Improving the Diagnostic Ability of Oximetry Recordings in Pediatric Sleep Apnea-Hypopnea Syndrome by Means of Multi-Class AdaBoost

Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, *Member, IEEE*, Verónica Barroso-García, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, *Senior Member, IEEE*

***Abstract*— Pediatric sleep apnea-hypopnea syndrome (SAHS) is a highly prevalent respiratory disorder that may impose many negative effects on the health and development of children. Due to the drawbacks of overnight polysomnography (PSG), the gold standard diagnosis technique, automated analysis of nocturnal oximetry has emerged as a simplified alternative. In order to improve diagnosis ability of oximetry, we propose to evaluate the usefulness of AdaBoost, a classification boosting algorithm, in the context of pediatric SAHS. A database composed of 981 SpO_2 recordings from pediatric subjects was used. For this purpose, a signal processing approach divided into two main stages was conducted: (i) feature extraction, where 3% oxygen desaturation index (ODI3), spectral, and nonlinear features were computed from the oximetry signal, and (ii) AdaBoost classification, where an AdaBoost.M2 model was trained with these features in order to determine the severity of pediatric SAHS according to the apnea-hypopnea index (AHI): AHI<1 events per hour (e/h), 1≤AHI<5 e/h, and AHI≥5 e/h. Our AdaBoost.M2 model achieved a Cohen's kappa of 0.474 in an independent test set in the 3-class classification task. In addition, high accuracies were obtained when using the AHI cutoffs for diagnosis of mild (AHI=1 e/h) and moderate-to-severe (AHI=5 e/h) SAHS: 80.9% and 82.9%, respectively. These results achieved slightly higher diagnostic accuracies than ODI3 as well as state-of-the-art studies. Therefore, AdaBoost could help to enhance the diagnostic ability of the oximetry signal to assess pediatric SAHS severity.**

This work was supported 'Ministerio de Economía y Competitividad (MINECO)' and 'European Regional Development Fund' under projects DPI2017-84280-R, TEC2014-53196-R and RTC-2015-3446-1, by 'Consejería de Educación de la Junta de Castilla y León and FEDER' under project VA037U16, by ' European Commission ' and ' European Regional Development Fund ' under project 'POCTEP 0378_AD_EEGWA_2_P', and the project 66/2016 de la Sociedad Española de Neumología y Cirugía Torácica (SEPAR).

F. Vaquerizo-Villar was in receipt of a 'Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)' grant from the Ministerio de Educación, Cultura y Deporte. V. Barroso-García was in a receipt of a 'Ayuda para financiar la contratación predoctoral de personal investigador' grant from the Consejería de Educación de la Junta de Castilla y León and the European Social Fund. D. Álvarez was in receipt of a Juan de la Cierva grant from MINECO. L. Kheirandish-Gozal was supported by National Institutes of Health grant HL130984

F. Vaquerizo-Villar, G. C. Gutiérrez-Tobal, V. Barroso-García and R. Hornero, are with the Biomedical Engineering Group, Universidad de Valladolid, Paseo de Belén 15, 47011, Valladolid, Spain (phone: +34 983 423000, ext. 3708) (e-mail: fernando.vaquerizo@gib.tel.uva.es).

D. Álvarez, F. del Campo, and A. Crespo are with the Hospital Universitario Río Hortega of Valladolid, Spain (e-mail: fsas@telefonica.net).

L. Kheirandish-Gozal and D. Gozal are with The University of Chicago, USA (e-mail: dgozal@uchicago.edu).

I. Introduction

Pediatric sleep apnea-hypopnea syndrome (SAHS) is a sleep-related breathing disorder characterized by recurrent events of complete cessation (apnea) and/or significant reduction (hypopnea) of airflow during sleep [1]. Pediatric SAHS is highly prevalent (from 1 to 5%) while being under-diagnosed. This disease may cause major consequences in children's health and development, such as cardiometabolic dysfunction, neurocognitive deficits, and growth impairment [1].

The gold standard test for pediatric SAHS diagnosis is overnight polysomnography (PSG). This nocturnal test records multiple physiological signals from pediatric patients in a specialized sleep laboratory while sleeping [1]. However, PSG is costly, complex, and highly intrusive, especially for children [2], [3]. Furthermore, it results in long waiting lists, due to its limited availability and the high prevalence of pediatric SAHS.

Due to these limitations, the search for simplified diagnosis techniques has become a major challenge for the scientific community [2]. A common approach is the use of a reduced set of signals included in the PSG. In this sense, nocturnal oximetry has been frequently advocated as a simple and reliable test for children [2], [4]. In addition, oximetry only requires a finger or earlobe sensor to record the blood oxygen saturation (SpO_2) signal [4], [5].

Previous research has shown the potential usefulness of the automated analysis of nocturnal oximetry recordings to help in the screening of pediatric SAHS [5]–[9]. These studies applied different signal processing algorithms, together with different pattern recognition techniques, such as linear discriminant analysis (LDA) [5], logistic regression (LR) [6], [7], and multi-layer perceptron (MLP) neural networks [8], [9]. However, these techniques are based only on one classification model. By contrast, ensemble learning algorithms combine several classification models to outperform the results of each one separately [10]. In this sense, the adaptive boosting method (AdaBoost) is a boosting classification algorithm that has the ability to produce models

with high performance when they are validated in a new set of data [10]. In addition, AdaBoost has been successfully applied to improve the diagnostic ability of the airflow signal in SAHS screening in adult patients [11].

Based on the aforementioned considerations, we hypothesize that AdaBoost could enhance the diagnosis ability of the oximetry signal in the context of pediatric SAHS. Thus, our objective is to evaluate the usefulness of AdaBoost to assist in the screening of pediatric SAHS using the oximetry signal. To achieve this objective, a two steps-methodology was conducted: feature extraction and AdaBoost classification. In the feature extraction stage, 3% oxygen desaturation index (*ODI*3), spectral features from power spectral density (PSD), and nonlinear features were obtained from the SpO_2 signal. *ODI*3 is a conventional oximetric index commonly employed to quantify the number of SpO_2 desaturations. Spectral analysis is related to the duration and recurrence of respiratory events, whereas nonlinear analysis provides information about nonstationary properties of the SpO_2 signal associated to apneic events. These signal processing approaches have proven its usefulness to characterize the oximetry signal in pediatric SAHS screening [5]–[9]. In the second phase, an AdaBoost classifier was fed with these features (*ODI*3, spectral and nonlinear features) in order to determine the severity of SAHS according to the apnea-hypopnea index (AHI) from nocturnal PSG.

II. Subjects And Signals Under Study

The dataset analyzed in this study was composed of 981 children (602 boys and 379 girls) ranging from 2 to 13 years of age. 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 SAHS. An informed consent was obtained for each patient in order to be included in this study. In addition, the Ethical Committee approved the protocol.

Children's sleep was monitored using a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA). SpO_2 recordings were acquired at sampling rates of 25, 200, and 500 Hz. SpO_2 recordings were subsequently resampled to a common sample rate of 25 Hz, as recommended by the American Academy of Sleep Medicine (AASM), and were also rounded to the second decimal place [9]. Artifacts due to patient's movements were automatically discarded from each recording by removing sudden changes between consecutive SpO_2 sampling intervals ≥4%/s and SpO_2 values below 50% [8]. Sleep and apneic events were quantified and the AHI, which is the number of apnea and hypopnea events per hour (e/h) of sleep, was estimated using the rules of the AASM [12]. According to previous studies, children showing an AHI<1 e/h were classified as no-SAHS, those with 1≤AHI<5 e/h were classified as mild SAHS, and children showing an AHI≥5 e/h were classified as moderate-to-severe SAHS [13], [14].

The dataset was randomly divided into two sets: training set (60%), used to optimize the feature extraction stage as well as train the AdaBoost model, and test set (40%), used to assess the diagnostic performance of the AdaBoost model in an independent dataset. Table I shows clinical and demographic data of the population under study.

TABLE I. DEMOGRAPHIC AND CLINICAL DATA.

	All	Training	Test
Subjects (n)	981	589	392
Age (years)	6 [3-9]	6 [3-8]	6 [3-9]
Males (n)	602 (61.4%)	347 (58.9%)	255 (65.1%)
BMI (kg/m²)	17.9 [15.8-21.9]	17.6 [15.9-22.0]	18.1 [18.1-21.7]
AHI (e/h)	3.8 [1.5-9.3]	4.1 [1.7-9.9]	3.3 [1.4-7.8]
AHI≥1 (e/h)	806 (82.2%)	491 (83.4%)	315 (80.4%)
AHI≥5 (e/h)	405 (41.3%)	259 (44.0%)	146 (37.2%)

Data are presented as median [interquartile range], n or %. BMI: Body Mass Index; AHI: Apnea-Hypopnea Index

III. Methodology

Our proposal was developed in two stages. First, *ODI*3, spectral and nonlinear features were computed from each oximetry recording. Then, an AdaBoost classifier was trained using these features to estimate the severity of SAHS.

A. *Feature Extraction*

The following features subsets were composed to obtain an initial set of 10 features:

- Oxygen desaturation index. *ODI*3 was computed in order to obtain information about the number of SpO_2 desaturation events [15]. It is defined as the number of desaturations greater than or equal 3% per hour of recording [15].
- Spectral features. Spectral analysis was applied to explore differences in the spectral information associated to the duration and periodicity of SAHS events [6]. PSD was estimated using the Welch's method (2^{13}-sample Hamming window, 50% overlap and 2^{14}-points DFT). A band of interest was determined as the region of the PSD with more statistical differences in the PSD amplitude between the three severity groups: 0.020-0.044 Hz. Then, the following features were obtained from the PSD in this band: first-to-fourth statistical moments (*M*1*f*-*M*4*f*), maximum (*MA*) and minimum amplitude (*mA*).
- Nonlinear parameters. Nonlinear analysis has proven to be useful to extract additional information of oximetry dinamics [6]. In this sense, central tendency measure (*CTM*), Lempel-Ziv complexity (*LZC*) and sample entropy (*SampEn*) were computed to measure the variability, complexity and irregularity of the SpO_2 signal, respectively.

B. *AdaBoost Classification: AdaBoost.M2*

Boosting algorithms employ models of the same type that complement one another. These models are obtained in an iterative way and their individual outputs are combined using a weighted vote scheme [10]. AdaBoost is a widely used boosting method that can be employed with any classifier. However, its prediction ability on a new set of data decreases if it is employed with complex classifiers [10]. Thus, it is preferable to use AdaBoost with a simple classification

algorithm as weak learner [10]. In this study, the well-known LDA algorithm has been chosen as weak learner, which has previously shown its usefulness in the context of adult SAHS [11]. Since our problem is a 3-class classification task (AHI<1 e/h, 1≤AHI<5 e/h, and AHI≥5 e/h), the AdaBoost.M2 version of the algorithm has been used [10].

AdaBoost.M2 assigns a weight w_j^k to every instance x_j in the training set (j=1,2, …, M, being M the number of instances) at each k iteration. The *k-th* weak classifier is then trained using the weighted instances. Then, the error ϵ_k of the classifier is calculated using a weighted pseudo-loss [16]:

$$\epsilon_k = \frac{1}{2} \cdot \sum_{j=1}^{M} \sum_{c \neq c_{true}} w_{j,c}^k \cdot \left(1 - h_k(x_j, c_{true}) + h_k(x_j, c)\right) \quad (1)$$

where c is the number of classes, c_{true} is the class of the instance x_j, and h_k is the confidence of the prediction of the weak classifier for an instance x_j and a class from c.

This error ϵ_k is used to obtain the weighted vote α_k for this *k-th* classifier [16]:

$$\alpha_k = ln\left(\frac{(1-\epsilon_k)}{\epsilon_k}\right) \quad (2)$$

Classifiers with smaller error ϵ_k have a higher weight α_k, thus having more importance on the final model. At the end of each iteration, the weights of the training instances are updated, assigning a higher weight w_j^{k+1} to those instances j that have not been rightly classified [10]. These weights are normalized so that their sum remains the same as in the previous iteration. Thus, the classifier focuses more on instances with a high weight in the next iteration in order to not misclassify them again [10].

At the end of the iterative process, the algorithm AdaBoost.M2 performs the final classification task for each instance x_j. It obtains, for each x_j, the class where the sum of the votes of all the classifiers is maximum, taking into account their weighted votes α_k [10].

C. Statistical Analysis

The Kruskal-Wallis test was used to search for significant statistical differences (p-value <0.01) among SAHS severity groups (AHI<1 e/h, 1≤AHI<5 e/h, and AHI≥5 e/h). Diagnostic ability of the AdaBoost classifier was assessed by means of sensitivity (Se, percentage of SAHS positive patients rightly classified), specificity (Sp, percentage of SAHS negative children rightly classified), accuracy (Acc, percentage of subjects rightly classified), and Cohen's kappa index (kappa). SMOTE was applied to the no-SAHS group (98 subjects) in the training set in order to compensate for imbalance with respect to the mild SAHS group (232 subjects), and moderate-to-severe SAHS (259 subjects) when training the AdaBoost.M2 classifier. Therefore, synthetic samples were obtained for each feature to reach 196 no-SAHS subjects.

IV. Results

A. Training Set

The proposed 10 features were computed in the training group to conform the initial feature set. All features except *M3f* and *M4f* reached significant statistical differences (p-value <0.01) between SAHS severity groups. Then, the 687 samples (589 originals and 98 synthetic) of the training set, each one composed of these features, fed the AdaBoost.M2 algorithm. AdaBoost.M2 was trained in a 100 steps iteration procedure, in order to minimize the overfitting chances.

B. Test Set

The AdaBoost.M2 model obtained was further validated in the independent test set. Table II shows the confusion matrices for this AdaBoost.M2 model, as well as the conventional oximetric index *ODI*3. These models rightly classified 66.3% (AdaBoostM2) and 62.5% (*ODI3*) of the subjects in the test set, whereas the 3-class kappa were 0.474 (AdaBoostM2) and 0.410 (*ODI*3), respectively. Table III shows the diagnostic performance parameters derived from the confusion matrices when assessing individual AHI cutoffs values of 1 and 5 e/h. Notice that the results obtained with AdaBoost.M2 slightly improved the performance of *ODI3* in terms of Acc for both AHI cutoffs: 80.9% *vs*. 79.1% (AHI≥1 e/h) and 82.9% *vs*. 81.6% (AHI≥5 e/h).

V. Discussion

In this study, the utility of an AdaBoost.M2 model was assessed to determine the severity of pediatric SAHS using the oximetry signal. This model was fed with *ODI*3, spectral, and nonlinear features. It outperformed *ODI*3 in the 3-class (no SAHS, mild SAHS, and moderate-to-severe SAHS) classification task in terms of overall Acc (66.3% *vs*. 62.5%), as well as kappa (0.474 *vs*. 0.410). Furthermore, AdaBoost.M2 reached higher Acc than *ODI*3 alone when assessing the AHI thresholds of 1 e/h (80.9% *vs*. 79.1%) and 5 e/h (82.9% *vs*. 81.6%). Nevertheless, *ODI*3 achieved a slightly higher Se for both cutoffs (1 e/h and 5 e/h). These cutoffs were not arbitrary selected. They define the presence of SAHS (mild SAHS, AHI≥1 (e/h)) and the need for surgical treatment (moderate-to-severe SAHS, AHI≥5 (e/h)) [13], [14]. In this regard, and according to the confusion matrices, 98.3% (115/117) of subjects predicted as moderate-to-severe SAHS (AHI≥5 e/h) by the AdaBoost.M2 model are at least mild patients (AHI≥1 e/h). Furthermore, 91.1% (82/90) of subjects predicted as no-SAHS (AHI<1 e/h) do not suffer from SAHS or are mild SAHS.

Previous studies also assessed the usefulness of the automated analysis of the oximetry signal in the screening of pediatric SAHS [5]–[9]. Garde et al. [5] assessed a LDA model trained with PSD and time domain features from a database of 146 SpO_2 and PR recordings, achieving 88.4% Se and 83.6% Sp (AHI≥5 e/h). Álvarez et al. [6] evaluated binary LR models trained with automatic features from 50 SpO_2 recordings for different AHI cutoffs (1, 3, and 5 e/h), reaching 85.5% Acc for AHI≥1 e/h, 83.4% Acc for AHI≥3 e/h, and 82.8% Acc for AHI≥5 e/h. Crespo et al. [7] employed a LR model fed with features from multiscales entropy and conventional oximetric indices computed from the SpO_2 signal. This model achieved 83.5% Acc (AHI≥3 e/h) in a database of 50 recordings. Nonetheless, these studies [5]–[7] only aimed at the detection of pediatric SAHS, thus only focused on binary classification. In addition, these studies used smaller databases than in our proposal [5]–[7].

In contrast to these studies [5]–[7], Vaquerizo-Villar et

TABLE II. CONFUSION MATRICES FOR ADABOOST.M2 AND *ODI3* IN THE TEST SET

		Estimated					
		AdaBoost.M2			***ODI3***		
		AHI<1	**1≤AHI<5**	**AHI≥5**	**AHI<1**	**1≤AHI<5**	**AHI≥5**
Actual	**AHI≤1**	**46**	29	2	**37**	37	3
	1≤AHI≤5	36	**116**	17	38	**107**	24
	AHI≥5	8	40	**98**	4	41	**101**

TABLE III. DIAGNOSTIC ABILITY OF ADABOOST.M2 AND *ODI3* IN THE TEST SET FOR AHI CUTOFFS = 1 E/H AND 5 E/H.

	AdaBoost.M2		***ODI3***	
	1	**5**	**1**	**5**
Se (%)	86.0	67.1	86.7	69.2
Sp (%)	59.7	92.3	48.1	89.0
Acc (%)	**80.9**	**82.9**	**79.1**	**81.6**
kappa (%)	**0.474**		**0.410**	

al. [8] and Hornero et al [9] developed MLP models in order to determine pediatric SAHS severity. Vaquerizo-Villar et al [8] assessed a classification MLP model fed with *ODI*3, PSD, bispectral, and anthropometric variables using a database of 298 SpO_2 recordings, reaching 81.3% and 85.3% Acc in the detection of moderate-to-severe (AHI ≥5 e/h) and severe (AHI ≥10 e/h) SAHS, respectively. Hornero et al [9], in a multicenter international study, developed a regression MLP model built with automated features computed from 4191 SpO_2 recordings from 13 sleep laboratories, reaching 75.2%, 81.7%, and 90.2% Acc for common AHI severity cutoffs of 1 e/h, 5 e/h, and 10 e/h, respectively. However, our proposal outperformed the results obtained in these studies for the detection of mild and moderate-to-severe SAHS, especially in the former case (AHI ≥1 e/h).

Despite our proposal reached promising results, there exist some limitations that must be taken into account. First, in spite of the number of subjects is high, more subjects with an AHI<1 e/h would be necessary to improve the AdaBoost.M2 model training. Additionally, further validation of our proposal would be required in order to assess the effects of regularization and different number of LDA classifiers for the AdaBoost.M2 algorithm as well as to compare the Adaboost.M2 model with the performance of other classification algorithms. Finally, it would also be appropriate to validate this methodology in unsupervised oximetry recordings obtained at patient's home.

VI. CONCLUSIONS

In summary, a multiclass Adaboost.M2 model fed with automated parameters computed from the oximetry signal achieved high performance in an independent test set in order to assess pediatric SAHS severity. Furthermore, this model achieved slightly higher diagnostic accuracies than the classical clinical index *ODI3*, as well as other state-of-the-art studies, in the detection of mild and moderate-to-severe SAHS. Therefore, we can conclude that Adaboost could help to improve the diagnostic ability of the oximetry signal in the context of pediatric SAHS.

REFERENCES

[1] C. L. Marcus *et al.*, "Diagnosis and management of childhood obstructive sleep apnea syndrome.," *Pediatrics*, vol. 130, no. 3, pp. 576–84, 2012.

[2] G. M. Nixon, A. S. Kermack, G. M. Davis, J. J. Manoukian, A. Brown, and R. T. Brouillette, "Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry," *Pediatrics*, vol. 113, no. 1, pp. e19–e25, 2004.

[3] E. S. Katz, R. B. Mitchell, and C. M. D. Ambrosio, "Obstructive Sleep Apnea in Infants," *Am. J. Respir. Crit. Care Med.*, vol. 185, no. 8, pp. 805–816, 2012.

[4] N. Netzar, A. H. Eliasson, C. Netzar, and D. A. Kristo, "Overnight Pulse Oximetry for Sleep- Disordered Breathing in Adults," *CHEST J.*, vol. 120, no. 2, pp. 625–633, 2001.

[5] A. Garde, P. Dehkordi, W. Karlen, D. Wensley, J. M. Ansermino, and G. A. Dumont, "Development of a screening tool for sleep disordered breathing in children using the phone oximeterTM," *PLoS One*, vol. 9, no. 11, 2014.

[6] D. Álvarez *et al.*, "Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry : An Alternative to Respiratory Polygraphy in Unattended Settings," *J Clin Sleep Med*, vol. 13, no. 5, pp. 7–11, 2017.

[7] A. Crespo *et al.*, "Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home," *Entropy*, vol. 19, no. 6, p. 284, 2017.

[8] F. Vaquerizo-Villar *et al.*, "Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings," *Comput. Methods Programs Biomed.*, vol. 156, pp. 141–149, 2018.

[9] R. Hornero *et al.*, "Nocturnal Oximetry-based Evaluation of Habitually Snoring Children.," *Am. J. Respir. Crit. Care Med.*, vol. 196, no. 12, pp. 1591–1598, 2017.

[10] I. Witten, E. Frank, and M. Hall, *Data Mining*. 2011.

[11] G. C. Gutierrez-Tobal, D. Álvarez, F. del Campo, and R. Hornero, "Utility of AdaBoost to Detect Sleep Apnea-Hypopnea," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 636–646, 2016.

[12] R. B. Berry, R. Budhiraja, D. . Gottlieb, and E. Al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, 2012.

[13] A. Kaditis, L. Kheirandish-Gozal, and D. Gozal, "Pediatric OSAS : Oximetry can provide answers when polysomnography is not available," *Sleep Med. Rev.*, vol. 27, pp. 96–105, 2016.

[14] P. E. Brockmann, C. Schaefer, A. Poets, C. F. Poets, and M. S. Urschitz, "Diagnosis of obstructive sleep apnea in children : A systematic review," *Sleep Med. Rev.*, vol. 17, no. 5, pp. 331–340, 2013.

[15] U. J. Magalang and J. Dmochowski, "Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry," *CHEST J.*, vol. 124, no. 5, pp. 1694–1701, 2003.

[16] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," vol. 55, no. 1, pp. 23–37, 1995.