Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings

Fernando Vaquerizo-Villar[a,*], Daniel Álvarez[a,b], Leila Kheirandish-Gozal[c], Gonzalo C. Gutiérrez-Tobal[a], Verónica Barroso-García[a], Andrea Crespo[a,b], Félix del Campo[a,b], David Gozal[c], and Roberto Hornero[a,d]

[a] Biomedical Engineering Group, Universidad de Valladolid, Valladolid, Spain.

[b] Servicio de Neumología, Hospital Universitario Rio Hortega, Valladolid, Spain.

[c] Dept. of Pediatrics, Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, United States of America.

[d] IMUVA, Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, Spain.

* Corresponding author: **Fernando Vaquerizo-Villar**

Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, Universidad de Valladolid, Campus Miguel Delibes, Paseo Belén 15, 47011 – Valladolid, Spain.

Tel. +34 983 423000 ext. 3708

E-mail address: fervaquerizo@gmail.com

URL: www.gib.tel.uva.es

Abstract- *Background and objective.* The aim of this study was to assess the utility of bispectrum-based oximetry approaches as a complementary tool to traditional techniques in the screening of pediatric sleep apnea-hypopnea syndrome (SAHS). *Methods:* 298 blood oxygen saturation (SpO_2) signals from children ranging 0-13 years of age were recorded during overnight polysomnography (PSG). These recordings were divided into three severity groups according to the PSG-derived apnea hypopnea index (AHI): AHI<5 events per hour (e/h), 5≤AHI<10 e/h, AHI≥10 e/h. For each pediatric subject, anthropometric variables, 3% oxygen desaturation index (*ODI*3) and spectral features from power spectral density (PSD) and bispectrum were obtained. Then, the fast correlation-based filter (FCBF) was applied to select a subset of relevant features that may be complementary, excluding those that are redundant. The selected features fed a multiclass multi-layer perceptron (MLP) neural network to build a model to estimate the SAHS severity degrees. *Results:* An optimum subset with features from all the proposed methodological approaches was obtained: variables from bispectrum, as well as PSD, *ODI3*, *Age*, and *Sex*. In the 3-class classification task, the MLP model trained with these features achieved an accuracy of 76.0% and a Cohen's kappa of 0.56 in an independent test set. Additionally, high accuracies were reached using the AHI cutoffs for diagnosis of moderate (AHI=5 e/h) and severe (AHI=10 e/h) SAHS: 81.3% and 85.3%, respectively. These results outperformed the diagnostic ability of a MLP model built without using bispectral features. *Conclusions:* Our results suggest that bispectrum provides additional information to anthropometric variables, *ODI3* and PSD regarding characterization of changes in the SpO_2 signal caused by respiratory events. Thus, oximetry bispectrum can be a useful tool to provide complementary information for screening of moderate-to-severe pediatric SAHS.

Keywords- Sleep apnea-hypopnea syndrome (SAHS), children, oximetry, bispectrum, feature selection, feature classification

1. INTRODUCTION

Childhood sleep apnea-hypopnea syndrome (SAHS) is a breathing disorder characterized by recurrent episodes of complete cessation (apnea) and/or significant reduction (hypopnea) of airflow during sleep due to the presence of increased upper airway collapsibility [1]. According to the American Academy of Pediatrics (AAP), SAHS has a prevalence in the range of 1% to 5% and it may impose many adverse effects on the health and development of infants and young children, such as neurocognitive deficits, cardio-metabolic dysfunction, and somatic growth stunting [1].

The gold standard test for pediatric SAHS diagnosis is overnight polysomnography (PSG). PSG requires the patient to spend the night in a specialized sleep laboratory while recording a wide range of biomedical signals [2,3]. Thus, PSG is costly and complex due to the necessary expensive equipment and intensive labor of medical personnel. It is also especially intrusive for children, due to the use of multiple sensors, and shows limited availability in most places around the world [4,5].

These drawbacks, together with the relatively high prevalence of the disease, have led the scientific community to explore the use of simplified screening tests [4]. The guidelines of the AAP recommend performing alternative tests when PSG is not available while requiring more conclusive evidences about the efficacy of these tests [1]. Thus, a commonly used approach has been the assessment of a reduced set of cardiorespiratory recordings. Cardiorespiratory signals contain essential information about the alterations produced by apneic events in the electrocardiogram (ECG) [6], pulse rate variability (PRV) [7,8], airflow (AF) [9], photopletysmography [10], oximetry [7,9,11–19], and acoustic pulmonary sounds [20,21]. Among these approaches, nocturnal oximetry is the alternative most frequently advocated. In the

nocturnal oximetry, pulse rate and blood oxygen saturation (SpO_2) signals are recorded with a pulse oximeter probe, typically placed on the earlobe, finger, or toe [22]. Moreover, SpO_2 signals can be recorded in an unsupervised way at the patient's home due to the development of commercial portable pulse oximeters [4,7]. Previous studies have shown the utility of the SpO_2 signal to assist in the SAHS diagnosis in both adults [23,24] and children [7,9,11–19]. In this study, we aim at gaining further insights into the diagnostic ability of the SpO_2 signal in the screening of pediatric SAHS.

Different techniques have been reported to automatically analyze biomedical signals in the context of SAHS. Several studies have assessed the performance of frequency domain features, which reflects the duration and periodicity of respiratory events in children [6–10,18–21]. Power Spectral Density (PSD) is the most common spectral analysis technique in these studies [6–9,18–21]. However, the information present in the PSD cannot characterize phase relationships and deviations from gaussianity in a signal [25]. By contrast, bispectrum is a frequency domain technique defined as the spectral representation of the third order statistic that contains information about the phase of the Fourier transform of a time series. It can detect deviations from linearity, stationarity, and gaussianity in the signal, such as those produced in physiological recordings by respiratory events [25].

Based on the aforementioned considerations, we hypothesized that bispectrum analytic could provide additional information about respiratory events, thus being a complementary tool to *ODI*3, anthropometric variables, and PSD parameters. Therefore, the aim of this study was to evaluate the complementarity of bispectrum to traditional approaches in the screening of pediatric SAHS using SpO_2 recordings.

We conducted our study in three phases: feature extraction, feature selection, and feature classification. First, anthropometric variables, *ODI*3 [26], and spectral features

from PSD and bispectrum were obtained. Then, the fast correlation-based filter (FCBF) method [27] was applied to select a smaller subset of relevant and non-redundant features. Finally, a multi-layer perceptron (MLP) neural network [23] was applied to this optimum subset for multiclass (3-class) classification in order to estimate the SAHS severity degrees according to the apnea-hypopnea index (AHI) from standard PSG.

2. SUBJECTS AND SIGNALS UNDER STUDY

The dataset was composed of 298 children (166 boys and 132 girls) ranging 0-13 years of age. All children were consecutively and prospectively referred to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) due to clinical symptoms and physical examination findings leading to the clinical suspicion of SAHS. In all participants, an informed consent was obtained as a prerequisite to be included in the study. The Ethical Committee of the University of Chicago Medicine Comer Children's Hospital approved the protocol.

Sleep was monitored using a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA). SpO_2 recordings were acquired during overnight PSG at a sampling rate of 25 Hz. They were exported and processed offline. Artifacts were discarded from oximetric recordings by removing SpO_2 values below 50% and sudden changes between consecutive SpO_2 sampling intervals ≥4%/second [28].

Sleep and cardiorespiratory events were scored and quantified by specialized technologists who were unaware of the study purpose, and AHI was estimated according to the American Academy of Sleep Medicine guidelines. In this regard, there is no consensus about the AHI cutoff used to determine SAHS in infants [1–3,29,30]. However, a wide range of studies typically classify children showing 5≤AHI<10 e/h as moderate SAHS and children with AHI ≥10 e/h as severe SAHS [2,3,29,30]. Hence, we

have classified the subjects under study into the three groups defined by these commonly used thresholds (AHI<5 e/h, 5≤AHI<10 e/h, and AHI≥10 e/h).

The dataset was randomly divided into three sets: feature optimization set (25%), training set (50%), and test set (25%). The first set (feature optimization set) was employed to optimize the feature extraction stage and obtain an optimum subset of features with FCBF. A bootstrap procedure was applied to select the optimum features in order to select a generalizer optimum subset of features [31]. The second set (training set) was used to select the optimal design parameters of the MLP classifier as well as train the MLP model. Ten-fold cross validation was used to emulate a different dataset when optimizing the MLP design parameters [32]. Finally, the third set (test set) was employed to assess the diagnostic performance of our proposal in an independent dataset (unknown data). Table 1 shows clinical and demographic data of the population under study. No statistically significant differences (p-value < 0.05) were found in the Age and Body Mass Index (BMI) between the three groups.

PLEASE, DISPLAY TABLE 1 AROUND HERE

3. METHODS

3.1. Feature Extraction

Four clinical and signal processing approaches were applied to each subject: anthropometric, *ODI*3, and PSD and bispectrum, which lead to an initial feature set composed of 22 features.

3.1.1. Anthropometric variables

Age, sex and BMI were acquired for each child since the prevalence of childhood SAHS has been associated with these factors in previous studies [1].

3.1.2. Oxygen desaturation index

In order to obtain information about the number of desaturations produced by respiratory events, *ODI*3 was computed for each SpO_2 recording [2]. The definition of a desaturation event employed for computing *ODI*3 is based in the study developed by Taha et al. [26]. In this study, a desaturation event occurs when SpO_2 value decreases at least by 3% with respect to the preceding baseline levels, at a rate between 0.1% and 4% per second, and the SpO_2 value subsequently returns to the baseline level or increases by at least 3% with respect to the preceding minimum value. The total duration of the event must be between 10 and 60 seconds.

3.1.3. Power Spectral Density (PSD)

PSD was estimated for each SpO_2 recording to explore differences in the spectral information of SpO_2 signals associated to the duration and recurrence of apneic events. Welch's method was used to estimate PSDs [33], using a Hamming window of 5 minutes (7500 samples) with 50% overlap and a discrete Fourier transform (DFT) of 2^{14} points.

According to previous research in the context of childhood SAHS diagnosis [9], a frequency band of interest was determined as the frequency region of the PSD where there were statistically significant differences (*p*-value under 0.05) between severity groups (AHI<5 e/h, 5≤AHI<10 e/h, and AHI≥10 e/h) in the feature optimization set. *P*-value was computed between the PSD amplitudes for each pair of severity groups at each frequency using the non-parametric Mann-Whitney *U* test. Accordingly, our band of interest was 0.018-0.050 Hz. In this band, higher PSD amplitude is obtained as the severity of SAHS increases.

The following parameters of the PSD were computed in the band of interest:

- First to fourth-order statistical moments (*M*1*f*-*M*4*f*) of the PSD amplitudes.

The mean (*M1f*), variance (*M2f*), skewness (*M3f*) and kurtosis (*M4f*) quantify the central tendency, dispersion, asymmetry and peakedness of the power spectrum, respectively.

- Relative power (P_R). P_R is defined as the ratio between the power (area enclosed under the PSD) in the band of interest and the total signal power.
- Maximum amplitude (*MA*) and minimum amplitude (*mA*) of the PSD.
- Spectral entropy (*SE*), is a irregularity measure which quantifies the flatness of the PSD [34].
- Mobility (*Mb*). It is a Hjorth descriptor, which measures the concentration of the signal power. It is defined as the squared root of the ratio between the variance (*M2f*) and the signal power [35].

3.1.4. Bispectrum

High order spectra (HOS) are representations in the frequency domain of high order cumulants of a random process [25]. PSD is the Fourier transform of the second-order cumulant, while bispectrum and trispectrum are the spectral representations of the third- and fourth-order cumulant, respectively [25]. Bispectrum can be described as a spectral decomposition of the skewness of a signal over frequency. In contrast to conventional PSD, bispectrum contains additional information about the phase relationships and deviations from gaussianity, linearity, and stationarity of a signal [25].

Let be $x(n)$ a deterministic and zero-mean signal, the bispectrum is expressed in terms of the Fourier transform of the signal $X(f)$ [25]:

$$\begin{aligned} B(f_1, f_2) &= \left\{ \sum_{m=-\infty}^{\infty} x(m) \exp[-j(f_1 m)] \right\} \times \left\{ \sum_{k=-\infty}^{\infty} x(k) \exp[-j(f_2 k)] \right\} \\ &\times \left\{ \sum_{n=-\infty}^{\infty} x(n) \exp[+j(f_1 + f_2) n] \right\} = X(f_1)\, X(f_2)\, X^*(f_1 + f_2), \end{aligned} \tag{1}$$

where f_1 and f_2 are the frequency indices. Due to the symmetry conditions of the bispectrum, it is sufficient to evaluate the bispectrum in the triangular region (Ω) that satisfies $f_2 \geq 0$, $f_2 \geq f_1$, $f_1 + f_2 \leq f_s/2$, where f_s is the sampling frequency of the signal [25]. In this study, bispectrum was estimated with a non-parametric approach using a Hamming window of 5 minutes with 50% overlap and a DFT of 2^{14} points. Figure 1 shows the averaged magnitude of the bispectrum for the three severity groups. Notice that higher amplitude in the bispectrum is observed at frequencies below 0.03 Hz, as the SAHS severity increases.

PLEASE, DISPLAY FIGURE 1 AROUND HERE

The following bispectral features were extracted in the region Ω to quantify the differences in the bispectrum between groups [36,37]:

- Mean amplitude of the bispectrum ($MB1$). This parameter is intended to differentiate between signals with similar PSD but different bispectrum [37]:

$$MB1 = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} |B(f_1, f_2)|, \tag{2}$$

  where L is the number of points in region Ω.

- Sum of logarithmic amplitudes of the bispectrum ($H1$), sum of logarithmic amplitudes of elements in the diagonal of the bispectrum ($H2$), and first-order spectral moment of amplitudes of elements in the diagonal of the bispectrum ($H3$) [36]. These parameters are related to the moments of the bispectrum [36]:

$$H1 = \sum_{f_1, f_2 \in \Omega} \log\left(\left|B\left(f_1, f_2\right)\right|\right), \tag{3}$$

$$H2 = \sum_{f_k \in \Omega_{diagonal}} \log\left(\left|B\left(f_k, f_k\right)\right|\right), \tag{4}$$

$$H3 = \sum_{f_k \in \Omega_{diagonal}} k \log\left(\left|B\left(f_k, f_k\right)\right|\right), \tag{5}$$

where $\Omega_{diagonal}$ is the diagonal of the bispectrum.

- Normalized bispectral entropy (*BE*1) and normalized bispectral squared entropy (*BE*2), which quantify regularity in the amplitude of the bispectrum [37]:

$$BE1 = -\sum_{j \in \Omega} p_j \ln\left(p_j\right), \tag{6}$$

where

$$p_j = \frac{\left|B\left(f_1, f_2\right)\right|}{\sum_{f_1, f_2 \in \Omega} \left|B\left(f_1, f_2\right)\right|}, \tag{7}$$

$$BE2 = -\sum_{j \in \Omega} q_j \ln\left(q_j\right), \tag{8}$$

where

$$q_j = \frac{\left|B\left(f_1, f_2\right)\right|^2}{\sum_{f_1, f_2 \in \Omega} \left|B\left(f_1, f_2\right)\right|^2}. \tag{9}$$

- Phase entropy (*PE*), which measures regularity in the phase of the bispectrum [37]:

$$PE = -\sum_{n\in\Omega} p(\Psi n)\ \ln\left(p(\Psi n)\right), \tag{10}$$

where

$$p(\Psi n) = \frac{1}{L}\sum_{f_1,f_2\in\Omega} Ind\left(\phi\left(B(f_1,f_2)\right)\in\Psi n\right), \tag{11}$$

$$\Psi n = \left\{\phi \left| -\pi + \frac{2\pi}{n} \le \phi < -\pi + \frac{2\pi\,(n+1)}{N}\right.\right\}, \quad n = 0,1,...,N-1, \tag{12}$$

where ϕ is the phase angle of the bispectrum, $Ind(\cdot)$ is the indicator function, whose value is 1 if ϕ is within the range of histogram bins ψ_n, and N is the number of bins of the histogram, being calculated according to Doane's rule [38].

- Mean (*meanPa*) and variance (*varPa*) of the bispectrum invariant ($P(a)$). These features identify a chaotic process with third-order time correlations or phase coupling between spectral components [37]. $P(a)$ is the phase of the integrated bispectrum along a radial with slope a [37]:

$$P(a) = \arctan\left(\frac{I_i(a)}{I_r(a)}\right), \tag{13}$$

where $I_r(a)$ and $I_i(a)$ are the real and imaginary part of $I(a)$:

$$I(a) = \int_{f_1=0^+}^{1/1+a} B(f_1, af_1)\,df_1 = I_r(a) + j\,I_i(a), \tag{14}$$

for $0\le a\le 1$.

3.2. Feature selection: Fast Correlation-Based Filter (FCBF)

After the feature extraction stage, FCBF is applied to select a subset of relevant and non-redundant features [27]. FCBF has previously shown its utility in the context of adult SAHS diagnosis [39]. It is based on symmetrical uncertainty (*SU*), which is a normalization of the information gain (*IG*) between two variables [27]. First, features are ranked from the most relevant ones (highest *SU* with the AHI). Then, the features considered redundant with respect to features that are more relevant were discarded. Thus, an optimum subset with the most relevant and non-redundant features is obtained [27].

In order to compose an optimum feature subset independent of a particular dataset, 1000 bootstrap replicates were built from our feature optimization set so that the FCBF method was applied to each bootstrapping subset [31,40]. An optimum subset composed of the variables that appear at least in 500 runs was selected.

3.3. Feature classification: Multi-layer Perceptron (MLP) neural network

Artificial neural network are mathematical models inspired in the human brain [41]. MLP are probably the most widely used neural network and it has already proven its usefulness in the context of adult SAHS diagnosis with SpO_2 recordings [23]. Its architecture consists on several interconnected layers (input, hidden, and output layers) composed of simple units called perceptrons. Each unit is characterized by an activation function and adaptive weights representing connections with units from the subsequent layer.

Since our problem is a 3-class classification task, the output layer has three output neurons, each one representing the posterior probability of belonging to each group. In addition, a configuration with a single hidden layer has been implemented, which may provide universal approximation to any function [41]. Weights of the network are

randomly initialized. Then, they are optimized using the scaled conjugate gradient with weight-decay regularization. It is used to minimize the cross-entropy error function and achieve good generalization, as recommended for classification tasks [41].

The Netlab toolbox was used to implement our MLP classifier [42]. A very complex MLP model leads to overfitting, whereas a very simple model leads to under-fitting. Thus, the design parameters of the MLP (the number of units in the hidden layer (N_H) and the regularization parameter (α)) were optimized by means of 10-fold cross-validation using the training set. Then, the MLP model was built using the whole training set with the optimum design parameters.

3.4. Statistical Analysis and Diagnostic Performance

The Mann Whitney *U* test and the Kruskal Wallis test were used to assess statistical differences (p-value <0.05) between groups. The Bonferroni correction was applied to deal with multiple comparisons. Diagnostic ability of the MLP network was assessed by means of sensitivity (Se, percentage of SAHS positive patients correctly classified), specificity (Sp, percentage of SAHS negative children correctly classified), positive predictive value (PPV, proportion of subjects classified as positive that are true positives), negative predictive value (NPV, proportion of subjects classified as negative that are true negatives), positive likelihood ratio (LR+, likelihood ratio for subjects classified as positive), negative likelihood ratio (LR-, likelihood ratio for subjects classified as negative), accuracy (Acc, percentage of subjects correctly classified), and Cohen's kappa index (kappa) [43].

4. RESULTS

4.1. Feature optimization and selection

A total of 22 features were obtained for each subject: 3 anthropometric variables,

ODI3, 9 parameters from PSD, and 9 bispectral features. Table 2 displays the values of these features for each SAHS severity group in the feature optimization set (median [interquartile range]), along with their corresponding *p*-values. *ODI3*, 6 out of 9 features from PSD (*M*1*f*, *M*2*f*, *MA*, *mA*, *SE*, and *Mb*) and 4 out of 9 features from bispectrum (*MB*1, *H*1, *H*2, and *H*3) showed statistical significant differences (*p*-value <0.05). These features showed higher values as the severity of SAHS increased.

PLEASE, DISPLAY TABLE 2 AROUND HERE

In order to assess the complementarity of bispectrum with respect to the other methodological approaches, two different feature sets were composed. The first one (set_{nobis}) consisting of all but bispectrum features and the second one (set_{bis}) consisting of all extracted features. FCBF was applied to each bootstrap replication generated with these feature sets (set_{nobis} and set_{bis}) using only the feature optimization group. In both cases, an optimum subset composed of the features selected at least 500 times was obtained. The optimum subset derived when applying FCBF to set_{nobis} ($subset_{nobis}$) was composed of 3 anthropometric features (*Age*, *Sex*, and *BMI*), *ODI*3, and 5 features from the PSD (*M*1*f*, *M*2*f*, *MA*, P_R, and *Mb*). Regarding the optimum subset obtained when applying FCBF to set_{bis} ($subset_{bis}$), it was composed of 2 anthropometric features (*Age* and *Sex*), *ODI*3, 5 features from the PSD (*M*1*f*, *M*2*f*, *MA*, P_R, and *Mb*), and 2 bispectral features (*MB*1 and *meanPa*). Notice that two bispectral features were selected: one amplitude bispectral feature (*MB*1) and one phase bispectral feature (*meanPa*).

4.2. Model optimization and training

Two MLP networks fed with these optimum subsets of features obtained with FCBF (MLP_{nobis}: $subset_{nobis}$; MLP_{bis}: $subset_{bis}$) were designed and trained using the

training set. N_H was varied from 2 up to 50, while α was varied from 0 up to 5. Kappa was obtained through ten-fold cross validation for each N_H-α pair, and the optimum values for N_H and α were obtained as those for which kappa was higher. Due to the dependence of the network to the initial random values of the weights, kappa was computed and averaged for a total of 10 runs for each N_H-α pair. Finally, user-dependent network parameters $N_H = 3$ and $\alpha = 1$ were chosen using $\text{subset}_{\text{nobis}}$ and $N_H = 4$ and $\alpha = 2$ were chosen using $\text{subset}_{\text{bis}}$, since those pairs reached the highest kappa. The entire training set was used to train the corresponding MLP models in both cases ($\text{MLP}_{\text{nobis}}$ and MLP_{bis}).

4.3. Diagnostic performance assessment

Table 3 shows the confusion matrices of the MLP models ($\text{MLP}_{\text{nobis}}$ and MLP_{bis}) in the test group. These matrices show the class estimated by our MLP classifiers for each subject versus the actual SAHS severity group of the subjects in the test set. The overall accuracies (sum of the main diagonal elements) of these models in the test set were 69.3% ($\text{MLP}_{\text{nobis}}$) and 76.0% ($\text{MLP}_{\text{bis}}$), whereas the 3-class kappa values were 0.45 ($\text{MLP}_{\text{nobis}}$) and 0.56 ($\text{MLP}_{\text{bis}}$). Table 4 shows the diagnostic ability of these models for AHI cutoffs = 5 and 10 e/h.

Notice that the results obtained with the model MLP_{bis} outperformed $\text{MLP}_{\text{nobis}}$ in terms of Se, Sp, PPV, NPV, LR, LR-, Acc, and kappa for both cutoffs.

PLEASE, DISPLAY TABLE 3 AROUND HERE

PLEASE, DISPLAY TABLE 4 AROUND HERE

5. DISCUSSION

This study assessed the usefulness of bispectrum to provide additional information from SpO_2 recordings in the screening of pediatric SAHS. The results obtained with our proposed approach suggest that the information provided by bispectrum is relevant and complementary.

Our findings showed that significantly higher values in 4 out of 9 features from bispectrum (*MB*1, *H*1, *H*2, and *H*3) are present in the subjects with the most severe degrees of SAHS. The statistical differences between groups of these bispectral features are consistent with the higher values of the bispectrum observed in Fig 1. MLP_{bis}, which was fed with optimum features from all signal processing approaches, outperformed the neural network without information from bispectrum (MLP_{nobis}). Regarding the optimum feature subset, FCBF automatically selected *Age*, *Sex* (anthropometric); *ODI*3 (oximetric index); *M*1*f*, *M2f*, *MA*, P_R, *Mb* (PSD); *MB*1 and *meanPa* (bispectrum). Moreover, the results obtained in this stage suggest that information from bispectrum-based variables is complementary to that obtained from conventional approaches.

In the test set, the proposed 3-class neural network (MLP_{bis}) achieved an overall Acc of 76%, as well as kappa=0.56, with 81.3% Acc and 85.3% Acc for the common cutoffs AHI=5 e/h and AHI=10 e/h, respectively. It is remarkable to say that, with our MLP model (MLP_{bis}), a high positive predictive value (95.5%) is obtained for an AHI cutoff of 5 e/h, whereas a high negative predictive value (86.7%) is obtained for an AHI cutoff of 10 e/h. These cutoffs (5 and 10 e/h) were not arbitrary selected. They are commonly employed in clinical settings to define the boundary for moderate ($5 \leq AHI < 10$ e/h) and severe ($AHI \geq 10$ e/h) SAHS [2,3,29,30]. For patients with an $AHI \geq 5$ e/h, treatment with adenotonsillectomy is recommended [2]. Furthermore, children with an $AHI \geq 10$ e/h have an increased risk for cardiac strain and overnight

observation is recommended after treatment. In this sense, continuous positive airway pressure (CPAP) is recommended in these cases when other treatment strategies such as surgery have failed [2].

To the best of our knowledege, this is the first study in the context of pediatric SAHS using bispectrum. Two parameters from bispectrum, *MB*1 and *meanPa*, were involved in the optimum feature subset obtained with FCBF. These features contain information about the amplitude (*MB*1) and the phase (*meanPa*) of the bispectrum. Thus, the changes in the amplitude and phase of the bispectrum of oximetric recordings detected by *MB*1 and *meanPa* can provide additional information about oximetry recordings to assist in pediatric SAHS screening.

Table 5 shows the performance of previous research focused on the use of SpO_2 recordings in the screening of pediatric SAHS [7,9,11–19]. ODI has been used for this task [11,12]. Kirk et al. [11] reached 67% Se, 60% Sp, and 64% Acc (AHI≥5) using *ODI3*. Tsai et al. [12] reported 79.0% Acc for AHI≥1 (77.7% Se and 88.9% Sp), 85.1% Acc for AHI≥5 (83.8% Se and 86.5% Sp), and 87.1% Acc for AHI≥10 (89.1% Se and 86.0% Sp) using 4% ODI (*ODI4*) in a multiclass task. Nevertheless, in this study, *ODI4* cutoff values for each severity group were optimized and validated using the same population [12].

Clusters of desaturations have been also assessed [13–15]. Brouillette et al. [13] achieved 42.9% Se, 97.8% Sp, and 64.8% Acc (AHI≥1), whereas Velasco et al. [14] reached 86.6% Se, 98.9% Sp, and 93.4% Acc (AHI≥1). However, the latter study only included patients with adenotonsillar hypertrophy, which limits its generalization [14]. Moreover, Van Eyck et al. [15] achieved moderate Acc results when validating the methodologies proposed by Brouillette et al. [13] (58% Se, 88% Sp, and 78% Acc) and Velasco et al. [14] (57% Se, 73.0% Sp, and 68% Acc) in a sample of obese patients

using AHI≥2 e/h as cutoff. Van Eyck et al. [15] also assessed *ODI*3 reaching low diagnostic ability (66% Se, 69% Sp, and 68% Acc). However, these studies only assess the presence of SAHS in children without taking into account of its severity.

Common symptoms and clinical history have been also involved in pediatric SAHS screening tools [16,17]. Chang et al. [16] used *ODI*3 and common symptoms to assess both a discriminative score and a logistic regression (LR) classifier [16]. The LR model achieved 76.6% Acc, whereas the discriminative score reached 60% Se, 86% Sp, and 72% Acc (AHI≥5). Recently, Villa et al. [17] developed a multiclass algorithm using both clinical history and the McGill oximetry score, which was defined by Nixon et al. [4]. This paper reported 57.4% Acc in the multiclass classification task (AHI<1 e/h, 1≤AHI<5 e/h, and AHI≥5 e/h). From their confusion matrix, diagnostic performance metrics were computed: 85.8% Acc for AHI≥1 (91.6% Se and 40.6% Sp), 69.4% Acc for AHI≥5 (40.6% Se and 97.9% Sp), and overall kappa=0.30.

Previous studies assessed the joint use of parameters from SpO_2 and other cardiorespiratory signals [7,9,18]. Cohen & de Chazal [18] applied linear discriminant analysis (LDA) to automatic features computed from SpO_2 and ECG recordings. This model achieved 58.1% Se, 67.0% Sp, and 66.7% Acc (AHI≥5). Gutiérrez-Tobal et al. [9] assessed a LR model built with *ODI3* from SpO_2 and PSD features from AF, achieving average 85.9% Se, 87.4% Sp, and 86.3% Acc (AHI≥3) using a bootstrap validation approach. Garde et al. [7] built a LDA model using features from PRV and SpO_2 recordings. This model was validated using 4-fold cross validation and achieved 88.4% Se, 83.6% Sp, and 85.0% Acc (AHI≥5). In contrast to these studies, our methods reached high diagnostic ability by the exclusive use of single-channel SpO_2 as the only signal involved.

Finally, Álvarez et al. [19] assessed oximetry-based LR models for different AHI

cutoffs (1, 3, and 5 e/h), reaching 85.5% Acc (89.6% Se and 71.5% Sp), 83.4% Acc (82.9% Se and 84.4% Sp), and 82.8% Acc (82.2% Se and 83.6% Sp), respectively. They used bootstrapping to validate results from a small sample size (50 children). While they focused on low severity AHI cutoffs, our current proposed methodology reached high diagnostic performance in the detection of moderate-to-severe pediatric SAHS (AHI ≥ 5, 10 e/h) in an independent test set from a large database (298 children).

PLEASE, DISPLAY TABLE 5 AROUND HERE

Although we present compelling evidence on the usefulness of our method, some limitations have to be taken into account. First, there were less subjects showing an AHI in the ranges 5≤AHI<10 and AHI≥10 e/h in the cohort. This is one possible reason for the slight tendency of the MLP classifier to underestimate for lower SAHS severity groups. A larger sample size, balancing the proportion of subjects among classes, would likely minimize this effect. Another limitation concerns the only detection of moderate (5≤AHI<10 e/h) to severe (AHI≥10 e/h) patients, while avoiding the evaluation of the presence of SAHS in subjects with AHI<5 e/h. However, while moderate to severe subjects are treated regardless the presence of co-morbidities, this group only requires treatment if neurocognitive or developmental deficits are concurrently present, and this latter feature cannot be evaluated by the AHI or any other PSG-derived variable [44]. Furthermore, since our methodology aims at simplifying the detection of pediatric SAHS, it would be also useful to validate this proposal using oximetry recordings obtained in unsupervised children at home. Finally, the only use of MLP for classification is another limitation of our study. In this sense, the application of more advanced machine learning algorithms for classification, such as ensemble learning classifiers, could be potentially useful to enhance our methodology.

In summary, a high diagnostic performance was achieved with a multiclass MLP model built with bispectral features, together with anthropometric variables, *ODI*3, and PSD parameters, in an independent set using a large database of oximetry recordings. Thus, bispectrum contains additional and complementary information to the other methodological approaches when aiming to further characterize desaturation events in the context of SAHS screening in children. Furthermore, this model outperformed previous results obtained by state-of-the-art studies. Therefore, bispectrum could be potentially used as a complementary tool in the analysis of oximetry recordings to help in the screening of moderate-to-severe childhood SAHS.

ACKNOWLEDGEMENTS

This work was supported by 'Consejería de Educación de la Junta de Castilla y León and FEDER' under project VA037U16, by 'Ministerio de Economía y Competitividad' and 'European Regional Development Fund' under projects TEC2014-53196-R and RTC-2015-3446-1, by 'European Commission' and 'European Regional Development Fund' under project 'Análisis y correlación entre el genoma completo y la actividad cerebral para la ayuda en el diagnóstico de la enfermedad de Alzheimer' ('Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020'), and the project 153/2015 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.

CONFLICT OF INTEREST

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

ETHICAL APPROVAL

In all participants, the informed consent to be included in the research was obtained and the Ethical Committee of the University of Chicago Medicine approved the protocol.

AUTHORSHIP RESPONSIBILITY

- The material in this manuscript is original and contains no matter libelous or otherwise unlawful.

- The manuscript represents valid work and that neither this manuscript nor any other with substantially similar content under my authorship has been published or is being considered for publication elsewhere.

- I have participated sufficiently in the work to take public responsibility for all its content.

REFERENCES

[1] C.L. Marcus, L.J. Brooks, K.A. Draper, D. Gozal, A.C. Halbower, J. Jones, M.S. Schechter, S.H. Sheldon, K. Spruyt, S.D. Ward, C. Lehmann, R.N. Shiffman, A.A. of Pediatrics, Diagnosis and management of childhood obstructive sleep apnea syndrome., Pediatrics. 130 (2012) 576–84. doi:10.1542/peds.2012-1671.

[2] A. Kaditis, L. Kheirandish-Gozal, D. Gozal, Pediatric OSAS : Oximetry can provide answers when polysomnography is not available, Sleep Med. Rev. 27 (2016) 96–105. doi:10.1016/j.smrv.2015.05.008.

[3] M.L. Alonso-Álvarez, T. Canet, M. Cubell-Alarco, E. Estivill, E. Fernández, D. Gozal, M.J. Jurado-Luque, M.A. Lluch-Roselló, F. Martínez-Pérez, M. Merino-Andren, G. Pin-Arboledas, N. Roure, F.X. Sanmartí, O. Sans-Capdevila, J. Segarra-Isern, Documento de consenso del síndrome de apneas-hipopneas durante el sueño en niños, Arch Bronconeumol. 47 (2011) 2–18. doi:10.1016/S0300-2896(11)70026-6.

[4] G.M. Nixon, A.S. Kermack, G.M. Davis, J.J. Manoukian, A. Brown, R.T. Brouillette, Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry, Pediatrics. 113 (2004) e19–e25. doi:10.1542/peds.113.1.e19.

[5] E.S. Katz, R.B. Mitchell, C.M.D. Ambrosio, Obstructive Sleep Apnea in Infants, Am. J. Respir. Crit. Care Med. 185 (2012) 805–816. doi:10.1164/rccm.201108-1455CI.

[6] R.B. Shouldice, L.M.O. Brien, C.O. Brien, P. De Chazal, D. Gozal, C. Heneghan, Detection of Obstructive Sleep Apnea in Pediatric Subjects using Surface Lead Electrocardiogram Features, Sleep. 27 (2004) 784–792.

[7] A. Garde, P. Dehkordi, W. Karlen, D. Wensley, J.M. Ansermino, G.A. Dumont, Development of a screening tool for sleep disordered breathing in children using the phone oximeterTM, PLoS One. 9 (2014). doi:10.1371/journal.pone.0112959.

[8] J. Park, H. Lee, J. Lee, T. Penzel, K. Kesper, I. Pinnow, Evaluation of cardiac modulation in children in response to apnea / hypopnea using the Phone Oximeter TM, Physiol. Meas. 37 (2016) 187–202. doi:10.1088/0967-3334/37/2/187.

[9] G.C. Gutiérrez-Tobal, M.L. Alonso-Álvarez, D. Álvarez, F. del Campo, J. Terán-Santos, R. Hornero, Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home, Biomed. Signal Process. Control. 18 (2015) 401–407. doi:10.1016/j.bspc.2015.02.014.

[10] E. Gil, R. Bailón, J.M. Vergara, P. Laguna, PTT Variability for Discrimination of Sleep Apnea Related Decreases in the Amplitude Fluctuations of PPG Signal in Children, IEEE Trans. Biomed. Eng. 57 (2010) 1079–1088. doi:10.1109/TBME.2009.2037734.

[11] V.G. Kirk, S.G. Bohn, W.W. Flemons, J.E. Remmers, Comparison of Home Oximetry Monitoring With Laboratory Polysomnography in Children, CHEST J. 124 (2003) 1702–1708. doi:10.1378/chest.124.5.1702.

[12] C. Tsai, C. Kang, M. Su, H. Lin, E. Huang, C. Chen, J. Hung, C. Niu, D. Liao, H. Yu, Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children, Int. J. Pediatr. Otorhinolaryngol. 77 (2013) 1286–1290. doi:10.1016/j.ijporl.2013.05.011.

[13] R.T. Brouillette, A. Morielli, A. Leimanis, K.A. Waters, F.M. Ducharme, R.T.

Brouillette, A. Morielli, A. Leimanis, K.A. Waters, R. Luciano, F.M. Ducharme, Nocturnal pulse oximetry as an abbreviated testing modality for pediatric obstructive sleep apnea, Pediatrics. 105 (2000) 405–412. doi:10.1542/peds.105.2.405.

[14] C.T. Velasco, M.D. Suárez, J.M. Figueroa, M.D. Turienzo, F.L. Len, E. Mansilla, Pulse oximetry recording in children with adenotonsillar hypertrophy : usefulness in the diagnosis of obstructive sleep apnea syndrome, Arch. Argent. Pediatr. 111 (2013) 196–201. doi:10.1590/S0325-00752013000300004.

[15] A. Van Eyck, C. Lambrechts, L. Vanheeswijck, The role of nocturnal pulse oximetry in the screening for obstructive sleep apnea in obese children and adolescents, Sleep Med. 16 (2015) 1409–1412. doi:10.1016/j.sleep.2015.07.023.

[16] L. Chang, J. Wu, L. Cao, Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea, Int. J. Pediatr. Otorhinolaryngol. 77 (2013) 365–371. doi:10.1016/j.ijporl.2012.11.028.

[17] M.P. Villa, N. Pietropaoli, M.C. Supino, O. Vitelli, J. Rabasco, M. Evangelisti, M. Del Pozzo, A.G. Kaditis, Diagnosis of Pediatric Obstructive Sleep Apnea Syndrome in Settings With Limited Resources, JAMA Otolaryngol. Neck Surg. 141 (2015) 990–996. doi:10.1001/jamaoto.2015.2354.

[18] G. Cohen, P. De Chazal, Automated detection of sleep apnea in infants : A multi-modal approach, Comput. Biol. Med. 63 (2015) 118–123. doi:10.1016/j.compbiomed.2015.05.007.

[19] D. Álvarez, M.L. Alonso-Álvarez, G.C. Gutiérrez-Tobal, A. Crespo, L. Kheirandish-Gozal, D. Gozal, J. Terán-Santos, F. Del Campo, Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal

Oximetry : An Alternative to Respiratory Polygraphy in Unattended Settings, J Clin Sleep Med. 13 (2017) 7–11. doi:10.5664/jcsm.6586.

[20] R. Palaniappan, K. Sundaraj, S. Sundaraj, N. Huliraj, S.S. Revadi, A novel approach to detect respiratory phases from pulmonary acoustic signals using normalised power spectral density and fuzzy inference system, Clin. Respir. J. 10 (2016) 486–494. doi:10.1111/crj.12250.

[21] R. Palaniappan, K. Sundaraj, S. Sundaraj, Adaptive neuro-fuzzy inference system for breath phase detection and breath cycle segmentation, Comput. Methods Programs Biomed. 145 (2017) 67–72. doi:10.1016/j.cmpb.2017.04.013.

[22] N. Netzar, A.H. Eliasson, C. Netzar, D.A. Kristo, Overnight Pulse Oximetry for Sleep- Disordered Breathing in Adults, CHEST J. 120 (2001) 625–633. doi:10.1378/chest.120.2.625.

[23] J.V. Marcos, R. Hornero, D. Álvarez, F. del Campo, C. Zamarrón, M. López, Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry, Comput. Methods Programs Biomed. 92 (2008) 79–89. doi:10.1016/j.cmpb.2008.05.006.

[24] D. Álvarez, R. Hornero, J.Ví. Marcos, N. Wessel, T. Penzel, M. Glos, F. del Campo, Assessment of Feature Selection and Classification Approaches To Enhance Information From Overnight Oximetry in the Context of Apnea Diagnosis, Int. J. Neural Syst. 23 (2013) 1350020. doi:10.1142/S0129065713500202.

[25] K. Chua, V. Chandran, U.R. Acharya, C. Min, Application of higher order statistics/spectra in biomedical signals—A review, Med. Eng. Phys. 32 (2010) 679–689. doi:10.1016/j.medengphy.2010.04.009.

[26] B.H. Taha, J.A. Dempsey, S.M. Weber, M.S. Badr, J.B. Skatrud, T.B.. Young, A.J. Jacques, K.C. Seow, Automated detection and classification of sleep-disordered breathing from conventional polysomnography data, Sleep. 20 (1997) 991–1001. doi:10.1093/sleep/20.11.991.

[27] L. Yu, H. Liu, Efficient Feature Selection via Analysis of Relevance and Redundancy, J. Mach. Learn. Res. 5 (2004) 1205–1224. doi:10.1145/1014052.1014149.

[28] U.J. Magalang, J. Dmochowski, Prediction of the Apnea-Hypopnea Index From Overnight Pulse Oximetry, CHEST J. 124 (2003) 1694–1701. doi:10.1378/chest.124.5.1694.

[29] G.D. Church, The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients, Curr. Probl. Pediatr. Adolesc. Health Care. 42 (2012) 22–25. doi:10.1016/j.cppeds.2011.10.001.

[30] H.-L. Tan, D. Gozal, H.M. Ramirez, H.P.R. Bandla, L. Kheirandish-Gozal, Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea, Sleep. 37 (2014) 255–260. doi:10.5665/sleep.3392.

[31] I. Guyon, An Introduction to Variable and Feature Selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[32] M. Hall, I. Witten, E. Frank, Data mining: Practical machine learning tools and techniques, 2011.

[33] P.D. Welch, The Use of Fast Fourier Transform fot he Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms, IEEE Trans. Audio Electroacoust. 15 (1967) 70–73.

doi:10.1109/TAU.1967.1161901.

[34] J. Poza, R. Hornero, D. Abásolo, A. Fernández, Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease, Med. Eng. Phys. 29 (2007) 1073–1083. doi:10.1016/j.medengphy.2006.11.006.

[35] M. Blanco-Velasco, F. Cruz-Roldan, J.I. Godino-Llorente, K.E. Barner, Nonlinear Trend Estimation of the Ventricular Repolarization Segment for T-Wave Alternans Detection, IEEE Trans. Biomed. Eng. 57 (2010) 2402–2412. doi:10.1109/TBME.2010.2048109.

[36] S. Zhou, J.Q. Gan, F. Sepulveda, Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface, Inf. Sci. (Ny). 178 (2008) 1629–1640. doi:10.1016/j.ins.2007.11.012.

[37] K. Chua Chua, V. Chandran, U.R. Acharya, C. Min Lin, Cardiac state diagnosis using higher order spectra of heart rate variability Cardiac state diagnosis using higher order spectra of heart rate variability, J. Med. Eng. Technol. 32 (2009) 145–155. doi:10.1080/03091900601050862.

[38] D.P. Doane, Aesthetic frequency classifications, Am. Stat. 30 (1976) 181–183. doi:10.1080/00031305.1976.10479172.

[39] G.C. Gutiérrez-Tobal, D. Álvarez, J.V. Marcos, F. del Campo, R. Hornero, Pattern recognition in airflow recordings to assist in the sleep apnoea-hypopnoea syndrome diagnosis, Med. Biol. Eng. Comput. 51 (2013) 1367–1380. doi:10.1007/s11517-013-1109-7.

[40] B. Efron, R.J. Tibshirani, Introduction to the Bootstrap, CRC press., 1994.

[41] C.M. Bishop, Neural Networks for Pattern Recognition, 1995.

[42] I. Nabney, NETLAB: algorithms for pattern recognition, Springer Science & Business Media, 2002.

[43] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46. doi:10.1177/001316446002000104.

[44] H.-L. Tan, M.L. Alonso Alvarez, M. Tsaoussoglou, S. Weber, A.G. Kaditis, When and why to treat the child who snores?, Pediatr. Pulmonol. 412 (2016) 399–412. doi:10.1002/ppul.23658.

TABLES

Table 1. Clinical and demographic data of the population under study. Data are presented as median [interquartile range], n or %.

Characteristics	All	AHI<5	5≤AHI<10	AHI≥10
All subjects				
Subjects (n)	298	164	56	78
Age (years)	6.0[4.0,9.0]	7.0[5.0,10.0]	5.0[3.0,8.0]	5.5[3.0,9.0]
Males (%)	55.7	55.5	57.1	55.1
BMI(kg/m^2)	18.4 [16.3,23.0]	18.2 [16.3,22.3]	18.1 [16.3,22.6]	19.1 [16.5,25.7]
Time (h)	7.8 [7.3,8.4]	7.9 [7.3,8.4]	7.9 [7.3,8.4]	7.7 [7.3,8.3]
AHI(e/h)	-	1.9 [1,3.5]	7.0 [5.9,8.5]	17.7 [11.7,27.3]
Feature optimization set (25%)				
Subjects (n)	74	41	14	19
Age (years)	6.0 [3.0,10.0]	7.0 [4.8,11.0]	4.0 [2.0,8.0]	6.0 [3.0,9.0]
Males (%)	58.1	52.6	71.4	56.1
BMI(kg/m^2)	18.2 [16.3,22.5]	18.2 [16.8,22.6]	19.0 [16.2,23.0]	17.7 [15.4,22.3]
Time (h)	8.0 [7.5,8.3]	8.0 [7.5,8.3]	7.9 [7.6,8.7]	7.7 [7.2,8.2]
AHI(e/h)	-	1.7 [1.0,3.4]	6.9 [5.9,8.1]	17.9 [11.5,26.4]
Training set (50%)				
Subjects (n)	149	82	28	39
Age (years)	7.0[4.0,9.3]	7.0[5.0,10.0]	6.0[3.0,8.0]	5.0[3.0,9.8]
Males (%)	51.7	46.2	46.4	56.1
BMI(kg/m^2)	18.5 [16.4,23.2]	18.5 [16.3,21.7]	17.5 [16.1,21.1]	20.1 [17.2,27.8]
Time (h)	7.7 [7.2,8.4]	7.8 [7.2,8.4]	7.9 [7.3,8.4]	7.7 [7.3,8.4]
AHI(e/h)	-	2.0 [1.0,3.6]	7.0 [5.9,8.5]	18.2 [12.0,27.3]
Test set (25%)				
Subjects (n)	75	41	14	20
Age (years)	6.0 [4.0,8.0]	7.0 [5.0,8.5]	5.0 [4.0,6.0]	5.5 [3.0,8.5]
Males (%)	61.3	75.0	64.3	53.7
BMI(kg/m^2)	18.1 [16.0,23.6]	18.0 [15.6,23.7]	18.5 [16.7,23.6]	18.5 [16.2,24.4]
Time (h)	7.8 [7.3,8.4]	7.8 [7.2,8.3]	7.9 [7.3,8.4]	7.9 [7.3,8.4]
AHI(e/h)	-	1.8 [0.9,3.2]	7.0 [5.9,8.7]	17.0 [11.8,30.2]

BMI: Body Mass Index; AHI: Apnea Hypopnea Index

Table 2. Feature values for the SAHS severity groups (median [interquartile range]) in the feature optimization set.

Features	AHI<5	5≤AHI<10	AHI≥10	*p*-value*
Age	7.0 [4.8,11.0]	4.0 [2.0,8.0]	6.0 [3.0,9.0]	0.24
Sex	-	-	-	0.52
BMI **(10^1)**	1.82 [1.68,2.26]	1.90 [1.62,2.30]	1.77 [1.54,2.23]	0.76
ODI3	2.08 [0.77,3.93]	5.82 [3.79,9.28]	8.72 [7.23,19.65]	<< 0.05
M1f	2.95 [1.93,4.23]	5.67 [5.10,8.58]	14.73 [7.51,28.45]	<< 0.05
M2f **(10^1)**	0.18 [0.05,0.32]	1.11 [0.56,1.54]	4.07 [1.34,14.39]	<< 0.05
M3f	0.48 [0.29,0.75]	0.68 [0.33,1.04]	0.54 [0.32,0.83]	0.54
M4f	2.10 [1.79,2.73]	2.33 [1.87,3.06]	2.20 [1.60,2.70]	0.65
P_R	0.31 [0.25,0.35]	0.24 [0.21,0.34]	0.34 [0.30,0.39]	0.05
MA **(10^1)**	0.60 [0.33,0.87]	1.27 [1.00,1.54]	2.92 [1.68,4.59]	<< 0.05
mA	1.21 [0.84,2.12]	2.30 [1.91,2.88]	3.46 [2.56,8.59]	<< 0.05
SE	4.33 [4.26,4.37]	4.30 [4.24,4.32]	4.20 [4.12,4.30]	<< 0.05
Mb	0.17 [0.11,0.20]	0.28 [0.23,0.32]	0.42 [0.25,0.56]	<< 0.05
MB1 **(10^{-1})**	0.05 [0.02,0.21]	0.23 [0.09,0.63]	0.49 [0.17,1.22]	<< 0.05
H1 **(10^8)**	-5.78 [-5.90,-5.57]	-5.59 [-5.73,-5.45]	-5.40 [-5.58,-5.28]	<< 0.05
H2 **(-10^5)**	-1.38 [-1.41,-1.33]	-1.32 [-1.36,-1.29]	-1.28 [-1.32,-1.25]	<< 0.05
H3 **(-10^8)**	-2.96 [-3.02,-2.86]	-2.87 [-2.93,-2.79]	-2.77 [-2.86,-2.71]	<< 0.05
BE1	8.51 [7.88,9.43]	8.17 [7.71,8.71]	8.60 [8.11,8.79]	0.34
BE2	6.08 [4.88,6.62]	5.26 [4.43,5.58]	6.70 [5.68,7.07]	0.13
PE	2.14 [2.08,2.15]	2.12 [2.05,2.14]	2.11 [2.08,2.13]	0.64
meanPa **(10^{-2})**	-1.60 [-3.19,0.81]	0.89 [-0.47,2.21]	0.92 [-1.96,3.20]	0.29
varPa	0.38 [0.26,0.43]	0.34 [0.22,0.37]	0.33 [0.22,0.44]	0.87

* *P*-values obtained after Bonferroni correction

Table 3. Confusion matrices of the MLP models in the test set. Regarding the model MLP_{nobis} average Acc=69.3% and kappa=0.45, whereas for model MLP_{bis} average Acc=76.0% and kappa=0.56.

		Estimated					
		MLP_{nobis}			**MLP_{bis}**		
		AHI<5	**5≤AHI<10**	**AHI≥10**	**AHI<5**	**5≤AHI<10**	**AHI≥10**
Actual	**AHI<5**	**37**	3	1	**40**	0	1
	5≤AHI<10	6	**6**	2	7	**5**	2
	AHI≥10	7	4	**9**	6	2	**12**

Table 4. Diagnostic ability of the MLP models in the test set for AHI cutoffs= 5 e/h and 10 e/h.

			AHI cutoff=5 e/h				
Features	**Se**	**Sp**	**PPV**	**NPV**	**LR+**	**LR-**	**Acc**
MLP_{nobis}	61.8	90.2	84.0	74.0	6.33	0.42	77.3
MLP_{bis}	61.8	97.6	95.5	75.5	25.32	0.39	81.3
			AHI cutoff=10 e/h				
Features	**Se**	**Sp**	**PPV**	**NPV**	**LR+**	**LR-**	**Acc**
MLP_{nobis}	45.0	94.5	75.0	82.5	8.25	0.58	81.3
MLP_{bis}	60.0	94.5	80.0	86.7	11.00	0.42	85.3

Table 5. Summary of previous relevant published studies in the context of automated analysis of SpO_2 recordings to assist in the diagnosis of pediatric SAHS.

Studies	Subjects (n)	Signal	AHI cutoff	Methods	Validation	Se (%)	Sp (%)	Acc (%)
Kirk *et al.* [11]	58	SpO_2	5	*ODI3*	Direct validation**	67	60	64*
Tsai *et al.* [12]	148	SpO_2	1	*ODI4*	No	77.7	88.9	79.0*
			5			83.8	86.5	85.1*
			10			89.1	86.0	87.1*
Brouillette *et al.* [13]	349	SpO_2	1	Clusters of desaturations	Direct validation**	42.9	97.8	64.7
Velasco *et al.* [14]	167	SpO_2	1	Clusters of desaturations	Direct validation**	86.6	98.9	93.4*
Van Eyck *et al.* [15]	130	SpO_2	2	*ODI3*, Brouillete *et al.* [13], and Velasco *et al.* [14]	Train-test for ODI3	57	73	68*
						58	88	78*
						66	69	68*
Chang *et al.* [16]	141	SpO_2	5	*ODI3* and sympthoms	Direct validation**	60	86	72*
Pia-Villa *et al.* [17]	268	SpO_2	1	McGill oximetric score and clinical history	Direct validation**	91.6*	40.6*	85.8*
			5			40.6*	97.9*	69.4*
Cohen & de Chazal [12]	288	ECG and SpO_2	Event detection	Statistical parameters, classical indices, and PSD	Loocv	58.1	67.0	66.7
Gutiérrez-Tobal *et al.* [12]	50	FA and SpO_2	3	PSD (AF) and ODI3 (SpO_2)	Bootstrap 0.632	85.9	87.4	86.3
Garde *et al.* [7]	146	PRV and SpO_2	5	Statistical parameters, nonlinear features, classical indices, and PSD	4-fold cross validation	88.4	83.6	85.0
Álvarez et al. [19]	50	SpO_2	1	Statistical parameters, nonlinear features, classical indices, and PSD	Bootstrap 0.632	89.6	71.5	85.5
			3			82.9	84.4	83.4
			5			82.2	83.6	82.8
Our proposal	298	SpO_2	5	Bispectrum, PSD, *ODI3*, anthropometric variables	Feature optimization-training-test	61.8	97.6	81.3
			10			60.0	94.5	85.3

loocv: leave-one-out cross validation; ECG: Electrocardiogram; AF. Airflow; PRV: Pulse rate variability.

* computed from reported data,

** direct validation of a scoring criteria against AHI from PSG.

FIGURE LEGENDS

Figure 1. Averaged magnitude of the bispectrum for the three SAHS severity groups: (a) AHI<5 e/h, (b) 5≤AHI<10 e/h, and (c) AHI≥10 e/h in the feature optimization set.