SLEEP BREATHING PHYSIOLOGY AND DISORDERS • ORIGINAL ARTICLE

Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea

Andrea Crespo[1,2] · Daniel Álvarez[1,2] · Leila Kheirandish-Gozal[3] · Gonzalo C. Gutiérrez-Tobal[2] ·
Ana Cerezo-Hernández[1] · David Gozal[3] · Roberto Hornero[2] · Félix del Campo[1,2]

Abstract
Purpose A variety of statistical models based on overnight oximetry has been proposed to simplify the detection of children with suspected obstructive sleep apnea syndrome (OSAS). Despite the usefulness reported, additional thorough comparative analyses are required. This study was aimed at assessing common binary classification models from oximetry for the detection of childhood OSAS.
Methods Overnight oximetry recordings from 176 children referred for clinical suspicion of OSAS were acquired during in-lab polysomnography. Several training and test datasets were randomly composed by means of bootstrapping for model optimization and independent validation. For every child, blood oxygen saturation (SpO_2) was parameterized by means of 17 features. Fast correlation-based filter (FCBF) was applied to search for the optimum features. The discriminatory power of three statistical pattern recognition algorithms was assessed: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR). The performance of each automated model was evaluated for the three common diagnostic polysomnographic cutoffs in pediatric OSAS: 1, 3, and 5 events/h.
Results Best screening performances emerged using the 1 event/h cutoff for mild-to-severe childhood OSAS. LR achieved 84.3% accuracy (95% CI 76.8–91.5%) and 0.89 AUC (95% CI 0.83–0.94), while QDA reached 96.5% PPV (95% CI 90.3–100%) and 0.91 AUC (95% CI 0.85–0.96%). Moreover, LR and QDA reached diagnostic accuracies of 82.7% (95% CI 75.0–89.6%) and 82.1% (95% CI 73.8–89.5%) for a cutoff of 5 events/h, respectively.
Conclusions Automated analysis of overnight oximetry may be used to develop reliable as well as accurate screening tools for childhood OSAS.

Keywords Pediatric obstructive sleep apnea syndrome · Overnight oximetry · Unattended oximetry · Automated signal processing · Pattern recognition

✉ Félix del Campo
    fsas@telefonica.net

1   Servicio de Neumología, Hospital Universitario Río Hortega de
    Valladolid, c/ Dulzaina 2, 47012 Valladolid, Spain

2   Biomedical Engineering Group, University of Valladolid, Paseo de
    Belén 15, 47011 Valladolid, Spain

3   Section of Sleep Medicine, Department of Pediatrics, Pritzker School
    of Medicine, Biological Sciences Division, The University of
    Chicago, Chicago, IL 60637, USA

Introduction

Childhood obstructive sleep apnea syndrome (OSAS) is a prevalent yet relatively under-diagnosed condition [1]. According to the American Academy of Pediatrics, OSAS affects 1 to 5% of children in the general pediatric population [2]. Untreated OSAS has been associated with adverse consequences affecting multiple organ systems in infants and young children, reducing overall health and quality of life while increasing health care use and associated costs [1, 2]. Thus, it is important to screen for the presence of the disease in children showing symptoms in order

to perform early and timely diagnosis and treatment, an issue that has been recently re-emphasized by the American Academy of Pediatrics [2]. In-laboratory polysomnography (PSG) is the gold standard test for OSAS [2, 3]. However, the availability and accessibility of specialized sleep laboratories is too limited to match clinical needs [4, 5]. In addition, PSGs are onerous and also intrusive and relatively inconvenient for children and caretakers, which hampers their widespread implementation [6].

The aforementioned drawbacks have led to an extensive search for screening tools over the last decade [1, 2, 7, 8]. Overnight oximetry has emerged as a potentially useful screening approach due to their simplicity and suitability in children [3, 9–14]. In this context, the oxygen desaturation index (ODI) is the most widely used oximetry marker [5, 10, 12]. Nevertheless, besides the number of desaturations, there is a lot of additional information present in the oximetric recordings. In this regard, automated signal processing and pattern recognition techniques are able to optimize the diagnostic capability of oximetry leading to a more reliable and effective approach [9, 14]. A recent multi-center study evaluating more than 4000 overnight oximetry recordings indicated that machine-learning approaches enable accurate identification of children with OSAS [15].

In the framework of automated pattern recognition, statistical classifiers model the characteristics of the problem under study, leading to automated tools able to assist physicians in their daily practice. Particularly, linear discriminant analysis (LDA) [9, 16–18] and logistic regression (LR) [10, 14, 19, 20] have been previously used with relative success for binary classification of children suspected of suffering from OSAS. Similarly, quadratic discriminant analysis (QDA) also provides a suitable alternative, but has been marginally assessed in this context [21]. Nevertheless, a thorough review reveals that no previous studies have extensively assessed the performance of these classification approaches when using different cutoff criteria for the diagnosis of childhood OSAS. We hypothesized that a comprehensive comparative analysis could provide essential information that would enable widespread use of automated processing of overnight oximetric recordings as a screening tool for pediatric OSAS. Therefore, the aim of this study was to design and assess several statistical binary classifiers based on overnight oximetry using different clinically used cutoffs for pediatric OSAS.

Materials and methods

Participants

A total of 176 consecutive otherwise healthy children (97 boys and 79 girls) ranging from 1 to 13 years of age composed our dataset. All children were referred by pediatricians to the Pediatric Sleep Unit at the University of Chicago Medicine Comer Children's Hospital (Chicago, IL, USA) for evaluation of habitual snoring and suspected OSAS. Enrolment criteria included habitual snoring and/or witnessed breathing pauses during sleep as reported by their parents or caretakers. The Institution's Ethical Review Committee approved the study protocol (#IRB14-1241) and informed consent was obtained from all caretakers prior to the enrolment.

Sleep studies

Children's sleep was monitored using a digital polysomnography system (Polysmith; Nihon Kohden America Inc., CA, USA) [22]. Blood oxygen saturation (SpO_2) recordings using pulse oximetry were acquired during overnight PSG (sampling frequency 25 Hz) and subsequently exported for offline processing.

All PSGs were manually scored by trained sleep technologists to derive the apnea-hypopnea index (AHI). The 2012 American Academy of Sleep Medicine scoring rules for children were used to quantify sleep and cardiorespiratory events [23]. In order to extensively assess the screening ability of the proposed classification models, we used several AHI cutoff values routinely used for establishing the diagnosis of OSAS [2, 8, 14], namely, 1, 3, and 5 events/h. Table 1 summarizes the main characteristics of the dataset.

Automated processing of oximetric recordings

Three signal processing stages were implemented: feature extraction, feature selection, and pattern recognition.

Feature extraction

The aim of this stage was to exhaustively characterize oximetry dynamics in every child. To conduct this critical initial step, several oximetric indices were computed:

i. Desaturations due to apneic events modify the shape and thus the characteristics of the oximetric data distribution. Conventional first- to fourth-order statistical moments, i.e., mean ($M1t$), variance ($M2t$), skewness ($M3t$), and kurtosis ($M4t$), are suitable as well as simple measures able to parameterize these changes in the histogram of SpO_2 amplitudes [14, 24].

ii. The power spectral density (PSD) function is able to reflect the occurrence of (pseudo)periodic desaturations at each frequency [9, 14]. The PSD of each SpO_2 signal was parameterized by means of first- to fourth-order 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 [14]. Similarly, the total signal power (P_T) as well as the peak amplitude (PA) and relative power (P_R) in a common apnea frequency band for children (0.021–0.040 Hz) were

Table 1 Demographic and clinical characteristics of the cohort using three different AHI cutoff values

	AHI ≥ 1 event/h		AHI ≥ 3 events/h		AHI ≥ 5 events/h	
	OSAS negative	OSAS positive	OSAS negative	OSAS positive	OSAS negative	OSAS positive
No. (%)	30 (17.1)	146 (82.9)	79 (44.9%)	97 (55.1%)	105 (59.7%)	71 (40.3%)
No. males (%)	17 (56.7%)	79 (54.1%)	45 (57.0%)	51 (52.6%)	57 (54.3%)	39 (54.9%)
Age (years)	8.0 [5.0]	7.0 [5.0]	7.0 [5.5]	6.0 [6.0]	7.0 [5.3]	6.0 [6.0]
BMI (kg/m^2)	17.7 [6.6]	18.3 [7.5]	18.1 [6.9]	18.4 [7.8]	18.3 [6.4]	18.1 [8.1]
AHI (events/h)	0.6 [0.4]	4.7 [8.9]	1.3 [1.2]	9.3 [12.5]	1.7 [2.1]	11.2 [12.1]

Data are provided as median and interquartile range and *n* (%)

AHI apnea-hypopnea index, *OSAS* obstructive sleep apnea syndrome, *BMI* body mass index

computed to measure changes in the PSD function due to desaturations [14, 24, 25].

iii. In the framework of long-term overnight oximetric recordings, non-linear measures have widely demonstrated to provide complementary information to conventional spectral techniques. Particularly, sample entropy (*SampEn*), central tendency measure (*CTM*), and Lempel-Ziv complexity (*LZC*) were computed due to their ability to quantify slight changes in irregularity, variability, and complexity of nocturnal SpO$_2$ [14, 26].

iv. Finally, the conventional 3% oxygen desaturation index (ODI3) was automatically scored and included in the initial feature space in order to quantify the number of drops per hour of recording greater than or equal to 3% from baseline [12].

Feature selection

For each child, all 17 features derived from the oximetry signal were jointly used to characterize the presence of pediatric OSAS. This initial feature set gathers valuable information able to detect the disease. Nevertheless, an improved as well as reduced feature subset can be derived by applying a feature selection algorithm. Previous works have shown that feature selection and dimensionality reduction algorithms enhance the prediction ability of oximetric features in the context of OSAS diagnosis [24, 26, 27]. In this study, the fast correlation-based filter (FCBF) was applied. FCBF is a filter methodology for feature selection independent of the classifier used in the subsequent classification stage [28]. FCBF automatically inspects the original feature space to search for the most relevant as well as non-redundant variables. In order to guide this search, the symmetrical uncertainty (SU) is used [28]. SU is a measure of predictability between two variables based on the information shared between them; i.e., it quantifies the amount of knowledge we can infer about the first variable using the information present in the second one. The higher the SU is, the higher the predictability is, i.e., the information shared between variables. In the context of pediatric OSAS, we considered that a variable is more relevant for characterizing the disease if it shares as

much information as possible with the AHI, which is the clinical index used by physicians to establish the presence of the disease and characterize its severity.

FCBF implements two consecutive filtering stages [28]: firstly, a relevance analysis and secondly, a redundancy-based variable selection. In order to measure the degree of relevancy of every variable from the original feature space, the association between each oximetric feature and the severity of the disease was estimated: the SU_i between each oximetric feature (X_i) and the actual AHI from standard PSG (Y) was computed. Subsequently, all the variables from the original feature space are ranked in decreasing order of SU_i, i.e., in decreasing order of relevancy for the problem under study. In this research, all the oximetric features were considered potentially relevant. Finally, the redundancy filtering stage is implemented as follows: (i) $SU_{i,j}$ is computed between each pair of features according to the previous relevancy-based ranking so that feature i is ranked higher (more relevant) than feature j, and (ii) if $SU_{i,j} \geq SU_i$, then feature j is removed because it is highly correlated with feature i and its valuable information linked with OSAS is smaller and can be derived from feature i. Therefore, the feature subset built by means of FCBF was composed of the most relevant and non-redundant variables from the proposed oximetric feature space.

In order to increase the generalizability of our results, a bootstrap procedure was embedded within the feature selection stage [29]. Accordingly, several datasets were composed by means of resampling with replacement, so that the FCBF method was repeated 1000 times. At each iteration of the proposed bootstrap approach, the variables automatically selected by FCBF were saved. Finally, only those features selected a number of times significantly higher (> 90%) than the remaining ones composed our optimum feature subset from oximetry.

Feature classification

In this comparative study, the performance of LDA, LR, and QDA binary classifiers is extensively assessed in the context of pediatric OSAS diagnosis. In order to decrease the model complexity and improve the performance, these widely known statistical classifiers assume that the probability

density function of the classes under study (OSAS-negative vs. OSAS-positive) is known a priori [30]:

i.   LDA assumes both data normality (Gaussian or normal distribution) and homoscedasticity (equal variances) to model each class-conditional density function for an input feature pattern and each class. Under these assumptions, a linear decision threshold minimizes the classification error.

ii. QDA does not presume homoscedasticity. Then, the Bayes classification rule that minimizes the classification error function establishes a quadratic decision boundary between classes in the feature space.

iii. In a more general context, LR does not assume a priori neither normality nor homoscedasticity of the probability distribution of variables involved in the model. A binary LR classifier models the probability density function as a Bernoulli distribution.

Statistical analysis

MATLAB R2015a (The MathWorks Inc., Natick, MA, USA) was used to implement the proposed signal processing methods, as well as to perform both statistical and performance analyses. Descriptive analysis of variables was presented in terms of their median and interquartile range. The nonparametric Mann-Whitney U test was applied to search for statistical differences between groups, and $p < 0.05$ was considered statistically significant. A separate analysis was carried out to assess statistical differences between the three classification approaches under study. The Bonferroni correction was applied to manage multiple comparisons, and a p value < 0.05 was regarded as significant.

The common bootstrap 0.632 was applied for performance assessment [14, 19, 31]. Given an original dataset of size N, resampling with replacement is applied to build m_i ($1 \le i \le M$) bootstrap replicates of size N. For each replicate, every oximetric pattern from the original dataset may be selected several times with equal probability (uniform distribution). Therefore, all bootstrap replicates m_i will likely contain repeated instances, whereas a number of cases from the original dataset are not selected. According to bootstrap 0.632, the replicates m_i are used for training, whereas instances not included in m_i are used for validation. At each iteration i ($1 \le i \le M$), a performance metric is obtained as the contribution of both the training replicate m_i (0.632 times the metric in the training dataset) and its corresponding validation set (0.368 times the metric in the test dataset) [31]. Finally, each metric is estimated as the average of the M bootstrap values. The following diagnostic performance metrics were computed: sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR−), accuracy (Acc), and

area under the receiver operating characteristics (ROC) curve (AUC). The AHI from in-lab PSG was the gold standard for evaluation. The 95% confidence interval (95% CI) was provided per each performance metric. In order to obtain a proper estimation of the 95% CI, the number of bootstrap replicates was set to $M = 1000$ [31].

Results

Table 2 summarizes the main polysomnographic and oximetric characteristics of the cohort based on the proposed PSG-derived AHI cutoff values used for establishing the presence of OSAS. Table 3 shows the optimum features automatically selected from FCBF using the proposed bootstrap procedure. For each cutoff, LDA, QDA, and LR models were optimized in each training bootstrap replicate and subsequently assessed in the remaining validation instances according to bootstrap 0.632. Figure 1 shows the influence of each variable in every model for each diagnostic AHI cutoff. Axes represent the dimensions of the optimum feature space for each cutoff, whereas vertices of each polygon are proportional to the absolute value of each coefficient in the optimized models, i.e., the overall influence of each oximetric variable.

Table 4 summarizes the performance metrics of the models involved in this comparative study for all the AHI cutoffs. Using an AHI ≥ 1 event/h for OSAS, LR achieved the highest diagnostic accuracy of the three models, with 84.3% Acc and a highly unbalanced sensitivity-specificity pair (93.9% Se, 37.8% Sp). Similarly, for an AHI cutoff of 3 events/h, the LR model achieved 77.7% Acc (74.6% Se, 81.7% Sp). Finally, using the AHI cutoff of 5 events/h, both LR and QDA reached similar high accuracy, with 82.7% Acc (70.0% Se, 91.4% Sp) and 82.1% Acc (62.3% Se, 95.5% Sp), respectively. It is noteworthy that QDA achieved significantly high PPV and LR+ values regardless of the cutoff point for OSAS. Regarding the performance of the conventional ODI3, our analyses showed a significant imbalance between sensitivity and specificity using fixed cutoffs for the ODI3 (≥ 1, 3, and 5 desaturations per hour of recording) for all the clinical thresholds under study. In this regard, the higher specificity commonly reported in the literature was also confirmed.

Figure 2 depicts the average ROC curves for all the models under study from the bootstrap approach. We can observe that differences between the different approaches in this comparative study decrease as the cutoff for positive pediatric OSAS increases. It is important to note that the QDA model reached the highest AUC using the most restrictive clinical threshold for mild OSAS (AHI ≥ 1 event/h), achieving an area of 0.91. For a cutoff of 3 events/h, both LR and ODI3 achieved an AUC equal to 0.88. Considering a cutoff of 5 events/h for moderate-to-severe childhood OSAS, LR reached the highest AUC (0.89), similar to QDA (0.88) and ODI3 (0.87). A

Table 2 Summary of the main polysomnographic and oximetric variables for the whole cohort using three defined AHI cutoffs for positive OSAS designation

	Cutoff AHI ≥ 1 event/h			Cutoff AHI ≥ 3 event/h			Cutoff AHI ≥ 5 event/h		
	OSAS negative	OSAS positive	p value[a]	OSAS negative	OSAS positive	p value[a]	OSAS negative	OSAS positive	p value[a]
TRT (min)	464.8 [79.8]	481.8 [60.8]	0.194	472.5 [69.5]	485.0 [62.0]	0.118	479.0 [68.0]	485.0 [60.3]	0.219
TST (min)	414.8 [80.1]	423.0 [70.0]	0.218	413.5 [82.5]	429.5 [69.5]	< 0.05	416.0 [74.5]	427.5 [76.3]	0.158
WASO (min)	18.3 [36.0]	18.8 [34.6]	0.430	17.5 [35.5]	22.0 [34.5]	0.323	17.5 [33.5]	20.5 [41.5]	0.358
Sleep eff. (%)	87.4 [11.6]	88.8 [8.7]	0.647	87.7 [9.7]	89.6 [8.9]	0.165	88.1 [9.8]	88.9 [9.1]	0.388
Sleep onset lat. (min)	17.5 [32.8]	18.8 [28.9]	0.381	21.0 [30.3]	13.5 [27.0]	0.122	21.0 [31.0]	12.5 [27.3]	0.166
REM onset lat. (min)	109.5 [69.8]	127.5 [79.0]	0.122	124.0 [84.5]	123.5 [79.0]	0.774	124.0 [88.0]	123.5 [75.0]	0.724
Awakenings (n)	2.5 [2.8]	3.0 [4.0]	0.780	3.0 [4.0]	3.0 [4.0]	0.432	3.0 [4.0]	3.0 [4.0]	0.257
Supine (min)	46.3 [56.6]	43.3 [45.0]	0.703	42.6 [45.6]	45.6 [49.3]	0.401	39.4 [50.5]	52.3 [46.4]	0.113
Prone (min)	3.3 [18.7]	3.9 [25.3]	0.425	3.5 [21.1]	4.2 [25.4]	0.491	5.5 [24.3]	3.0 [22.6]	0.729
N1 (%)	1.7 [2.4]	3.4 [4.1]	< 0.05	2.5 [2.7]	3.8 [4.7]	< 0.05	2.5 [2.7]	4.4 [5.8]	< 0.05
N2 (%)	50.3 [18.3]	49.2 [11.4]	0.599	51.4 [13.4]	48.5 [11.4]	0.713	51.0 [12.1]	46.9 [11.7]	0.313
N3 (%)	29.4 [11.3]	27.4 [10.5]	0.650	27.9 [11.9]	27.4 [9.8]	0.413	27.7 [10.7]	27.2 [10.4]	0.636
REM (%)	19.2 [9.6]	18.3 [7.9]	0.412	18.5 [9.9]	18.3 [7.5]	0.988	18.5 [9.5]	18.6 [6.9]	0.868
AHI (event/h)	0.6 [0.4]	4.7 [8.9]	< 0.05	1.3 [1.2]	9.3 [12.5]	< 0.05	1.7 [2.1]	11.2 [12.1]	< 0.05
REM AHI (event/h)	0.3 [1.4]	7.5 [18.7]	< 0.05	1.8 [3.8]	13.3 [28.2]	< 0.05	2.4 [4.5]	22.5 [30.9]	< 0.05
NREM AHI (event/h)	0.3 [0.5]	3.6 [6.8]	< 0.05	1.1 [1.2]	6.8 [9.4]	< 0.05	1.4 [1.9]	8.8 [10.9]	< 0.05
Supine AHI (event/h)	0.4 [0.9]	5.1 [9.0]	< 0.05	1.1 [2.3]	7.6 [10.6]	< 0.05	1.7 [3.2]	11.3 [13.1]	< 0.05
Prone AHI (event/h)	0.0 [0.0]	0.0 [5.4]	< 0.05	0.0 [0.7]	2.1 [10.2]	< 0.05	0.0 [1.4]	1.1 [14.9]	< 0.05
Side AHI (event/h)	0.2 [0.6]	3.6 [9.2]	< 0.05	0.8 [1.6]	8.4 [13.7]	< 0.05	1.1 [2.1]	10.4 [14.3]	< 0.05
ODI3 (event/h)	0.5 [1.4]	4.2 [7.8]	< 0.05	1.3 [2.1]	7.7 [10.2]	< 0.05	1.8 [2.8]	9.2 [12.8]	< 0.05
REM ODI (event/h)	0.8 [2.4]	7.5 [16.2]	< 0.05	2.4 [4.4]	13.0 [23.6]	< 0.05	3.0 [6.1]	18.9 [30.2]	< 0.05
NREM ODI (event/h)	0.5 [1.5]	3.1 [6.2]	< 0.05	1.1 [1.8]	5.0 [7.6]	< 0.05	1.4 [2.3]	7.5 [11.2]	< 0.05
Supine ODI (event/h)	0.9 [2.2]	5.1 [9.4]	< 0.05	1.9 [2.6]	8.4 [12.6]	< 0.05	2.3 [3.8]	10.2 [15.0]	< 0.05
Prone ODI (event/h)	0.0 [1.1]	0.7 [5.5]	< 0.05	0.0 [1.5]	0.9 [9.3]	< 0.05	0.0 [2.2]	0.0 [14.6]	0.068
Side ODI (event/h)	0.8 [2.2]	4.2 [9.6]	< 0.05	1.4 [3.0]	6.3 [13.3]	< 0.05	1.7 [3.7]	10.0 [15.1]	< 0.05
Mean SpO2 (%)	97.0 [1.0]	96.0 [2.0]	< 0.05*	97.0 [1.0]	96.0 [2.0]	< 0.05	97.0 [1.0]	96.0 [2.0]	< 0.05
Lowest SpO2 (%)	92.0 [3.5]	87.0 [10.0]	< 0.05	91.0 [4.0]	85.0 [12.0]	< 0.05	91.0 [5.0]	83.0 [11.5]	< 0.05
90–100 SpO2 (%)	99.0 [4.0]	98.0 [6.0]	0.112	99.0 [4.0]	98.0 [7.0]	< 0.05*	99.0 [4.0]	97.0 [8.0]	< 0.05

Data are provided as median and interquartile range

AHI apnea-hypopnea index, *OSAS* obstructive sleep apnea syndrome, *OSAS negative* OSAS negative group, *OSAS positive* OSAS positive group, *PSG* polysomnography, *TRT* total recording time, *TST* total sleep time, *WASO* wakefulness after sleep onset, *sleep eff* sleep efficiency, *sleep onset lat* sleep onset latency, *REM* rapid eye movement, *REM onset lat* REM onset latency, *N1* N1 sleep stage, *N2* N2 sleep stage, *N3* N3 sleep stage, *NREM* no REM sleep, *ODI3* 3% oxygen desaturation index, *SpO2* blood oxygen desaturation

*p value > 0.01

[a] Non-parametric Mann-Whitney *U* test

statistical analysis was carried out to assess differences between the ROC curves of the four classification approaches. Using a clinical threshold of 1 event/h for positive OSAS, significant statistical differences ($p < 0.05$) were found between LDA and all the remaining classifiers (QDA, LR, and ODI3), suggesting a real dependence on the classification model. In addition, visual differences observed between ROC curves of QDA and ODI3 led to slight differences ($p < 0.05$

Table 3 Optimum features automatically selected using FCBF and bootstrap as well as median values for each AHI cutoff point under study

		AHI ≥ 1 event/h		AHI ≥ 3 events/h		AHI ≥ 5 events/h	
Optimum features	No. of times selected	OSAS negative	OSAS positive	OSAS negative	OSAS positive	OSAS negative	OSAS positive
M1t	950	97.7 [1.1]	97.3 [1.9]	97.7 [1.1]	96.9 [1.9]	97.7 [1.2]	96.7 [1.8]
M4t	921	2.83 [2.76]	2.18 [1.70]	2.23 [1.93]	2.21 [1.72]	2.24 [1.76]	2.17 [2.06]
PA	966	4.03 [2.45]	8.55 [15.09]	4.90 [3.38]	11.38 [19.27]	5.49 [4.06]	18.02 [29.78]
P_R	981	0.17 [0.05]	0.18 [0.07]	0.17 [0.06]	0.19 [0.08]	0.17 [0.06]	0.20 [0.07]
SampEn (×10^{-3})	994	2.44 [0.99]	4.06 [2.99]	2.74 [1.29]	4.84 [2.88]	2.86 [1.55]	5.19 [4.24]
ODI3	1000	0.49 [1.36]	4.24 [7.99]	1.34 [2.14]	7.68 [10.33]	1.82 [2.76]	9.22 [12.86]

Data are provided as median and interquartile range

AHI apnea-hypopnea index, *OSAS* obstructive sleep apnea syndrome, *OSAS negative* OSAS negative group, *OSAS positive* OSAS positive group, *M1t* mean of SpO2 overnight profile in the time domain, *M4t* kurtosis of SpO2 overnight profile in the time domain, *PA* peak amplitude in the frequency band of interest, P_R relative power in the frequency band of interest, *SampEn* sample entropy, *ODI3* oxygen desaturation index of 3%
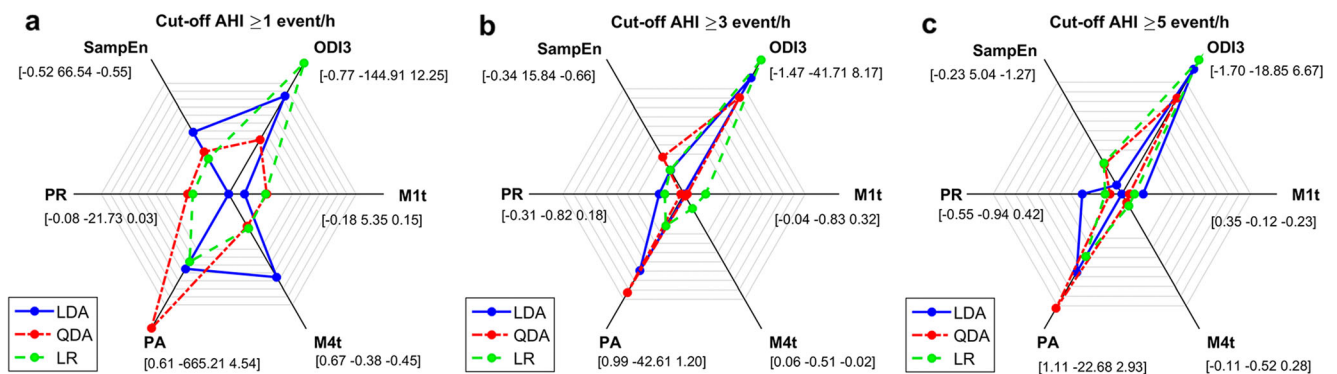
Fig. 1 Optimum models for each AHI-derived diagnostic cutoff for OSAS: **a** AHI ≥ 1 event/h, **b** AHI ≥ 3 events/h, and **c** AHI ≥ 5 events/h. For each statistical multivariate model, polygon vertices are proportional to the influence (weight) of each oximetric variable. For the sake of better representation within the same figure, the magnitudes of the coefficients are proportional to their influence within the same model but not between models

and $p = 0.21$ before and after the Bonferroni correction, respectively). Regarding the cutoffs of 3 and 5 events/h, no significant differences were found neither between the models nor with the conventional ODI3.

Discussion

This prospective study provides a comparative assessment of the most frequently employed analytical techniques, namely, LDA and LR, for determination of suitability of overnight oximetry in the context of OSAS screening in children. In addition, an alternative approach in the framework of statistical classifiers, the QDA binary classifier, was also evaluated due to its potential advantages when compared to LDA or LR. Our results provide additional and important support to the conceptual framework that automated overnight oximetry is a consistent, unbiased, and

effective method as an abbreviated screening tool for pediatric OSAS [32], and further confirm the validity of a recent multicenter study that examined 4191 overnight oximetry recordings [15].

Conventional statistical classifiers are commonly proposed as automated tools to assist in the detection of both adult and pediatric OSAS. They allow clinicians to combine different sources of medical information, and they have demonstrated their usefulness to maximize the diagnostic ability of oximetry by merging several oximetric features. In the present study, a thorough assessment of the diagnostic performance of LDA, QDA, and LR has been performed. We found that differences among these classifiers decreased when the clinical threshold for positive childhood OSAS increased. LR showed a more stable behavior than LDA and QDA. Moreover, LR reached the highest diagnostic accuracy for all the thresholds under study. Nevertheless, QDA stands out for reaching PPV values greater than 90% regardless the cutoff, which increases its usefulness as screening

Table 4 Performance metrics of each optimum statistical model and the ODI3 from oximetry for each AHI cutoff value for positive OSAS

	Se (%)	Sp (%)	PPV (%)	NPV (%)	LR+	LR−	Acc (%)
Cutoff AHI ≥ 1 event/h							
LDA	70.2 (54.0, 84.6)	63.0 (31.7, 96.7)	90.5 (81.8, 99.1)	30.3 (16.1, 46.2)	2.31 (1.15, 6.23)	0.51 (0.28, 0.97)	69.0 (57.5, 9.1)
QDA	61.3 (47.6, 76.0)	88.6 (69.6, 100)	96.5 (90.3, 100)	32.0 (19.4, 45.7)	8.77 (4.53, 14.80)	0.45 (0.29, 0.61)	65.8 (55.1, 76.8)
LR	93.9 (84.8, 99.8)	37.8 (10.3, 67.1)	88.1 (81.3, 94.5)	59.0 (28.9, 96.2)	1.62 (1.06, 2.78)	0.18 (0.01, 0.55)	84.3 (76.8, 91.5)
ODI ≥ 1	9.7 (3.4, 17.4)	100 (100, 100)	100 (100, 100)	18.5 (11.6, 25.6)	N.D. (−)	0.90 (0.83, 0.97)	25.1 (17.0, 33.8)
Cutoff AHI ≥ 3 events/h							
LDA	59.4 (45.4, 73.7)	88.2 (70.1, 99.5)	86.6 (71.2, 99.3)	63.9 (52.3, 75.6)	7.53 (2.20, 28.15)	0.46 (0.30, 0.64)	72.3 (62.6, 81.4)
QDA	57.5 (41.8, 73.3)	93.8 (83.9, 100)	92.3 (81.6, 100)	64.3 (52.8, 75.9)	15.00 (4.65, 33.07)	0.45 (0.30, 0.62)	73.8 (64.3, 82.7)
LR	74.6 (59.8, 89.4)	81.7 (57.4, 95.5)	83.9 (68.0, 95.2)	72.5 (59.4, 85.2)	5.18 (1.97, 14.24)	0.31 (0.16, 0.48)	77.7 (68.5, 85.5)
ODI ≥ 3	4.9 (0.8, 12.2)	100 (100, 100)	100 (100, 100)	46.1 (36.7, 55.1)	N.D. (−)	0.95 (0.88, 0.99)	47.6 (38.0, 56.5)
Cutoff AHI ≥ 5 events/h							
LDA	63.7 (48.1, 79.1)	90.1 (77.6, 99.3)	82.0 (64.3, 98.3)	78.6 (68.5, 87.6)	8.61 (3.02, 26.99)	0.40 (0.23, 0.58)	79.4 (70.9, 87.4)
QDA	62.3 (46.5, 77.9)	95.5 (87.3, 100)	90.8 (76.3, 100)	78.9 (69.0, 87.8)	19.83 (6.44, 46.42)	0.40 (0.24, 0.57)	82.1 (73.8, 89.5)
LR	70.0 (55.1, 84.1)	91.4 (80.3, 99.3)	85.1 (70.1, 98.4)	81.9 (72.3, 90.5)	10.60 (3.89, 30.85)	0.33 (0.18, 0.49)	82.7 (75.0, 89.6)
ODI ≥ 5	2.9 (0.0, 10.8)	100 (100, 100)	100 (100, 100)	60.3 (51.2, 69.1)	N.D. (−)	0.97 (0.89, 1.00)	60.8 (51.6, 69.6)

Data are presented as mean and 95% confidence interval from a bootstrap procedure of 1000 iterations

AHI apnea-hypopnea index, *LDA* linear discriminant analysis, *QDA* quadratic discriminant analysis, *LR* logistic regression, *Se* sensitivity, *Sp* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *LR+* positive likelihood ratio, *LR−* negative likelihood ratio, *Acc* accuracy, *N.D.* not defined due to division by zero
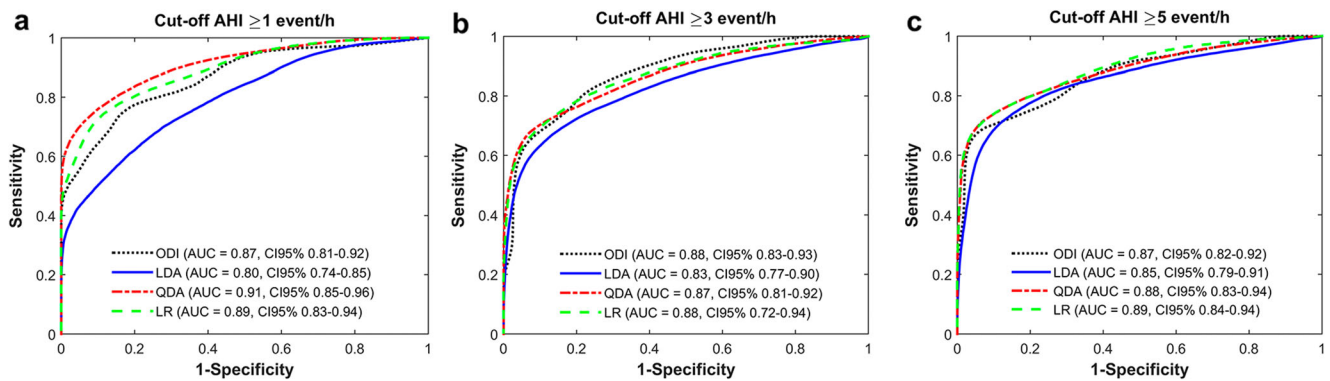
Fig. 2 ROC curves of each optimum statistical model from oximetry and the ODI3 from the bootstrap procedure for each cutoff for positive OSAS: **a** AHI ≥ 1 event/h, **b** AHI ≥ 3 events/h, and **c** AHI ≥ 5 events/h

test for pediatric OSAS in the clinical practice. In fact, oximetry is commonly proposed as a simplified screening test for OSAS due to its simplicity and availability as portable recording technique at home. Therefore, despite the higher overall performance of LR, QDA showed the best screening capability, particularly for a cutoff of AHI ≥ 1 event/h, where a PPV of 96.5% was reached while maintaining AUC of 0.91 and fair sensitivity and specificity pair. Regarding common indexes from oximetry, statistical binary classifiers outperformed the conventional ODI3 in terms of diagnostic accuracy. When using common cutoffs for classification (ODI3 ≥ 1, 3, and 5 events/h), the ODI3 reached a highly unbalanced sensitivity-specificity pair leading to poor accuracy. Furthermore, regarding the ROC curves, the ODI3 showed higher instability when varying the cutoff, which minimizes its usefulness in the clinical practice.

Some methodological issues deserve comment. First, all the variables composing the various models were automatically selected using unbiased data processing and all of the measures were identified regardless of the classification technique, a feature that expands their generalizability. In the present comparative study, *M1t*, *M4t*, *PA*, P_R, *SampEn*, and *ODI3* demonstrated unique robustness in characterizing OSAS because they were selected as optimum features a number of times significantly higher (> 90% of bootstrap replicates) than the remaining variables (< 50% of bootstrap replicates). As shown in Fig. 1, the influence of each oximetric variable in every optimum model illustrates this important attribute of machine classification of the retained measures. When using them jointly, automated multivariate classifiers are able to maximize the diagnostic capability of oximetry. On the other hand, we can observe that *M1t* and *M4t* did not show a completely stable behavior; i.e., the sign of their weights (the model coefficients) changed as the clinical cutoff for the disease varied. The remaining optimum features (*PA*, P_R, *SampEn*, and *ODI3*) showed different weight and sign for the different statistical approaches but a stable behavior as the cutoff changed. It is also important to note that *PA* and *ODI3* showed the largest weights regardless the model and the

cutoff, which highlights their significance in the characterization of oximetry in the context of pediatric OSAS.

In regard to the physiological and clinical interpretation of the optimum features, it is noticeable that all signal processing approaches were included in the optimum feature subset, i.e., time and frequency domains as well as linear and non-linear methods. Similarly, advanced signal processing methods demonstrated their complementarity with ODI3. Conventional ODIs just account for the number of the desaturations along the overnight recording. Nevertheless, *M1t*, *M4t*, *PA*, P_R, and *SampEn* provide additional information linked with the depth, duration, and repetition of the desaturations. *M1t* accounts for the overall influence of the desaturations in the average saturation level while *M4t* quantifies how SpO$_2$ values are distributed in a narrow (higher *M4t*) or in a broader (lower *M4t*) range according to a higher severity and number of desaturations. Similarly, both *PA* and P_R account for the pseudo-periodicity and depth of the desaturations along the overnight recording so that greater and repetitive desaturations (higher *PA* and P_R) are linked with a higher severity of the disease. Finally, *SampEn* quantifies the entropy, i.e., the disorder, in the overnight saturation profile due to changes caused by the desaturations so that the higher the severity of the disease is, the higher the irregularity is (higher *SampEn*).

Changes in the dynamics of oximetric profiles due to apneic events are smaller in children than in adults, thereby making oximetry-based screening tools for pediatric OSAS particularly challenging. The best screening ability emerged when using the cutoff of 1 events/h for detection of mild-to-severe OSAS. QDA and LR models showed higher generalization capability than LDA. Common assumptions such as normal distribution, homoscedasticity, linearity, or stationarity are not always consistent in real-world pattern classification problems. Therefore, our results suggest that more general modeling approaches such as LR and QDA, which both assume less restrictive conditions in the data distribution and implement a non-linear combination of input-dependent features, perform better than conventional LDA in the context of detection of childhood OSAS from

oximetry. Notice that QDA reached excellent AUC (> 0.90) for the lowest cutoff. Furthermore, a real dependence on the classification model was found.

Both dependence and intercorrelation of the input variables affects the way coefficients are fitted in statistical models. Intercorrelation may lead to significant coefficients with high values, although these models could not fit properly the problem under study. FCBF is a feature selection methodology able to minimize this issue since it removes redundant features before the classification stage, i.e., the variables with likely high intercorrelation. However, the selected non-redundant features could still show high intercorrelation. Regarding our optimum feature subset from oximetry, we observed that $M1t$, PA, $SampEn$, and $ODI3$ were significantly correlated ($p < 0.05$), whereas $M4t$ and P_R showed no significant correlations. This issue could affect the consistence and generalization of the models. Similarly, dependence of oximetric features should also be considered because they are all derived from the same signal and common statistical classifiers assume independent input variables.

Overfitting is also an important problem in the framework of automated pattern recognition, particularly when the size of the training dataset is small. In addition, some classifiers are more affected by overfitting than others. Particularly, QDA is more flexible in fitting the classes on the training data due to its quadratic characteristic, which increases the risk of overfitting. In this study, we compared the average accuracy computed in the training and in the test bootstrap replicates in order to assess whether the performance of the proposed statistical models was affected by overfitting. We observed that differences between training and test values were lower than 5% regardless the model and the cutoff. This slight difference fits with the common expected decrease between training and test stages, suggesting that there is no overfitting affecting our results. In addition, bootstrap 0.632 accounts for both the training and the test contributions when computing every performance metric, which minimizes these differences.

It is important to assess whether statistical models provided additional value compared to the simple ODI3. According to our results, differences in performance between the statistical automated models and with the conventional ODI3 were maximal for AHI ≥ 1 event/h. In addition, ROC curves of LDA, QDA, and LR were smooth while the curves for the ODI3 showed marked changes in the slope regardless the clinical diagnostic threshold for the disease. This characteristic reveals higher irregularity of ODI3 when the cutoff for classification changes. Using standard classification cutoffs for the ODI, i.e., ODI3 ≥ 1, 3, and 5 desaturations per hour of recording, this conventional index reached a diagnostic accuracy significantly lower than the statistical models, as well as sensitivity vs. specificity pairs highly unbalanced (Se < 10%). On the other hand, QDA reached PPV > 90% and LR+ > 5 with fair sensitivity and specificity balance (Se around 60%). Both

characteristics are essential for screening purposes, which is probably the main advantage of statistical models over ODI3. Therefore, our results suggest that automated modeling of overnight oximetry by means of QDA provides further and relevant information on the dynamics of oximetry compared to the conventional ODI3, particularly when AHI ≥ 1 event/h is used as the clinical cutoff for the disease.

As previously mentioned, the QDA model reached notably high PPV and LR+ values regardless of the cutoff for positive OSAS. Particularly, 96.5% PPV and 8.77 LR+ were reached using a threshold of 1 event/h. This agrees with previous studies reporting that oximetry is able to achieve high PPV values for the detection of an AHI ≥ 1 event/h [14, 32]. Most sleep laboratories use a clinical threshold of 1 event per hour during interpretation of PSG [2], which increases the usefulness of these automated techniques for screening purposes. Inspecting false positive cases using QDA, we observed that two no-OSAS (actual AHI < 1 event/h) children were misclassified as OSAS-positive, a number of times notably higher than the remaining patients throughout the 1000 iterations of the bootstrap algorithm. One of these children (actual AHI and ODI from PSG equal to 0 events/h) showed mean SpO_2 significantly lower (92%) than the average for the OSAS-negative group (97%), a feature characteristic of severe OSAS. The other one (actual AHI and ODI from PSG equal to 0.4 and 0 events/h, respectively) showed no sleep time with a saturation in the range of 90–100%, which could suggest that other factors may be specifically detracting from the validity of the classifiers in these children, e.g., underlying parenchymal lung disease or skin color artifacts in oximetry readouts. Considering a cutoff of three events per hour, three no-OSAS (actual AHI < 3 events/h) were significantly more frequently misclassified as OSAS-positive by the oximetry-based QDA model throughout the bootstrap samples. Two of them (actual AHI from PSG equal to 1.3 and 1.8 events/h) showed an ODI greater than 13 events/h and lowest SpO_2 significantly lower (74 and 56%, respectively) than the average for the OSAS-negative group (91%), while the other one was borderline (actual AHI equal to 2.98 events/h) and also showed significantly higher ODI (7.1 events/h) and lower minimum SpO_2 (81%) than the average for the no-OSAS population. Similarly, three children were wrongly classified as suffering from the disease by the QDA model using the cutoff of 5 events/h for positive OSAS. Two of them were also misclassified using a clinical threshold of 3 events/h (subjects with actual AHI equal to 1.3 and 1.8 events/h, respectively), whereas the other one (actual AHI equal to 4.0 events/h) showed an ODI of 7.6 events/h and minimum SpO_2 equal to 78%, which both are values characteristic of OSAS-positive children.

Table 5 shows previous studies using these statistical classifiers in the context of childhood OSAS diagnosis using pulse oximetry. LDA [9, 16–18] and LR [10, 14, 19, 20] were predominantly used. Nevertheless, to the best of our knowledge, there are no proper and extensive comparisons among these a

Table 5 Summary of the studies using statistical binary classifiers from overnight oximetry to assist in the detection of childhood OSAS

Author (year)	Dataset (n)	Gold standard (cutoff)	Proposed technique	Variables	Classification	Se (%)	Sp (%)	Acc (%)
Gil et al. (2010) [16]	21 children with suspected OSAS	In-lab PSG (AHI ≥ 5 events/h)	PPG and PTT from pulse oximetry and ECG	Time-frequency analysis	Linear discriminant	75.0	85.7	80.0
Chang et al. (2013) [10]	141 children with suspected OSAS	In-lab PSG (AHI ≥ 5 events/h)	Questionnaire and oximetry	Presence of mouth breathing, restless sleep, ODI4	Logistic regression	60.0	86.0	71.6
Garde et al. (2014) [9]	146 children with suspected OSAS	In-lab PSG (AHI ≥ 5 events/h)	Portable oximetry (attended)	Time and spectral: - SpO₂ - SpO₂ + PR	Linear discriminant	80.0 88.4	83.9 83.6	78. 84.9
Lázaro et al. (2014) [17]	21 children with suspected OSAS	In-lab PSG (AHI ≥ 5 events/h)	PPG and PTT from pulse oximetry	T-F maps	Linear discriminant	100	71.4	86.7
Gutiérrez-Tobal et al. (2015) [19]	50 children with suspected OSAS	In-lab PSG (AHI ≥ 3 events/h)	Airflow + oximetry from HRP	Spectral features from airflow + ODI3	Logistic regression	85.9	87.4	86.3
Cohen & De Chazal (2015) [18]	288 infants (< 27 weeks) with suspected OSAS	At-home PSG (epoch-based classification)	Portable pulse oximetry + ECG (unattended)	T-F maps ECG Time SpO₂ + T-F maps ECG	Linear discriminant	39.6 58.1	76.4 67.0	74.7 66.7
Álvarez et al. (2017) [14]	50 children with suspected OSAS	In-lab PSG AHI ≥ 1 events/h AHI ≥ 3 events/h AHI ≥ 5 events/h	Port. oximetry from HRP (unattended)	Time (statistics and non-linear) and spectral features	Logistic regression	89.6 82.9 82.2	71.5 84.4 83.6	85.5 83.4 82.8
Wu et al. (2017) [20]	311 children with suspected OSAS	In-lab PSG AHI ≥ 5 events/h	Diagnostic scale	History, demographic, and anthropometric	Logistic regression	94.8	25.0	78.2

AHI apnea-hypopnea index, *OSAS* obstructive sleep apnea syndrome, *PSG* polysomnography, *PPG* photoplethysmography, *PTT* pulse transit time, *ECG* electrocardiogram, *HRP* home respiratory polygraphy, *ODI3* oxygen desaturation index ≥ 3%, *T-F* time-frequency maps, *Se* sensitivity, *Sp* specificity, *Acc* accuracy

priori effective techniques in the framework of automated pediatric OSAS detection. In the study by Gil et al. [16], photoplethysmographic and pulse transit time recordings were parameterized and LDA was applied. An accuracy of 80% was reached using leave-one-out cross-validation. In a later similar study by the same group [17], the performance increased up to 86.7% using the pulse rate signal from oximetry to compute the pulse transit time. Garde et al. [9] analyzed SpO₂ and pulse rate portable recordings. LDA achieved 84.9% accuracy using 4-fold cross-validation. In an epoch-based classification scheme, Cohen and De Chazal [18] built two LDA models using features from SpO₂ and ECG. A maximum accuracy of 74.7% was reached using a leave-one-out cross-validation approach.

Regarding LR, Chang et al. [10] reported 71.6% accuracy combining questionnaire-based variables and conventional oximetric indices. The model was optimized and validated using the same population. Similarly, Wu et al. [20] built a LR model with clinical history, demographic, and other relevant variables, reaching 78.2% accuracy in an independent test set. In a recent study by our group [14], the accuracy of different LR models from unattended oximetry ranged from 82.8 to 85.5% in a bootstrapping validation scheme. Similarly, a LR model composed of spectral features from airflow portable recordings and the ODI3 achieved 86.3% accuracy [19]. In the present study, our analyses further validate the efficacy of oximetry-based classification models aimed at screening for OSAS in children and attest to their robustness and high predictive ability.

Despite the favorable screening capability of the proposed oximetry-based models, some limitations should be taken into account. The population cohort under study should be expanded to increase the generalization of our results such as to include different types of oximeters, and enable comparisons across diverse populations such as those with craniofacial or genetic syndromes, as well as in infants. However, when compared with similar previous studies in the context of pediatric OSAS, the cohort included herein was quite extensive, and included implementation of an appropriate bootstrap approach for independent optimization and further validation of the models using several datasets. Similarly, a larger dataset would lead to a more balanced OSAS-negative and OSAS-positive groups and thus to a more accurate parameter estimation.

Our results suggest that oximetry is able to provide relevant and useful information in the context of pediatric OSAS. On the other hand, using the proposed oximetric features, differences among the statistical classifiers under study decreased as the cutoff for the disease increased. Therefore, additional features from signal processing methods as well as advanced pattern recognition techniques, such as ensemble learning, or more complex approaches for classification, such as deep neural networks, would be needed to maximize the screening ability of oximetry in moderate-to-severe cases. In addition, it is understood that the ultimate goal of simplified tools such as overnight oximetry is to carry out abbreviated screening tests at home. Therefore, the validity of the proposed models needs to be reproduced in an

unattended setting. Finally, despite its usefulness for the screening of mild (96.5% PPV for AHI ≥ 1 event/h) and moderate-to-severe (90.8% PPV for AHI ≥ 5 events/h) childhood OSAS, there is still room for improvement regarding the effectiveness of oximetry to classify no-OSAS children (AHI < 1 event/h). Such limitations will likely require more than a single-step modeling approach and incorporate additional screening stages, whereby a fail/pass no-OSAS decision by the proposed oximetry-based model will then undergo a second scrutiny using an alternative classifier. If we want to screen for no-OSAS children to minimize the number of complete in-lab PSGs, an alternative classifier aimed at maximizing the negative predictive value should be implemented. In this regard, as oximetry is characterized by a high specificity, other sources of information would be needed, such as history and clinical data or additional biomedical recordings, e.g., airflow or pulse rate. On the other hand, additional useful information could be derived from the same sensor so that the simplicity of our proposal as abbreviated screening test for pediatric OSAS is not compromised. Pulse oximetry devices are able to provide information from both oxygen saturation and heart rate. In the present study, we used just the SpO_2 signal. Nevertheless, previous researchers demonstrated the usefulness of the pulse rate derived from the photoplethismographic signal to derive relevant information about the presence of pediatric OSAS [9, 16, 17]. Therefore, features derived from different signals but acquired using a single device could be used to improve the performance of automated classifiers without increasing the intrusiveness of the proposal for children.

In summary, automated analysis of SpO_2 recordings from nocturnal oximetry emerges as a simple as well as reliable alternative to complete PSG in the screening of children with high pre-test probability of OSAS. Our results provide additional insights on the effectiveness and limitations of statistical modeling of overnight SpO_2 recordings as an abbreviated screening tool for childhood OSAS.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Kadmon G, Shapiro CM, Chung SA, Gozal D (2013) Validation of a pediatric obstructive sleep apnea screening tool. Int J Pediatr Otorhinolaryngol 77(9):1461–1464. https://doi.org/10.1016/j.ijporl.2013.06.009
2. Marcus CL, Brooks LJ, Ward SD et al (2012) Diagnosis and management of childhood obstructive sleep apnea syndrome. Pediatrics 130(3):e714–e755. https://doi.org/10.1542/peds.2012-1672
3. Kirk VG, Bohn SG, Flemons WW, Remmers JE (2003) Comparison of home oximetry monitoring with laboratory polysomnography in children. Chest 124(5):1702–1708. https://doi.org/10.1378/chest.124.5.1702
4. Kheirandish-Gozal L (2010) What is "abnormal" in pediatric sleep? Respir Care 55:1366–1376
5. Lesser DJ, Haddad GG, Bush RA, Pian MS (2012) The utility of a portable recording device for screening of obstructive sleep apnea in obese adolescents. J Clin Sleep Med 8(3):271–277. https://doi.org/10.5664/jcsm.1912
6. Katz ES, Ron BM, D'Ambrosio CM (2012) Obstructive sleep apnea in infants. Am J Respir Crit Care Med 185(8):805–816. https://doi.org/10.1164/rccm.201108-1455CI
7. Kaditis AG, Alonso-Alvarez ML, Boudewyns A et al (2016) Obstructive sleep disordered breathing in 2- to 18-year-old children: diagnosis and management. Eur Respir J 47(1):69–94. https://doi.org/10.1183/13993003.00385-2015
8. Alonso-Álvarez ML, Terán-Santos J, Ordax-Carbajo E et al (2015) Reliability of home respiratory polygraphy for the diagnosis of sleep apnea in children. Chest 147:1020–1028
9. Garde A, Dehkordi P, Karlen W et al (2014) Development of a screening tool for sleep disordered breathing in children using the phone OximeterTM. PLoS One 9:e112959
10. Chang L, Wu J, Cao L (2013) Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. Int J Pediatr Otorhinolaryngol 77(3):365–371. https://doi.org/10.1016/j.ijporl.2012.11.028
11. Cohen G, de Chazal P (2013) Automated detection of sleep apnea in infants using minimally invasive sensors. In: Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE-EMBC), 3–7 July 2013, Osaka (Japan), pp 1652–1655. https://doi.org/10.1109/EMBC.2013.6609834
12. Tsai CM, Kang CH, Su MC, Lin HC, Huang EY, Chen CC, Hung JC, Niu CK, Liao DL, Yu HR (2013) Usefulness of desaturation index for the assessment of obstructive sleep apnea syndrome in children. Int J Pediatr Otorhinolaryngol 77(8):1286–1290. https://doi.org/10.1016/j.ijporl.2013.05.011
13. Sahadan DZ, Davey MJ, Horne RSC, Nixon GM (2015) Improving detection of obstructive sleep apnoea by overnight oximetry in children using pulse rate parameters. Sleep Breath 19(4):1409–1414. https://doi.org/10.1007/s11325-014-1108-4
14. Álvarez D, Alonso-Álvarez ML, Gutiérrez-Tobal GC, Crespo A, Kheirandish-Gozal L, Hornero R, Gozal D, Terán-Santos J, del Campo F (2017) Automated screening of children with obstructive sleep apnea using nocturnal oximetry: an alternative to respiratory polygraphy in unattended settings. J Clin Sleep Med 13(05):693–702. https://doi.org/10.5664/jcsm.6586
15. Hornero R, Kheirandish-Gozal L, Gutiérrez-Tobal GC et al (2017) Nocturnal oximetry-based evaluation of habitually snoring children. Am J Respir Crit Care Med 196:1591–1598

16. Gil E, Bailón R, Vergara JM et al (2010) PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children. IEEE Trans Biomed Eng 57:1079–1088

17. Lázaro J, Gil E, Vergara JM et al (2014) Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. IEEE J Biomed Health Inform 18:240–246

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

19. Gutiérrez-Tobal GC, Alonso-Álvarez ML, Álvarez D, del Campo F, Terán-Santos J, Hornero R (2015) Diagnosis of pediatric obstructive sleep apnea: preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home. Biomed Signal Process Control 18:401–407. https://doi.org/10.1016/j.bspc.2015.02.014

20. Wu D, Li X, Guo X, Qin J, Li S (2017) A simple diagnostic scale based on the analysis and screening of clinical parameters in paediatric obstructive sleep apnoea hypopnea syndrome. J Laryngol Otol 131(04):363–367. https://doi.org/10.1017/S0022215117000238

21. Shouldice RB, O'Brien LM, O'Brien C et al (2004) Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features. Sleep 27(4):784–792. https://doi.org/10.1093/sleep/27.4.784

22. Montgomery-Downs HE, O'Brien LM, Gulliver TE, Gozal D (2006) Polysomnographic characteristics in normal preschool and early school-aged children. Pediatrics 117(3):741–753. https://doi.org/10.1542/peds.2005-1067

23. Berry RB, Budhiraja R, Gottlieb DJ, et al for the American Academy of Sleep Medicine (2012). Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. Deliberations Sleep Apnea Definitions Task Force Am Academy Sleep Medicine J Clin Sleep Med 8:597–619

24. Álvarez D, Hornero R, Marcos JV et al (2013) Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of sleep apnea diagnosis. Int J Neural Syst 23:e13520

25. Alvarez D, Gutiérrez-Tobal GC, Alonso-Álvarez ML, et al (2014) Análisis espectral y no lineal de la señal de oximetría domiciliaria en la ayuda al diagnóstico de la apnea infantil. In proc 32th Annu Nac Conf Spanish Biomedical Engineering Society, Barcelona (Spain), pp 1–4

26. Álvarez D, Hornero R, Marcos JV et al (2010) Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. IEEE Trans Biomed Eng 57:2816–2824

27. Gutiérrez-Tobal GC, Álvarez D, Marcos JV et al (2013) Pattern recognition in airflow recordings to assist in the sleep apnoea–hypopnoea syndrome diagnosis. Med Biol Eng Comput 51:1367–1380

28. Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5:1205–1224

29. Crespo A, Álvarez D, Gutiérrez-Tobal GC et al (2017) Multiscale entropy analysis of unattended oximetric recordings to assist in the screening of paediatric sleep apnoea at home. Entropy 19:284

30. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York

31. Witten IH, Frank E, Hall MA (2011) Data mining practical machine learning tools and techniques. Burlington, Morgan Kaufmann/Elsevier

32. Kaditis A, Kheirandish-Gozal L, Gozal D (2016) Pediatric OSAS: oximetry can provide answers when polysomnography is not available. Sleep Med Rev 27:96–105