tonsils.

*Research supported by the project VA059U13 (Consejería de Educación de la Junta de Castilla y León). G. C. Gutiérrez-Tobal was in receipt of a PIRTU grant (Consejería de Educación de la Junta de Castilla y León and the European Social Fund).

G. C. Gutiérrez-Tobal, D. Álvarez, and R. Hornero, are with the Biomedical Engineering Group (University of Valladolid, Spain) (e-mail: Gonzalo.gutierrez@gie.tel.uva.es)

L. Kheirandish-Gozal, M. F. Philby, M. Mohammadi, and D. Gozal are with the Section of Sleep Medicine, Department of Pediatrics, Pritzker School of Medicine, Biological Sciences Division (The University of Chicago, Chicago, IL, USA) (e-mail: dgozal@peds.bsd.uchicago.edu).

A. Crespo and F. del Campo are with the sleep unit of Hospital Universitario Rio Hortega (Valladolid, Spain) (e-mail: fsas@telefonica.net).
and adenoids [2]. Therefore, we here evaluate the spectrum of the SpO2 recordings from children in the three groups looking for discriminative features. Additionally, we use 3% ODI (ODI1) for comparison purposes. Finally, we combine the spectral information and ODI1 by means of an artificial neural network, a multi-layer percepton (MLP) [12], in order to classify children into one of the three groups. This multiclass approach lets us define a protocol which includes doubtful subjects, as well as allows the evaluation, at the same time, of both AHI=1 e/h and AHI=5 e/h cutoffs from a binary classification point of view.

II. SUBJECTS AND SIGNALS UNDER STUDY

The study involved SpO2 recordings from 176 children (97 males and 79 females). All of them were clinically referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children’s Hospital (Chicago, IL, USA) due to suspicion of OSAHS. The Ethical Committee approved the protocol and an informed consent to participate in the study was obtained for each child. Overnight PSGs were conducted from 20.00 to 08.00. Recordings were acquired by means of a digital polysomnography system (Polysonith; Nihon Kohden America Inc., CA, USA). Detection and quantification of sleep and cardiorespiratory events were carried out according to the rules of the American Academy of Sleep Medicine [13]. Thus, apnea was defined as the absence of oronasal airflow during at least 2 respiratory cycles. Accordingly, hypopnea was defined as a decrease ≥30% in the nasal pressure airflow signal lasting at least 2 respiratory cycles, leading to a desaturation ≥3% and/or an arousal [13]. As previously stated, children were divided into three groups according to their corresponding AHI: AHI under 1 e/h (AHI<1), AHI in the range [1, 5) e/h (AHI[1,5]), and AHI equal or above 5 e/h (AHI≥5). Table I summarizes demographic and clinical data from subjects according to this division. No statistical significant differences (p-value<0.01) were found in age, gender, and body mass index (BMI) when applying the non-parametric Kruskal-Wallis test to compare the three groups.

The SpO2 recordings were acquired during PSG at a sampling rate of fs = 25 Hz. Artifacts due to children movements were automatically removed during preprocessing. Thus, SpO2 values equal to zero as well as differences between consecutive SpO2 samples ≥4% were considered artifacts [14]. Removed samples were substituted by interpolated data. ODI3 was estimated as the number of desaturations (at least 3%) per hour of sleep time.

III. METHODOLOGY

The methodology was divided into three steps. First, a spectral analysis of the SpO2 recordings was conducted to look for differences among the three groups. Then we extracted several spectral features according to this analysis. Finally, we combined the spectral data and ODI1 through MLP to classify the children into one of the three classes.

A. Spectral analysis and feature extraction

Power spectral density (PSD) was estimated for each SpO2 recording by means of the Welch’s method [15]. We used a Hamming window of 2^15 samples (5.5 minutes), 50% overlap, and a discrete Fourier transform of 2^16 samples. Fig. 1 shows the median PSD for each group of OSAHS severity. Higher PSDs can be observed as the severity increases. A band of interest (BW) is also shown in the range 0.0137-0.0473 Hz. This corresponds to the spectral bandwidth in which the three groups showed statistical significant differences (Mann-Whitney U test) in their PSD amplitude values (p-value = 0.01, p-value = 0.0033 after Bonferroni correction). In this case, BW is equivalent to the bandwidth in which AHI<1 and AHI[1,5] showed significant differences. Fig. 2 displays the p-value vs. frequency plots for each of the three possible comparisons. The limits of BW are easily located as the crosspoints between the AHI<1 vs. AHI[1,5] p-value (f) curve and the p-value significance level line (in red).

We extracted the following features from the BW of each PSD: maximum PSD value (MA), minimum PSD value (mA), spectral power (PS), as the area under the PSD at BW), and standard deviation of the PSD values (SD). According to Fig. 1, higher values were expected in these features as the OSAHS severity increases. After feature extraction, each subject under study is characterized by a vector x, (i = 1, 2…

<table>
<thead>
<tr>
<th>TABLE I. DEMOGRAPHIC AND CLINICAL DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td># Subjects</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Male (%)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
</tr>
<tr>
<td>AHI (e/h)</td>
</tr>
</tbody>
</table>

BMI: Body Mass Index; AHI: Apnea Hypopnea Index; *p-value=0.016; **p-value=0.816
posed of $M$, $M = 176$) whose 5 components are the corresponding values of the four spectral features and ODI$_3$.

**B. Multi-layer perceptron**

MLP is a supervised learning algorithm whose architecture is arranged in several interconnected layers (input, hidden, and output). These are composed of units known as neurons or perceptrons [12]. Each neuron is characterized by an activation function $g(\cdot)$ and their connections to neurons from other layers ($w_{ij}$). In our case, the input layer has five units, corresponding to the number of spectral features obtained for each subject ($MA, mA, P_S, SD$) and ODI$_3$. Moreover, since the purpose is to carry out a three-class classification, three output units with a logistic activation function were used. We also implemented a single hidden layer, composed of neurons with non-linear activation functions. This configuration is known to be able to provide a universal function approximation [12]. Since the number of neurons in the hidden layer ($N_{H}$) controls the effective complexity of the network [12], we chose a small number, $N_{H} = 5$, to prevent network from overfitting. Thus, the final input-hidden-output layer architecture was 5:5:3 neurons. The weights $w_{ij}$ were optimized using the sum of squares error function minimization criterion by means of the scaled conjugate gradient algorithm [12]. For each subject under study, the final classification task was performed by assigning the corresponding $x_i$ ($i = 1, 2…, M, M = 176$) to the class with the highest probability in the output layer.

**C. Statistical analysis**

The non-parametric Kruskal-Wallis test was used to assess statistical differences in the spectral features from the OSAHS severity groups. A confusion matrix was used to evaluate the performance of multiclass MLP. Also, to assess the output of MLP from a binary classification point of view; sensitivity (Se, percentage of OSAHS-positive subjects rightly classified), specificity (Sp, percentage of OSAHS-negative subjects rightly classified), accuracy (Acc, overall percentage of subjects rightly classified), positive predictive value (PPV, proportion of positive test results which are true positives), negative predictive value (NPV, proportion of negative test results which are true negatives), positive likelihood ratio (LR+, Se/(1-Sp)), and negative likelihood ratio (LR-, (1-Se)/Sp) measured the diagnostic ability for both AHI=1 e/h and AHI=5 e/h cutoffs. All these statistics were obtained after leave-one-out cross-validation (loo-cv).

### IV. RESULTS

Table II displays the values of the spectral features and ODI$_3$ for each of the three OSAHS severity groups (mean ± standard deviation). All of them showed large statistical significant differences when comparing the three groups by means of Kruskal-Wallis test. As expected, the five features are higher as the OSAHS severity increases.

Table III shows the confusion matrix resulting from the diagnostic ability assessment of the MLP network for the three-class classification task (results after loo-cv). A total of 125 out of 176 subjects were rightly classified in their actual class (71.0%). Per classes, 50.0% (15 out of 30) of the subjects in AHI$_{1,5}$, 80.0% (60 out of 75) in AHI$_{1,5}$, and 70.4% (50 out of 71) in AHI$_{1,5}$ were rightly classified.

### Table II. Values of the Spectral Features and ODI$_3$ (Mean±Standard Deviation)

<table>
<thead>
<tr>
<th>Features</th>
<th>AHI$_{1}$</th>
<th>AHI$_{1,5}$</th>
<th>AHI$_{1,5}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODI$_3$ (e/h)</td>
<td>1.01±1.10</td>
<td>3.21±2.80</td>
<td>16.18±20.78</td>
<td>&lt;10$^{-11}$</td>
</tr>
<tr>
<td>MA (W/Hz)</td>
<td>7.76±7.41</td>
<td>13.58±16.98</td>
<td>71.61±149.83</td>
<td>&lt;10$^{-12}$</td>
</tr>
<tr>
<td>mA (W/Hz)</td>
<td>1.44±0.59</td>
<td>2.91±5.23</td>
<td>19.51±70.43</td>
<td>&lt;10$^{-11}$</td>
</tr>
<tr>
<td>$P_S$ (W/Hz)</td>
<td>7.80±4.32</td>
<td>15.15±21.02</td>
<td>97.74±248.19</td>
<td>&lt;10$^{-13}$</td>
</tr>
<tr>
<td>SD$_3$ (W/Hz)</td>
<td>1.84±1.81</td>
<td>3.23±4.12</td>
<td>16.31±27.29</td>
<td>&lt;10$^{-11}$</td>
</tr>
</tbody>
</table>

### Table III. Confusion Matrix for the MLP Multiclass Task (After Loo-CV)

<table>
<thead>
<tr>
<th>actual</th>
<th>estimated AHI$_{1}$</th>
<th>AHI$_{1,5}$</th>
<th>AHI$_{1,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHI$_{1}$</td>
<td>15</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>AHI$_{1,5}$</td>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>AHI$_{1,5}$</td>
<td>1</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

### Table IV. Diagnostic Ability of MLP and ODI$_3$ (Binary Classification after Loo-CV)

<table>
<thead>
<tr>
<th>Se (%)</th>
<th>Sp (%)</th>
<th>Acc (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODI$_3$ (AHI=1)</td>
<td>78.1</td>
<td>80.0</td>
<td>78.4</td>
<td>95.0</td>
<td>48.9</td>
<td>3.91</td>
</tr>
<tr>
<td>ODI$_3$ (AHI=5)</td>
<td>69.0</td>
<td>81.9</td>
<td>76.7</td>
<td>79.6</td>
<td>72.1</td>
<td>3.81</td>
</tr>
<tr>
<td>MLP (AHI=1)</td>
<td>91.8</td>
<td>50.0</td>
<td>84.7</td>
<td>89.9</td>
<td>55.6</td>
<td>1.84</td>
</tr>
<tr>
<td>MLP (AHI=5)</td>
<td>70.4</td>
<td>96.2</td>
<td>85.8</td>
<td>92.6</td>
<td>82.8</td>
<td>18.5</td>
</tr>
</tbody>
</table>

Table IV shows the diagnostic ability of MLP and ODI$_3$ when assessing both the AHI = 1 e/h and AHI= 5 e/h cutoffs (results after loo-cv). MLP results are directly derived from the confusion matrix. For both cutoffs the global Acc of MLP is higher than the corresponding ODI$_3$ (84.7% vs. 78.4% and 85.8% vs. 76.7%, respectively). In the case of AHI=1 e/h, ODI$_3$ is much more specific than MLP, leading to higher PPV and LR+. In the case of AHI=5 e/h, however, MLP outperforms ODI$_3$ at each statistic.

### V. DISCUSSION AND CONCLUSIONS

In this work, we have developed an automatic diagnostic methodology for pediatric OSAHS severity based on the information contained in single-channel SpO$_2$. Features from a spectral band of interest and the clinical variable ODI$_3$ were combined by means of MLP to classify subjects into one out of the three OSAHS severity levels.

The spectral analysis of the SpO$_2$ signal revealed a band of interest (BW=0.0137-0.0473 Hz) in which statistically significant differences were found for the three classes. The lower limit of BW is consistent with the corresponding band of interest in adults (0.014-0.033 Hz, i.e., events lasting from 30 to 71 s) [16]. Conversely, a higher upper limit was found in children, suggesting shorter events as also significant for them. This agrees with the higher respiratory rate reported in children [17]. However, further analysis is required regarding the causes of the differences in both bands.

The spectral features extracted from BW showed statistically significant differences when comparing the three
classes. All of them reached higher values as the OSAHS severity increases. Since the ideal SpO\textsubscript{2} time series is a constant, close to 100\%, the higher PSD values in the frequencies correspond to more desaturations and recoveries to the baseline. Consequently, higher MA, m\textalpha, P\textbeta, and SD\textgamma suggest more desaturation events both in discrete frequencies (M\textalpha, m\textalpha) and in the whole band (P\textbeta, SD\textgamma), which is consistent with the clinically used severity classification of OSAHS.

Our multiclass MLP proposal correctly classified 71\% of the subjects. Although this overall accuracy is not high enough, a deeper study of the subjects wrongly classified reveals that the 11 children who belong to AHI\textsubscript{1,5}l, and were assigned to AHI\textsubscript{1,5}, present an AHI of 1.65 ± 0.42 e/h. This means that 96.3\% of subjects predicted as AHI\textsubscript{1,5}l have no OSAHS or a low severity degree. Additionally, the 4 children from AHI\textsubscript{1,5}l assigned to AHI\textsubscript{1,5} present an AHI of 3.0 ± 1.7 e/h, i.e., 100\% of children predicted as AHI\textsubscript{1,5}l have severe OSAHS or a higher severity degree comparing with the mean of the AHI\textsubscript{1,5}l class. Finally, children assigned to AHI\textsubscript{1,5}l come from the three classes: AHI\textsubscript{1,5}l (15.8\%), AHI\textsubscript{1,5}l (63.2\%), and AHI\textsubscript{1,5}l (21.0\%). Consequently, subjects assigned to this class should be regarded as inconclusive. A screening protocol could be generated from these results as follows: i) if MLP predicts AHI\textsubscript{1,5}l, discard OSAHS; ii) if MLP predicts AHI\textsubscript{1,5}l, consider treatment; iii) if MLP predicts AHI\textsubscript{1,5}l, send to overnight PSG. Since the SpO\textsubscript{2} signal is easily acquired from an oximeter, such a protocol would reduce the need by 46\% (81/176) of overnight PSGs.

Other studies analyzed physiological signals to help in pediatric OSAHS diagnosis. All of them reported results from a binary classification point of view. Shouldice et al. analyzed 50 ECG recordings, reaching 85.7\% Se, 81.8\% Sp, and 84.0\% Acc using a quadratic linear discriminant applied to 23 features (AHI cutoff = 1 e/h) [8]. Gil et al. assessed the diagnostic ability of information contained in 21 PPG time series, reporting 75.0 \% Se, 85.7\% Sp, and 80.0\% Acc (AHI cutoff = 5 e/h) [9]. Gutiérrez-Tobal et al. combined spectral features from 50 AF recordings with ODI\textsubscript{1} from SpO\textsubscript{2} to achieve 85.9\% Se, 87.4\% Sp, and 86.3\% Acc with a logistic regression methodology (AHI cutoff = 3 e/h) [10]. Finally, Garde et al. reported 83.6 \% Se, 88.4\% Sp, and 85.0\% Acc in a 146 subject database by combining 8 features from SpO\textsubscript{2} and PRV in a linear discriminant [11]. Our MLP methodology can be assessed for AHI =1 e/h and AHI = 5 e/h at the same time (84.7\% and 85.8\% Acc, respectively). Only Gutiérrez-Tobal et al. reported higher Acc. However, they used two channels and the AHI cutoff was 3 e/h.

This work has some limitations that are worthy of mention. First, although the number of subjects is not small when comparing to other similar studies, more children, particularly those with AHI<1 e/h, would be necessary for the sake of a proper MLP training. This would include a training-test strategy as well as the evaluation of a range of neurons in the hidden layer, which was arbitrary set to a low value in order to decrease the chances for overfitting.

Additionally, more subjects would also let us use a training set from which we could independently obtain the spectral band of interest. However, a loo-cv methodology was used to validate our results. Finally, the use of features from time domain could complement our findings. One future goal is the assessment of features and classification models other than those presented in this work using a larger dataset.

In summary, we have developed a multiclass MLP methodology with capability to help in pediatric OSAHS severity screening. The SpO\textsubscript{2} features obtained from a frequency band of interest, combined with ODI\textsubscript{1} through MLP, outperform the single diagnostic yield of this clinical variable. Our proposal can be also evaluated for binary classification purposes, reaching high diagnostic ability comparing with recent state-of-the-art studies. Thus, our results suggest that the information contained in single-channel SpO\textsubscript{2} is helpful to detect severity categories among children with OSAHS that are worthy of mention.

**REFERENCES**