Automated Analysis of Nocturnal Oximetry as Screening Tool for Childhood Obstructive Sleep Apnea-Hypopnea Syndrome

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

Abstract—Childhood obstructive sleep apnea-hypopnea syndrome (OSAHS) is a highly prevalent condition that negatively affects health, performance and quality of life of infants and young children. Early detection and treatment improves neuropsychological and cognitive deficits linked with the disease. The aim of this study was to assess the performance of automated analysis of blood oxygen saturation (SpO₂) recordings as a screening tool for OSAHS. As an initial step, statistical, spectral and nonlinear features were estimated to compose an initial feature set. Then, fast correlation-based filter (FCBF) was applied to search for the optimum subset. Finally, the discrimination power (OSAHS negative vs. OSAHS positive) of three pattern recognition algorithms was assessed: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and logistic regression (LR). Three clinical cut-off points commonly used in the literature for positive diagnosis of the disease were applied: apnea-hypopnea index (AHI) of 1, 3 and 5 events per hour (e/h). Our methodology reached 88.6% accuracy (71.4% sensitivity and 100.0% specificity, 100.0% positive predictive value, and 84.0% negative predictive value) in an independent test set using QDA for a clinical cut-off point of 5 e/h. These results suggest that SpO₂ nocturnal recordings may be used to develop a reliable and efficient screening tool for childhood OSAHS.

I. INTRODUCTION

Childhood obstructive sleep apnea-hypopnea syndrome (OSAHS) is a highly prevalent but under-diagnosed condition [1]. It is characterized by prolonged partial (hypopnea) and/or intermittent complete (apnea) upper airway obstruction during sleep, which lead to gas exchange abnormalities and sleep disruption [2]. According to the American Academy of Pediatrics, OSAHS affects 1% to 5% of children in the general pediatric population [2]. Untreated OSAHS has been associated with negative consequences in the development and performance of infants and young children, reducing overall health and quality of life [1, 3], while increasing healthcare use and associated costs [4].

Overnight laboratory polysomnography (PSG) is the gold standard test for OSAHS in children [2, 5]. Nevertheless, PSG is costly due to the need for a specialized sleep laboratory, expensive equipment and trained personnel [1, 6]. Furthermore, complete PSG is highly intrusive for children, which limits its effectiveness [7]. Thus, there is an increasing demand for novel screening tools in the context of childhood OSAHS. Main methodologies include sleep-related questionnaires, respiratory polygraphy (RP), nap PSG, and ambulatory PSG [1, 2, 8]. Despite recent improvements [1], questionnaires are usually not sensitive as single screening tools. On the other hand, RP and PSG-based approaches do not overcome the main limitations of standard PSG. Therefore, further research is still needed. An interesting approach is the analysis of single-channel sleep-related recordings, which reduces cost and complexity. In this regard, automated processing of oximetry signals is a promising alternative due to its reliability, simplicity, and suitability for children [5, 6, 9-12].

Previous oximetry-based studies in the context of OSAHS diagnosis assessed conventional indices [5, 6, 9-11], common statistics [6, 10] and conventional spectral features [10]. Similarly, the studies by Garde et al. [10] and Sahadan et al. [12] also used the information contained in pulse rate recordings from pulse oximetry. In the present research, blood oxygen saturation (SpO₂) recordings were analyzed. Statistical (first-to-fourth moments), spectral (amplitude, relative power and power distribution measures), nonlinear (irregularity, variability, and complexity measures), and conventional indices (number of desaturations from baseline) were computed. These metrics have been previously assessed in the context of OSAHS diagnosis both in adults [13, 14] and children [6]. Fast correlation-based filter (FCBF) is proposed for feature selection. FCBF is a variable ranking methodology for feature selection independent of the classifier subsequently used in the classification stage [15]. Linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR) are proposed for classification [16]. QDA and LR are suitable alternatives to conventional LDA in binary classification problems but their performances have been weakly assessed in the context of childhood OSAHS. We hypothesized that the proposed methodology could detect complementary variables and provide general classification models useful as screening tools for OSAHS in children. The aim of this study was to design and assess several binary classifiers using different clinical cut-offs for OSAHS in order to analyze the screening ability at different severity thresholds. To achieve this goal, independent training and test datasets were analyzed.
II. SUBJECTS AND SIGNALS UNDER STUDY

A total of 176 children (97 boys and 79 girls) composed our dataset. All children were referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children’s Hospital (Chicago, IL, USA) due to clinical suspicion of suffering from OSAHS. Informed consents to participate in the study were obtained and the Institution’s Ethical Review Committee approved the protocol.

Children’s sleep was monitored using a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA). SpO2 recordings from PSG (sampling frequency 25 Hz) were exported and processed offline. Artifacts were automatically removed by means of a preprocessing stage. SpO2 values equal to zero and differences between consecutive SpO2 samples were considered artifacts [17].

The American Academy of Sleep Medicine rules were used to quantify sleep and cardiorespiratory events and derive the apnea hypopnea index (AHI), which averages the number of events per hour of sleep. Apnea was defined as the absence of oronasal airflow during at least 2 respiratory cycles. Hypopnea was defined as a decrease ≥50% lasting at least 2 respiratory cycles, leading to a desaturation ≥3% and/or an arousal. In the present study, the AHI-based clinical threshold was varied in order to assess the performance of the proposed methodology as a screening tool for OSAHS using commonly used cut-off points. AHI ≥ 1, 3, and 5 events per hour (e/h) from PSG were considered as OSAHS-positive [2]. Table I displays demographic and clinical features of the dataset taking into account the proposed AHI cut-off thresholds for the disease. For each cut-off point, the whole population was randomly divided into independent training (60%) and test (40%) sets.

III. METHODOLOGY

Firstly, each SpO2 recording was parameterized computing 17 features: time domain statistics (4), frequency domain statistics (6), conventional spectral features (3), nonlinear measures (3), and conventional oximetric indices (1). Then, a feature selection stage was applied using FCFB. An optimum feature subset was derived for each OSAHS cut-off. Finally, LDA, QDA, and LR models were composed for each feature subset. The training set was used for feature selection and model optimization whereas the test set was used for assessing all classifiers in an independent dataset.

A. Feature extraction stage

The following feature subsets were computed:

- Time domain statistics [16]. Mean (M1t), variance (M2t), skewness (M3t), and kurtosis (M4t) were derived from the data histogram of SpO2 amplitudes.
- Frequency domain statistics [13, 14]. The distribution of power spectral density (PSD) amplitudes was parameterized by means of first-to-fourth statistical moments (M1f-M4f). In addition, the median frequency (MF) and spectral entropy (SE) were computed to quantify the degree of flatness of the power distribution.
- Conventional spectral features [13, 14]. Total signal power (Pr) and the peak amplitude (PA) and relative power (Pr) in the apnea frequency band (0.021 – 0.040 Hz) were computed from the PSD [18].
- Nonlinear measures [13]. Sample entropy (SampEn, m=1, r=0.25), central tendency measure (CTM, r=1) and Lempel-Ziv complexity (LZC) were computed to quantify irregularity, variability and complexity.
- Conventional oximetric indices [17]. Number of desaturations greater than or equal to 3% from baseline per hour of recording (ODI3). SpO2 recordings were segmented into 1-min length epochs before computing the time domain features (each feature average value was subsequently obtained) whereas the PSD average value was estimated using the Welch’s method (15000-sample Hanning window, 50% overlap and 214-points DFT).

B. Feature selection stage

FCFB computes the symmetric uncertainty (SU) to select relevant and non-redundant variables [15], SU between the i-th input feature (Xi) and the AHI (Y) is defined as follows:

\[ SU_i(X_i, Y) = 2 \frac{IG(X_i, Y)}{H(X_i) + H(Y)}, i = 1, \ldots, p \]  

where IG is the information gain and H is the well-known Shannon’s entropy. In the first step, FCFB ranks features according to their relevance (the higher SUi, the more relevant feature). Then, a threshold is used to discard irrelevant features. In this study, the log criterion was applied, where the cut-off is the SU value of the \([\log(N)]\)-th ranked feature. In the second step, redundant features are removed. In order to perform the redundancy analysis, SUi(feature, feature) between each pair of remaining ranked features (so that SUi ≥ SUj) is computed. Then, feature j is removed if SUi ≥ SUj due to redundancy.

C. Feature classification stage

Conventional statistical pattern recognition techniques were used for binary classification.
• Linear discriminant analysis (LDA) [16]. Statistical classification algorithms based on discriminant analysis assume normality to model each class-conditional density function \( p(x | c_j) \) for input pattern \( x \) and class \( c_j \). If homoscedasticity is also presumed, i.e. all the class covariance matrices are equal \((\Sigma_j = \Sigma)\), then the classification rule is called LDA and a linear decision threshold is assumed. Equation (2) shows the classification rule,

\[
y_j(x) = \mu_j^{\top} \Sigma_j^{-1} x - \frac{1}{2} \mu_j^{\top} \Sigma_j^{-1} \mu_j + \ln\left( P(c_j) \right),
\]

where \( \mu_j \) and \( \Sigma \) are the class \( c_j \) mean vector and covariance matrix, respectively.

• Quadratic discriminant analysis (QDA) [16]. In a more general context where it is not possible to presume homoscedasticity, the Bayes classification rule that minimizes the classification error function establishes a quadratic decision boundary between classes in the feature space. Equation (3) shows the classification rule under these assumptions,

\[
y_j(x) = -\frac{1}{2} (x - \mu_j)^{\top} \Sigma_j^{-1} (x - \mu_j) - \frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} \ln(2\pi) + \ln(\mathcal{L}(c_j)).
\]

• Logistic regression (LR) [16]. No a priori normality and homoscedasticity of data are presumed. A binary LR classifier models the probability density function as a Bernoulli distribution. The maximum likelihood criterion is used to optimize the coefficients of the logistic model. Equation (4) shows the logistic classification function:

\[
y(x, \beta) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_1 - \ldots - \beta_p x_p}},
\]

where \( \beta \) is the vector of coefficients of the LR model.

**TABLE III.** Diagonsitic assessment of each single variable from the initial feature space in the test set

<table>
<thead>
<tr>
<th>Performance (%)</th>
<th>cut-off AHI = 1 e/h</th>
<th>cut-off AHI = 3 e/h</th>
<th>cut-off AHI = 5 e/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se</td>
<td>Sp</td>
<td>PPV</td>
<td>NPV</td>
</tr>
<tr>
<td><strong>M1t</strong></td>
<td>51.3</td>
<td>57.5</td>
<td>59.8</td>
</tr>
<tr>
<td><strong>M2f</strong></td>
<td>71.8</td>
<td>83.3</td>
<td>79.8</td>
</tr>
<tr>
<td><strong>M3t</strong></td>
<td>78.7</td>
<td>81.3</td>
<td>77.5</td>
</tr>
<tr>
<td><strong>M4t</strong></td>
<td>71.4</td>
<td>82.9</td>
<td>77.1</td>
</tr>
<tr>
<td><strong>M1f</strong></td>
<td>89.3</td>
<td>86.1</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>M2f</strong></td>
<td>89.3</td>
<td>86.1</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>M3f</strong></td>
<td>89.3</td>
<td>86.1</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>M4f</strong></td>
<td>89.3</td>
<td>86.1</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
<tr>
<td><strong>PA</strong></td>
<td>60.6</td>
<td>65.0</td>
<td>60.3</td>
</tr>
</tbody>
</table>

Se: sensitivity (%); Sp: specificity (%); PPV: positive predictive value (%); NPV: negative predictive value (%); Acc: accuracy (%)

Features with the highest accuracy for each OSAHS cut-off are highlighted in bold.
performance in terms of accuracy was achieved using a cut-off for OSAHS equal to 5 e/h, where QDA reached 88.6% accuracy (71.4% Se, 100.0% Sp).

V. DISCUSSION AND CONCLUSIONS

Feature extraction, selection, and classification algorithms were assessed in the context of screening for pediatric OSAHS using SpO2 recordings obtained during overnight polysomnographic evaluations in a clinical setting. All feature extraction approaches (time, frequency, linear, and nonlinear) were present in all optimum feature subsets from FCBF, suggesting the complementarity of the proposed methods. Our results suggest that M2t, PA, SampEn, and ODI3 are relevant for the disease because they were always selected. Similarly, M2t, PA, and ODI3 achieved the highest individual performance using the cut-off points 5, 3, and 1 e/h, respectively. Optimum pattern recognition models improved individual features for a cut-off AHI = 5 e/h. The highest performance was reached by QDA, which achieved 71.4% Se, 100.0% Sp and 88.6% Acc in the test set. It is important to point out that using this model there are no false negatives: if children test positive, then they definitely have OSAHS (positive post-test probability of 100%).

Our results agree with recent studies focused on screening methods for OSAHS in children. The study by Sahadan et al. analyzed a population of 93 children and achieved 18% Se and 97% Sp (cut-off AHI=1 e/h) using pulse rate conventional measures from pulse oximetry recordings [12]. Similarly, the study by Garde et al. used SpO2 and pulse rate from a population composed of 146 children. The proposed LDA model achieved 88.4% Se and 83.6% Sp (cut-off AHI=5 e/h) in a test set [6]. In [1], Kadmon et al. assessed a simplified sleep-related questionnaire for screening OSAHS in a population of 85 children. Their method achieved 83% Se and 64% Sp (cut-off AHI=5 e/h).

We should take into account some limitations. The population cohort evaluated herein should be expanded in order to derive more generalizable conclusions. In addition, input parameters of spectral and nonlinear analyses should be thoroughly optimized. Finally, additional feature selection and classification methods should be assessed.

In summary, our results suggest that automated analysis of overnight SpO2 using suitable features and statistical pattern recognition models could improve the performance of oximetry as a screening tool for OSAHS in children.

REFERENCES


---

### Table IV. Diagnostic Performance in the Test Set of Each Optimum Oximetric Model from FCBF Using Different Cut-off Points

<table>
<thead>
<tr>
<th>Performance (%)</th>
<th>cut-off AHI = 1 e/h</th>
<th>cut-off AHI = 3 e/h</th>
<th>cut-off AHI = 5 e/h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Log criterion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>53.5</td>
<td>5.93</td>
<td>53.9</td>
</tr>
<tr>
<td>QDA</td>
<td>46.6</td>
<td>91.7</td>
<td>62.7</td>
</tr>
<tr>
<td>LR</td>
<td>91.4</td>
<td>83.8</td>
<td>77.1</td>
</tr>
</tbody>
</table>

Se: sensitivity (%); Sp: specificity (%); PPV: positive predictive value (%); NPV: negative predictive value (%); Acc: accuracy (%)

Features with the highest accuracy for each OSAHS cut-off are highlighted in bold.