

Article

Assessment of Airflow and Oximetry Signals to Detect Pediatric Sleep Apnea-Hypopnea Syndrome Using AdaBoost

Jorge Jiménez-García ^{1,*}, Gonzalo C. Gutiérrez-Tobal ^{1,2}, María García ^{1,2},
Leila Kheirandish-Gozal ³, Adrián Martín-Montero ¹, Daniel Álvarez ^{1,2,4},
Félix del Campo ^{1,2,4}, David Gozal ³ and Roberto Hornero ^{1,2}

¹ Biomedical Engineering Group, University of Valladolid, 47011 Valladolid, Spain; gonzalo.gutierrez@gib.tel.uva.es (G.C.G.-T.); maria.garcia@tel.uva.es (M.G.); adrian.martin@gib.tel.uva.es (A.M.-M.); dalvarezgo@saludcastillayleon.es (D.Á.); fsas@telefonica.net (F.d.C.); robhor@tel.uva.es (R.H.)

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, 28029 Madrid, Spain

³ Department of Child Health, The University of Missouri School of Medicine, Columbia, MO 65212, USA; gozall@health.missouri.edu (L.K.-G.); gozald@health.missouri.edu (D.G.)

⁴ Sleep-Ventilation Unit, Pneumology Service, Río Hortega University Hospital, 47012 Valladolid, Spain

* Correspondence: jorge.jimenez@gib.tel.uva.es; Tel.: +34-983-18-47-16

Received: 30 April 2020; Accepted: 15 June 2020; Published: 17 June 2020



Abstract: The reference standard to diagnose pediatric Obstructive Sleep Apnea (OSA) syndrome is an overnight polysomnographic evaluation. When polysomnography is either unavailable or has limited availability, OSA screening may comprise the automatic analysis of a minimum number of signals. The primary objective of this study was to evaluate the complementarity of airflow (AF) and oximetry (SpO₂) signals to automatically detect pediatric OSA. Additionally, a secondary goal was to assess the utility of a multiclass AdaBoost classifier to predict OSA severity in children. We extracted the same features from AF and SpO₂ signals from 974 pediatric subjects. We also obtained the 3% Oxygen Desaturation Index (ODI) as a common clinically used variable. Then, feature selection was conducted using the Fast Correlation-Based Filter method and AdaBoost classifiers were evaluated. Models combining ODI 3% and AF features outperformed the diagnostic performance of each signal alone, reaching 0.39 Cohens's kappa in the four-class classification task. OSA vs. No OSA accuracies reached 81.28%, 82.05% and 90.26% in the apnea-hypopnea index cutoffs 1, 5 and 10 events/h, respectively. The most relevant information from SpO₂ was redundant with ODI 3%, and AF was complementary to them. Thus, the joint analysis of AF and SpO₂ enhanced the diagnostic performance of each signal alone using AdaBoost, thereby enabling a potential screening alternative for OSA in children.

Keywords: sleep apnea-hypopnea syndrome; airflow; oximetry; AdaBoost; spectral analysis; nonlinear analysis

1. Introduction

Childhood Obstructive Sleep Apnea (OSA) syndrome is a sleep disorder in which airflow is intermittently interrupted or decreased during sleep, mainly due to the obstruction of the upper airway [1,2]. Events of absence (apnea) or reduction (hypopnea) in air exchange caused by these obstructions reduce the oxygenation of blood, and disturb the normal progression of sleep stages, which leads to restless sleep, daytime sleepiness and behavioral problems [1,2]. Untreated pediatric

OSA can lead to neurocognitive deficits, cardiovascular complications and negative consequences related to development and behavior [3–5]. OSA prevalence in children is between 1–5%, with a high proportion of cases likely undiagnosed [2,3]. The gold standard of OSA diagnosis is a nocturnal in-lab polysomnography (PSG) [2,3,6], which consists in the analysis of children's sleep by recording an extensive number of biomedical signals [6–8]. Physicians then calculate the apnea–hypopnea index (AHI), the rate of apneas and hypopneas per hour of sleep (e/h). Pediatric OSA severity is then classified into three levels: Mild ($1 \leq \text{AHI} < 5$ e/h), Moderate ($5 \leq \text{AHI} < 10$ e/h) or Severe ($\text{AHI} \geq 10$ e/h) [9]. Early and timely diagnosis of OSA is of crucial importance to allow children to be assessed for surgical treatment, depending on their severity and comorbidities, while also preventing the negative consequences of the disease [9].

Costs, complexity and scarce availability of pediatric sleep laboratory facilities are the main disadvantages of PSG [8,10]. Several alternatives have been proposed to overcome them. Mostly accepted approaches aim to reduce the complexity of PSG, and usually rely on the analysis of a reduced set of signals. Some examples include nocturnal oximetry and respiratory polygraphy, which includes the recording of airflow (AF) and oximetry (SpO_2) signals [7,10–13]. AF reflects the respiratory activity during sleep, while SpO_2 represents the arterial oxygen saturation [14]. Both signals are involved in the definition of apnea or hypopnea events and, therefore, provide useful information to detect OSA [3,14]. Thus, it is possible to retrieve most of the information related to apneas and hypopneas using the AF and SpO_2 signals included in the PSG [8,12]. Some studies have reported that oximetry alone is less reliable than respiratory polygraphy [7,13]. Hence, the joint analysis of AF and SpO_2 from PSG can enhance pediatric OSA screening with respect to using SpO_2 only.

Previous studies focused on the assessment of AF or SpO_2 signals as simplified alternatives to PSG [15–17]. They involved adult [18–22] and pediatric [23–32] subjects as well as different automated signal processing methods. Regarding automatic analysis of SpO_2 signals, several methods have been proposed to detect OSA in children. These studies comprised time and frequency domain analyses [23–26], as well as nonlinear methods and various oximetric indexes [25–29]. Studies involving AF signals comprised spectral analysis [30], irregularity and variability analyses [31], and nonlinear recurrence plots [32]. Some of these studies established the complementarity of AF with the 3% oxygen desaturation index (ODI) [30,32]. This clinical variable from SpO_2 counts the number of oxygen desaturations $>3\%$ per hour of recording, and it is used as simplified screening test in clinical practice due to pediatric sleep laboratories are not widely available to perform complete PSG in children [12,13]. Both ODI 3% and ODI 4% have been proposed in conjunction with other oximetric variables in the screening of OSA using nocturnal oximetry [12]. However, there is a scarcity of studies regarding the joint analysis of AF and SpO_2 signals in children [33]. Only one recent study in adults focused on a direct comparison of the diagnostic ability of AF and SpO_2 signals, which evidenced their complementarity [34]. Pediatric OSA differs from the adult disease in the symptom spectrum, negative outcomes and the PSG findings, such that adult criteria cannot be used for diagnosis in children [1,3,5,9,35]. Accordingly, the definition of apneic events and the AHI cutoffs used to define OSA severity are different in children [6,14,17]. Here, we propose to evaluate the diagnostic ability of AF and SpO_2 signals using a pediatric population.

From a machine learning point of view, previous studies focused on the automatic detection of pediatric OSA using simple and widespread algorithms. These approaches included Logistic Regression (LR) [23–25,27,28,30,31,36,37], Linear or Quadratic Discriminant Analysis (LDA, QDA) [38–40], and Support Vector Machines (SVM) [41]. Other recent studies relied on more complex Multilayer Perceptron (MLP) neural networks [26,29,32,42,43]. Most of these algorithms can be hardly generalizable due to their simplicity and susceptibility to overfitting the training data [44]. However, ensemble learning approaches have not yet been evaluated despite their high performance in studies involving AF or SpO_2 recordings from adults [19,22,45]. To avoid this limitation, we propose the use of AdaBoost, a well-known generalizable ensemble learning algorithm [46].

The main hypothesis of this study is that an adequate combination of the information from AF and SpO₂ signals can yield higher diagnostic performance than each of these signals separately. Therefore, our primary goal is to compare the information of AF and SpO₂ signals to detect OSA in children and evaluate their complementarity. Additionally, the secondary objective is to evaluate the diagnostic ability of AdaBoost classifiers using features from each signal separately and combined.

2. Database

The database used in this study comprised 974 pediatric subjects referred to the Comer Children's Hospital, University of Chicago Medicine (Chicago, IL, USA), with clinical suspicion of OSA. This study was conducted according to the Declaration of Helsinki. The legal caretakers of each subject provided the informed consent and the Ethics Committee of the University of Chicago Medicine approved the study protocol (#11-0268-AM017, # 09-115-B-AM031, and # IRB14-1241). In-laboratory sleep studies were performed with a digital polysomnography device (Nihon Kohden America Inc., Irvine, CA, USA). Subjects were evaluated according to the rules defined by the American Academy of Sleep Medicine (AASM) [14], including the computation of the AHI. The subjects of the database were classified according to OSA severity in four groups: No OSA (AHI < 1 e/h), Mild OSA (1 ≤ AHI < 5 e/h), Moderate OSA (5 ≤ AHI < 10 e/h) and Severe OSA (AHI ≥ 10 e/h). These severity groups were chosen in accordance with previous studies [8,9,26]. Table 1 summarizes sociodemographic—Age, number of males and females—And clinical data—Normalized body mass index (BMI z-score), AHI, number of patients with OSA—Of the subjects involved in this study. They were randomly split into a training set (60%) and a test set (40%). No significant differences were found in age, sex, BMI z-score and AHI between the two sets ($p > 0.01$, Mann–Whitney U test). The training set was used to fix the optimum values of the method parameters using a bootstrap approach and train the classifiers. The test set was used to evaluate the diagnostic performance of our algorithm.

Table 1. Sociodemographic and clinical data of the subjects involved in the study. Subjects distributions represented as N° (%). Age, normalized body mass index (BMI z-score) and apnea–hypopnea index (AHI) represented as the median (interquartile range).

	All	Training Set	Test Set
N° of Subjects	974	584 (60%)	390 (40%)
Age (years)	6.00 [3.00, 8.00]	6.00 [3.00, 8.00]	5.50 [3.00, 9.00]
N° of Males	599 (61.50%)	346 (59.25%)	253 (64.87%)
N° of Females	375 (38.50%)	238 (40.75%)	137 (35.13%)
BMI z-score	−0.22 [−0.60, 0.37]	−0.24 [−0.61, 0.43]	−0.17 [−0.58, 0.27]
AHI (events/hour)	3.80 [1.53, 9.35]	4.08 [1.71, 10.00]	3.30 [1.40, 7.87]
N° of No OSA	171 (17.56%)	96 (16.44%)	75 (19.23%)
N° of Mild OSA	398 (40.86%)	229 (39.21%)	169 (43.33%)
N° of Moderate OSA	176 (18.07%)	113 (19.35%)	63 (16.15%)
N° of Severe OSA	229 (23.51%)	146 (25.00%)	83 (21.28%)

Normalized body mass index (BMI z-score); apnea–hypopnea index (AHI); Obstructive Sleep Apnea (OSA).

3. Methods

In this study, AF and SpO₂ signals extracted from 974 PSG recordings were analyzed. AF signals were sampled at $f_s = 100$ Hz, while SpO₂ signals were obtained from the pulse oximeter at $f_s = 25$ Hz, as recommended by the AASM [14]. Figure 1 shows the workflow of the proposed methodology. After preprocessing, features were extracted using time and frequency-based analyses. This study was intended to assess the complementarity of the features extracted from AF and SpO₂ signals, and therefore different feature sets were assessed: AF-derived features, SpO₂-derived features and both AF and SpO₂ features. We also split the experiments in two situations: with and without ODI 3%. Six settings were thus investigated, namely: 'AF', 'SpO₂', 'AF + SpO₂', 'AF + ODI', 'SpO₂ + ODI' and

'AF + SpO₂ + ODI'. Feature selection was conducted in these feature sets independently to establish optimum subsets of features before the classification stage. Finally, the selected features were used to train and evaluate six independent AdaBoost classifiers.

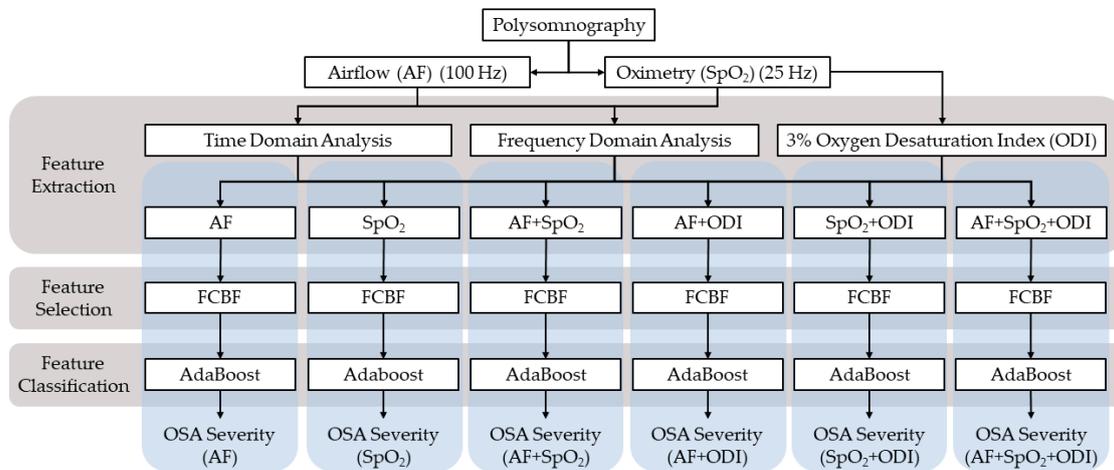


Figure 1. Workflow of the proposed methodology. Fast Correlation-Based Filter (FCBF); Obstructive Sleep Apnea (OSA).

3.1. Preprocessing

AF and SpO₂ signals were preprocessed in order to remove artifacts and signal loss intervals, as well as to normalize the amplitude values. In the case of SpO₂ signals, samples with values lower than 50% of saturation and intervals with abrupt changes of oxygen saturation greater than 4% per second were removed [26,29]. AF signals were filtered using a low-pass filter (cutoff frequency of 1.5 Hz) and subsequently normalized [31,32]. Artifacts in the AF signal were removed using a method based on the standard deviation and the kurtosis of 30 s segments, as in previous studies [32].

3.2. Feature Extraction: Time and Frequency Domain Analyses

The feature extraction stage comprised the characterization of AF and SpO₂ signals using automatic signal processing algorithms. In this study, analyses were performed in time and frequency domains. Extracted features summarize the information about the alterations of the signals properties and the recurrence of apneic events, and have been widely assessed in previous studies dealing with automatic detection of adult and pediatric OSA [19,20,26,31].

3.2.1. Time-Domain Moments and Nonlinear Analysis

Five time domain statistical moments were computed: mean (M1T), standard deviation (M2T), skewness (M3T), kurtosis (M4T) and median (MedT) [20,26]. Besides, three nonlinear features have also been obtained: Central Tendency Measure (CTM) [47], Lempel–Ziv Complexity (LZC) [48] and Sample Entropy (SampEn) [49]. Signals were segmented into epochs of 30 s prior to perform time-domain and nonlinear analyses, and each feature was calculated as the average value across all segments.

CTM is a measure of the variability of a signal [47]. It is based on plots of first order differences: given a signal $x(n)$ of length N , the values $x(n+2) - x(n+1)$ are represented against $x(n+1) - x(n)$ in a scatter plot. CTM is the rate of differences that lie inside a circle of fixed radius r [20,47]:

$$CTM = \frac{\sum_{i=1}^{N-2} \delta(i)}{N-2}; \delta(i) = \begin{cases} 1 & \text{if } \sqrt{[x(i+2) - x(i+1)]^2 + [x(i+1) - x(i)]^2} < r \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The computation of CTM relies on the parameter r . The values of r were independently set for AF and SpO₂ signals by maximizing the absolute value of the Spearman's correlation coefficient (ρ) between CTM and the AHI in the training set [26,31].

LZC is a nonparametric measure of the complexity of a time series. The LZC of a sequence increases as more subsequences are contained in it. To analyze the subsequences, the signal is converted to binary by applying a threshold, usually the median value of the samples [20,26,48]. Then, the binary data is scanned and a counter $c(n)$ is increased as more different sequences are found in the data. LZC is the coefficient [20,48]:

$$LZC = \frac{c(n)}{b(n)}; b(n) = \frac{n}{\log_2 n}. \quad (2)$$

In (2), the normalizing factor $b(n)$ is equal to the theoretical upper bound of $c(n)$. [48].

SampEn is a statistic used to measure irregularity in biomedical signals [49]. It has been widely employed to characterize fluctuations of AF and SpO₂ signals [18,20,27,49]. Given a signal of length N , *SampEn* is defined as the negative logarithm of the conditional probability of two similar sequences of length m remaining similar (distance lower than r) after the length of the sequence increases in one sample (length $m + 1$) [20,49]:

$$SampEn(m, r, N) = -\log \left[\frac{A^m(r)}{B^m(r)} \right], \quad (3)$$

where $A^m(r)$ and $B^m(r)$ are the average number of similar sequences of length $m + 1$ and m , respectively. For each signal, we fixed the optimum values of parameters m and r to those which maximized the absolute value of Spearman's ρ of *SampEn* with the AHI in the training set [26,31]. The trials were set with m in the range 1–3 and r in the range 0.05–0.3 times the standard deviation of the signal [50].

3.2.2. Spectral Analysis

Frequency domain features were obtained after the estimation of the Power Spectral Density (PSD) of the signals using the non-parametric Welch method [51]. The signals were segmented in epochs with 50% overlap using a Hamming window of 2^{14} and 2^{16} samples for SpO₂ and AF, respectively. Window lengths were defined as the minimum power of two that encompasses a segment duration greater than 10 min, a tradeoff between spectral resolution and number of segments [51]. The PSD estimation was then obtained by averaging the PSDs of segments [51].

Once the signals PSDs were estimated, we defined the spectral band of interest (BOI) of the AF signal as the band where the amplitude of the PSD of AF differs among severity groups. In the case of SpO₂ signal, we employed the spectral BOI between 0.020–0.044 Hz defined in previous studies [26]. Following the same methodology, we sought a spectral BOI for the AF signal. The Mann–Whitney U test was used to compare the values of the PSD and find frequency ranges showing the highest statistically significant differences among OSA severity groups [18,26,33]. Figure 2a shows the average PSDs of No OSA, Mild, Moderate and Severe OSA subjects in the training set. The p -values obtained for each frequency are shown in Figure 2b. Only frequencies between 0–0.36 Hz are displayed to allow a proper visualization of the BOI. As four severity groups were involved in the comparison, a total of six pairwise comparisons were conducted. A spectral BOI was found in 0.134–0.176 Hz, where the maximum number of comparisons showed statistically significant differences ($p < 0.05/6$, Bonferroni correction).

Seven features were obtained from the PSD values in the spectral BOI [26,30]: first to fourth statistical moments (M1F–M4F), median (MedF), maximum (MaxF) and minimum (MinF). Additionally, the full spectrum of the signals was characterized by obtaining four features: the median frequency (FreqM), the spectral entropy (SpecEn) and the quadratic and cubic spectral entropies (SpecEn² and

SpecEn³, respectively) [20,31]. FreqM is defined as the frequency (f) that accomplishes that 50% of the total power is below that frequency.

$$\sum_{f=0}^{f=\text{FreqM}} PSD(f) = \frac{1}{2} \sum_{f=0}^{f=f_s/2} PSD(f). \quad (4)$$

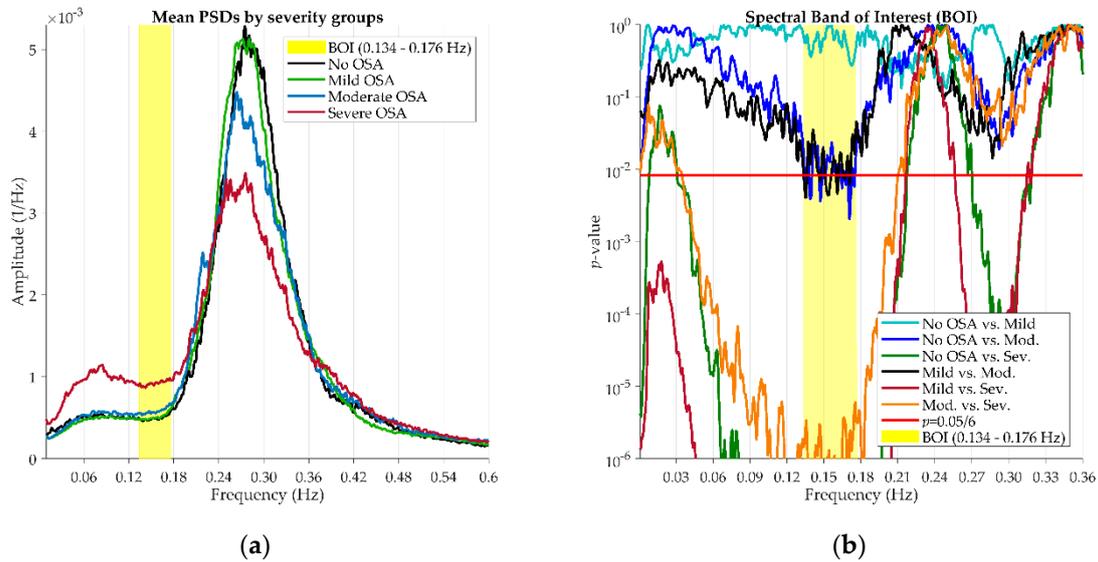


Figure 2. (a) Mean Power Spectral Densities of AF signals for each Obstructive Sleep Apnea (OSA) severity group. (b) Definition of the spectral Band of Interest (BOI).

SpecEnⁱ is defined as the Shannon entropy of the frequency distribution provided by the normalized i -th power of the PSD (PSD_n^i) [31]:

$$\text{SpecEn}^i = -\frac{1}{\log N} \sum_{f=0}^{f=f_s/2} PSD_n^i(f) \cdot \log[PSD_n^i(f)] \quad (i = 1, 2, 3), \quad (5)$$

where N is the number of samples of PSD_n in $0 - f_s/2$. SpecEn indirectly estimates the irregularity of a signal since higher SpecEn values are expected from flatter PSDs, with no dominant frequencies [31].

3.2.3. Oxygen Desaturation Index

Finally, ODI 3% was computed from the SpO₂ signal as the number of desaturations greater than or equal to 3% from the baseline per hour of recording. This oximetric index has been found useful in previous approaches focused on the detection of childhood OSA [25–27,34].

3.3. Feature Selection: Fast Correlation-Based Filter

In the feature extraction stage, 19 signal processing-derived features were extracted for each signal: five time-domain statistics, three nonlinear measures, seven statistics from the BOI and four from the full spectrum measures. ODI 3% was added to these 38 features, so the total number of features was 39. Feature selection was implemented in this study to identify relevant and complementary features of AF and SpO₂ signals and derive simpler models with reduced chances of overfitting [52]. We employed the Fast Correlation-Based Filter (FCBF) method [53] prior to the feature classification stage. This classifier-independent method identifies the most relevant features and removes redundant

ones to obtain the optimum subset of features [53]. FCBF is based on measures of the symmetrical uncertainty (SU) between features X_i and X_j . It is defined as [53]:

$$SU(X_i|X_j) = 2 \frac{H(X_i) - H(X_i|X_j)}{H(X_i) + H(X_j)}, \quad (6)$$

where $H(X_i)$ is the Shannon entropy of X_i , and $H(X_i|X_j)$ is the Shannon entropy of the feature X_i when X_j is observed. The relevance of X_i is defined as $SU(X_i|Y)$ —Being Y the AHI—And redundancy is defined as $SU(X_i|X_j)$. The criteria to remove X_i is the following [53]:

$$SU(X_j|Y) \geq SU(X_i|Y), \text{ and } SU(X_i|X_j) \geq SU(X_i|Y). \quad (7)$$

FCBF was combined with bootstrapping to reduce dependency on the training data and improve generalization [44,52]. We obtained 1000 bootstrap replicates from the training data and the FCBF algorithm was applied to each one [26,29,44]. Features selected at least 500 times formed the optimum subset of features [26,29,52].

3.4. Classification: Multiclass AdaBoost

The classification stage was aimed at predicting the severity of OSA using the features selected in the previous stage. As we described in Section 2, subjects were classified in four groups according to the severity of OSA. We employed the multiclass AdaBoost classifier, an ensemble learning method based on boosting [44,46]. The main idea behind ensembles is to combine several classifiers to build a robust one with an increased generalization ability [44]. Base classifiers used to construct ensembles are usually weak and simple decision rules, such as decision trees or LDA classifiers [46]. The crucial rule of ensembles is diversity, that is, weak classifiers need to be trained with different representations of the training set [46]. This way, each weak classifier becomes an expert in a certain area of the feature space and the ensemble makes its predictions based on a committee of diverse and complementary classifiers [46]. Boosting methods are ensembles characterized by sequential training of base classifiers. At each iteration, a new base classifier is trained giving higher weights to instances in which the previous base classifier failed to make a prediction. After a sufficient number of base classifiers are trained, final predictions are obtained by weighted vote of base classifiers [46].

AdaBoost is the most widespread boosting method [46]. In this study, we employed the algorithm AdaBoost.M2, which allows multiclass classification [54]. We used LDA as base classifier since it was proven useful in previous OSA-related studies [19,22]. The training data comprised N feature vectors, x_i , with labels, y_i ($i = 1, \dots, N$). The number of iterations, L , was experimentally tuned. For each iteration t ($t = 1, \dots, L$), a single base classifier is trained using a version of the training data with weights $w_i(t)$. First, the distribution $D_t(i)$ is calculated as [54]:

$$D_t(i) = \frac{W_i^t}{\sum_{i=1}^N W_i^t}, \quad (8)$$

where with W_i :

$$W_i^t = \sum_{y \neq y_i} w_{i,y}^t. \quad (9)$$

The base classifier is then trained with the distribution $D_t(i)$. The trained base classifier generates a weak prediction $h_t(x,y)$ and the pseudo-loss ε_t is calculated [54]:

$$\varepsilon_t = \frac{1}{2} \sum_{i=1}^N D_t(i) \left[1 - h_t(x_i, y_i) + \sum_{y \neq y_i} \frac{w_{i,y}^t}{W_i^t} h_t(x_i, y) \right]. \quad (10)$$

Then, the weight update coefficient of base classifier t , β_t , is obtained [54]:

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}. \quad (11)$$

Additional regularization was added using a modified β_t with a learning rate parameter ν :

$$\beta_t = \left(\frac{\varepsilon_t}{1 - \varepsilon_t} \right)^\nu, \quad (12)$$

where with ν in the range 0–1. Then, w_i of the instances x_i for the next iteration $t + 1$ are computed as [54]:

$$w_{i,y}^{t+1} = w_{i,y}^t \beta_t^{\frac{1}{2}(1+h_t(x_i,y_i)-h_t(x_i,y))}. \quad (13)$$

The final prediction of AdaBoost $H(x)$ is obtained by means of weighted vote [54]:

$$H(x) = \arg \max_y \sum_{t=1}^L \left(\log \frac{1}{\beta_t} \right) h_t(x, y). \quad (14)$$

3.5. Model Optimization and Training

Our database was split into a training set and a test set. The training set was used to derive the optimal number of iterations (L) and the learning rate (ν) of the AdaBoost algorithm, while the test set was intended to evaluate the models in new data. The hyperparameters L and ν are involved in the number of base classifiers to be trained and the calculation of the weight update coefficient β_t , respectively. We set trials to estimate the performance of AdaBoost in the training set using the Cohen's Kappa (κ) [55], with L in the range between two and 10,000 classifiers and ν in the range 0.1–1. Cohen's κ is less sensitive to class imbalance in comparison with the error rate [55]. We used the 0.632 bootstrap validation method to estimate κ with reduced chances of overfitting [44]. We obtained 1000 new bootstrap replicates from the training data and trained a model for each one. Repeated instances are frequent in a bootstrap replicate, whereas other instances are not selected [44]. Unselected instances formed a validation set used to evaluate the trained model. The estimate of κ using 0.632 bootstrap, $\kappa_B(i)$ ($i = 1, \dots, 1000$), is [44]:

$$\kappa_B(i) = 0.632 \cdot \kappa_{BValidation}(i) + 0.368 \cdot \kappa_{BTraining}(i), \quad (15)$$

where $\kappa_{BValidation}(i)$ and $\kappa_{BTraining}(i)$ are the values of κ obtained when the model is evaluated in the validation and training bootstrap datasets, respectively [44]. The final κ_B is the average of $\kappa_B(i)$ over i [44]. AdaBoost classifiers were trained using the overall training set with optimum L and ν fixed.

3.6. Statistical Analysis

A correlation analysis was conducted in the training set to evaluate the relationship between extracted features and the AHI using the Spearman's ρ . Statistically significant differences between severity groups were also examined in the training set using the Kruskal–Wallis test ($p < 0.01/6$, Bonferroni correction), since features did not pass the Lilliefors normality test. Results obtained in the test set were summarized in a four-class confusion matrix. The agreement between the predicted severity and the gold standard was assessed using the four-class accuracy (Acc_4) and κ . Diagnostic ability in the common AHI cutoffs was evaluated using Sensitivity (Se, percentage of diseased subjects correctly classified), Specificity (Sp, percentage of healthy subjects correctly classified), Accuracy (Acc, percentage of subjects correctly classified), Positive and Negative Predictive Value (PPV, NPV, percentage of subjects correctly classified as positives/negatives) and Positive and Negative Likelihood Ratios ($LR+ = Se/(1 - Sp)$, $LR- = (1 - Se)/Sp$).

4. Results

4.1. Preprocessing. Parameters Optimization in the Training Set

Artifacts in both AF and SpO₂ signals were removed in the preprocessing stage. The rates of rejected data—Median [interquartile range]—Were 5.65% [1.76%, 10.30%] and 5.36% [1.51%, 9.17%] of the total recording time for AF and SpO₂ signals, respectively. The amount of discarded data was low comparing to the length of the overnight recordings and both signals were similarly affected by artifacts ($\rho = 0.5394$, Spearman's rank correlation). In addition, no substantial differences were found between the rates of rejected data of AF and SpO₂ ($p = 0.4175$, Wilcoxon signed rank test). Figure 3 shows the absolute value of the Spearman's ρ of CTM with AHI for varying r in the training set. The maximum values of $|\rho(r)|$ were reached using $r = 0.0004$ in AF and $r = 0.025$ in SpO₂. Following the same criteria for SampEn, the optimum parameters were $m = 2$ and $r = 0.05$ for AF, and $m = 3$ and $r = 0.05$ for SpO₂ (Table 2).

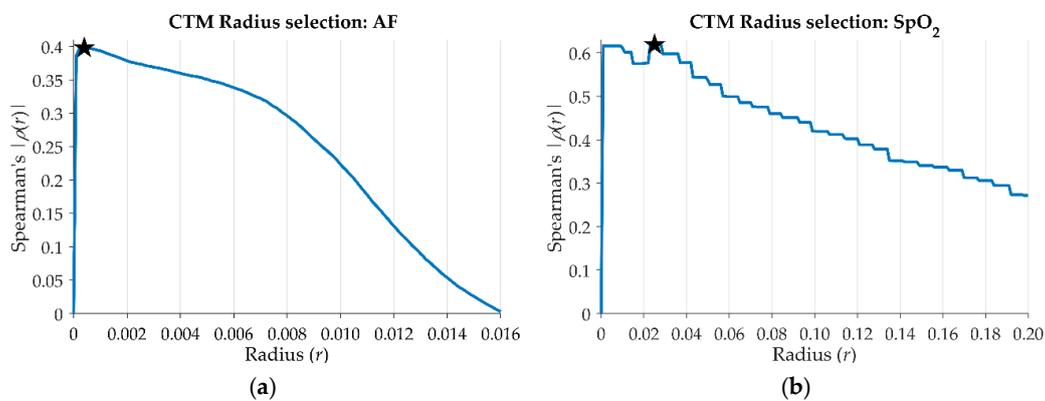


Figure 3. Absolute value of Spearman's correlation coefficient (ρ) between Central Tendency Measure (CTM) and the apnea–hypopnea index as a function of radius (r) in the training set. (a) Airflow (AF) signal; (b) oximetry (SpO₂) signal.

Table 2. Spearman's correlation coefficient of sample entropy with apnea–hypopnea index in the training set for different values of the parameters m and r .

	AF			SpO ₂		
	$m = 1$	$m = 2$	$m = 3$	$m = 1$	$m = 2$	$m = 3$
$r = 0.05$	−0.0872	−0.1187	0.0026	0.5502	0.5516	0.5586
$r = 0.10$	−0.0753	−0.0863	−0.1168	0.5123	0.5118	0.5134
$r = 0.15$	−0.0777	−0.0802	−0.0914	0.4786	0.4786	0.4784
$r = 0.20$	−0.0832	−0.0824	−0.0886	0.4395	0.4381	0.4399
$r = 0.25$	0.0897	−0.0880	−0.0910	0.3895	0.3899	0.3900
$r = 0.30$	0.0983	−0.0951	−0.0966	0.3341	0.3350	0.3367

Airflow (AF) signal; (b) oximetry (SpO₂). Maximum absolute values represented in bold.

4.2. Statistical Analysis in the Training Set: Individual Features

Table 3 summarizes the results of the correlation and statistical differences between severity groups in the training set. Several features extracted from both signals showed significant differences between severity groups. Two nonlinear features from AF as well as some time and frequency-domain measures from both signals showed no statistically significant differences. These features were generally associated with the lowest correlations obtained in this study. In general, SpO₂ features obtained the highest correlations with the AHI, whereas correlations of several AF features were weaker but significant. It is also remarkable that both time and frequency domain analyses showed statistically significant correlations with the AHI. CTM obtained the highest correlation among AF features and the

highest correlations among nonlinear features in both signals. Regarding the spectral analysis-derived features, correlations were also higher in the SpO₂ signal. Nevertheless, SpecEn-derived features showed higher correlations when they were applied in the AF signal in comparison with SpO₂. Overall, ODI 3% achieved the highest correlation with the AHI.

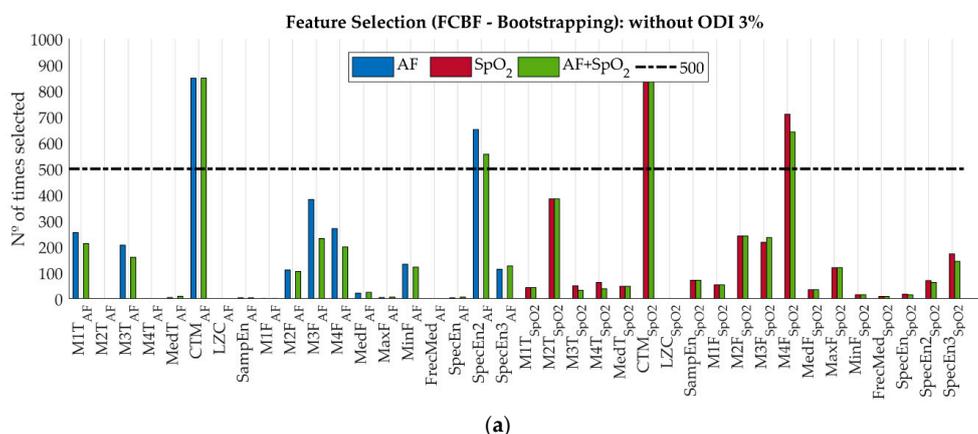
Table 3. Spearman’s correlation coefficients (ρ) with the apnea–hypopnea index and their corresponding p -values of features in the training set and p -values of the Kruskal–Wallis test.

Feature	AF			SpO ₂		
	Spearman		Kruskal–Wallis	Spearman		Kruskal–Wallis
	ρ	p -Value	p -Value	ρ	p -Value	p -Value
M1T	0.1693	<<0.01	0.0061 *	−0.4135	<<0.01	<<0.01
M2T	−0.2481	<<0.01	<<0.01	0.5145	<<0.01	<<0.01
M3T	−0.1655	<<0.01	0.0024 *	−0.1879	<<0.01	<<0.01
M4T	0.3580	<<0.01	<<0.01	0.0968	0.0194	0.0103 *
MedT	0.2070	<<0.01	<<0.01	−0.3467	<<0.01	<<0.01
CTM	0.3979	<<0.01	<<0.01	−0.6187	<<0.01	<<0.01
LZC	−0.0660	0.1111	0.0409 *	0.3871	<<0.01	<<0.01
SampEn	−0.1187	<0.01	0.0270 *	0.5586	<<0.01	<<0.01
M1F	0.3492	<<0.01	<<0.01	0.6773	<<0.01	<<0.01
M2F	0.2979	<<0.01	<<0.01	0.6352	<<0.01	<<0.01
M3F	−0.1418	<<0.01	<<0.01	0.0184	0.6574	0.4893 *
M4F	−0.0967	0.0195	0.0112*	0.0356	0.3899	0.4643 *
MedF	0.3591	<<0.01	<<0.01	0.6753	<<0.01	<<0.01
MaxF	0.3245	<<0.01	<<0.01	0.6646	<<0.01	<<0.01
MinF	0.3588	<<0.01	<<0.01	0.6504	<<0.01	<<0.01
FreqM	−0.1280	<0.01	0.0117 *	0.1209	<0.01	0.0073 *
SpecEn	0.3464	<<0.01	<<0.01	0.0060	0.8842	0.9340 *
SpecEn ²	0.2741	<<0.01	<<0.01	0.1247	<0.01	0.0234 *
SpecEn ³	0.1304	<0.01	0.0024 *	0.1075	<0.01	0.0742 *
ODI 3%	—	—	—	0.6918	<<0.01	<<0.01

*: Not lower than the Bonferroni corrected p -value ($p = 0.01/6$). Airflow (AF) signal; (b) oximetry (SpO₂).

4.3. Feature Selection in the Training Set

Figure 4 shows the histograms of the number of times each feature was selected using different groups of features in the training set: ‘AF’, ‘SpO₂’, ‘AF + SpO₂’, ‘AF + ODI’, ‘SpO₂ + ODI’ and ‘AF + SpO₂ + ODI’. Results of feature selection without ODI 3% are shown in Figure 4a. Selected features from the ‘AF’ (CTM_{AF}, SpecEn²_{AF}) and the ‘SpO₂’ (CTM_{SpO2}, M4F_{SpO2}) sets were selected again using the ‘AF + SpO₂’ set. In this case, no redundant features were found when both signals were combined. Results with ODI 3% are shown in Figure 4b. In these three cases, ODI 3% was found to be the most relevant feature and made SpecEn²_{AF} and CTM_{SpO2} redundant. Furthermore, CTM_{AF}, and M4F_{SpO2} were nonredundant.



(a)

Figure 4. Cont.

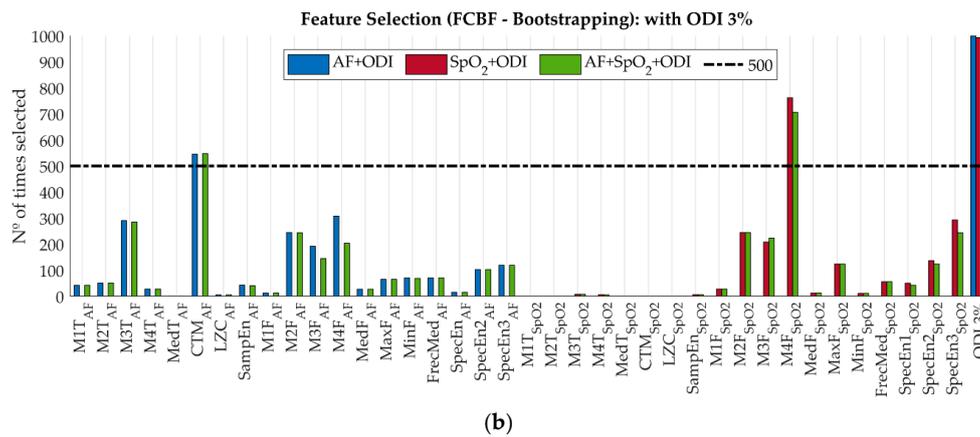


Figure 4. Results of the feature selection in the training set using Fast Correlation-Based Filter (FCBF). (a) sets without ODI 3%; (b) sets with ODI 3%.

4.4. Model Optimization in the Training Set

We trained an independent AdaBoost ensemble model for each of the six subsets of selected features using training data. Hence, different optimum values of L and ν were obtained in each case to optimize the performance. Figure 5 shows the bootstrap estimate of κ in the training set for the corresponding trials. AdaBoost models trained with features from the AF signal did not yield higher κ as L increased, as shown in Figure 5a. In this case, a large value of L was not necessary to retrieve the most useful information from AF. The remaining experiments showed increasing κ as L became higher until the maximum was reached. In general, the optimum κ was reached combining intermediate values of values of L and ν , except for the AF + SpO₂ subset. This setting reached the maximum κ with a large L and the lowest ν —Figure 5c. Nevertheless, differences between the maximum κ for different values of ν were not high.

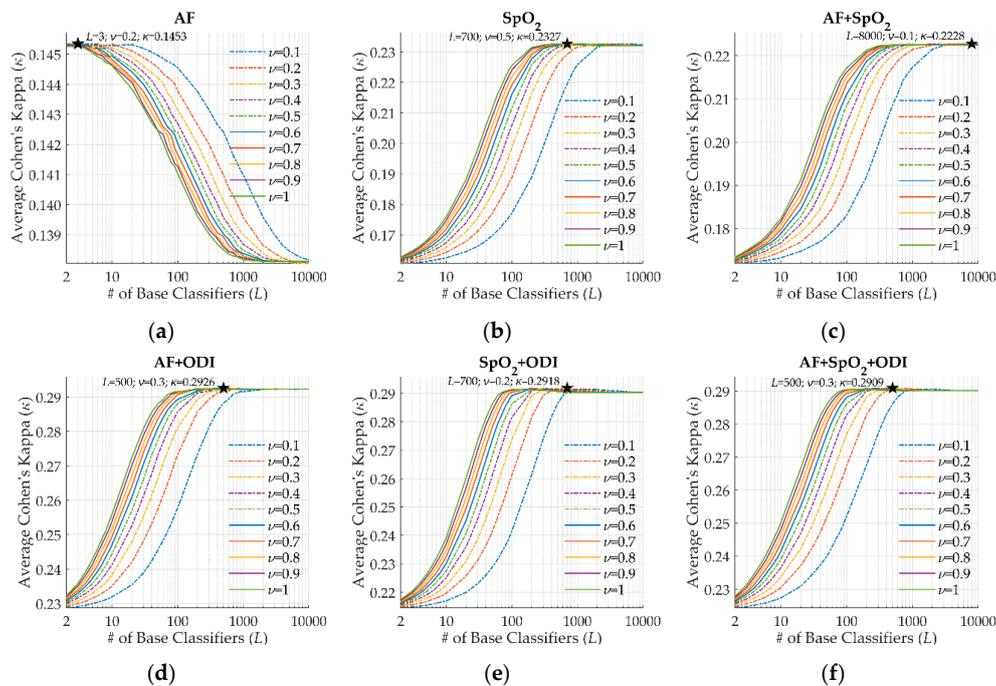


Figure 5. Performance in the training set of AdaBoost models as a function of the number of base classifiers (L) and the learning rate (ν), with their optimum values highlighted. (a) AF subset; (b) SpO₂ subset; (c) AF + SpO₂ subset; (d) AF + ODI subset; (e) SpO₂ + ODI subset; (f) AF + SpO₂ + ODI subset.

4.5. Diagnostic Ability Assessment in the Test Set

Tables 4 and 5 show the confusion matrices along with their respective κ and Acc_4 values obtained on the test set. Besides, the classification results of ODI 3% in the test set are shown in Table 6. Regarding multiclass classification, both Acc_4 and κ increased when features from both signals were combined. The highest performances were obtained when ODI 3% was also included. The highest overall Acc_4 and κ were achieved using the AF + SpO₂ + ODI subset, although the same Acc_4 but slightly lower κ were reached using AF + ODI. It is important to note that AdaBoost models were more accurate than ODI 3%—Except for the AF model.

Table 4. Confusion matrices of the predictions of AdaBoost models in the test set using the subsets AF, SpO₂ and AF + SpO₂.

Severity Levels		AdaBoost (Without ODI 3%)											
		Estimated: AF				Estimated: SpO ₂				Estimated: AF + SpO ₂			
		No	Mild	Mod.	Sev.	No	Mild	Mod.	Sev.	No	Mild.	Mod.	Sev.
Actual	No	1	55	16	3	17	50	8	0	19	47	8	1
	Mild	1	97	53	18	19	119	30	1	21	111	35	2
	Mod.	1	29	22	11	5	29	24	5	6	24	27	6
	Sev.	0	25	25	33	3	12	34	34	2	8	36	37
$Acc_4 = 39.23\%; \kappa = 0.1143$					$Acc_4 = 49.74\%; \kappa = 0.2646$				$Acc_4 = 49.74\%; \kappa = 0.2781$				

Table 5. Confusion matrices of the predictions of AdaBoost models in the test set using the subsets AF + ODI, SpO₂ + ODI and AF + SpO₂ + ODI.

Severity Levels		AdaBoost (With ODI 3%)											
		Estimated: AF + ODI				Estimated: SpO ₂ + ODI				Estimated: AF + SpO ₂ + ODI			
		No	Mild	Mod.	Sev.	No	Mild	Mod.	Sev.	No	Mild	Mod.	Sev.
Actual	No	27	44	3	1	26	45	3	1	28	43	3	1
	Mild	23	115	30	1	23	113	32	1	25	113	30	1
	Mod.	2	24	32	5	7	18	32	6	7	18	33	5
	Sev.	0	9	22	52	1	8	22	52	2	8	21	52
$Acc_4 = 57.95\%; \kappa = 0.3930$				$Acc_4 = 57.18\%; \kappa = 0.3864$				$Acc_4 = 57.95\%; \kappa = 0.3984$					

Table 6. Confusion matrix of the predictions of ODI 3% in the test set.

Severity Levels		ODI 3%			
		Estimated			
		No	Mild	Mod.	Sev.
Actual	No	65	7	1	2
	Mild	110	35	11	13
	Mod.	18	14	8	23
	Sev.	6	6	3	68
$Acc_4 = 45.13\%; \kappa = 0.2833$					

Table 7 shows the diagnostic performance in the test set for each setting in terms of their ability to predict the presence of OSA using the reference AHI cutoffs. Despite the lower κ in the four-class classification task, the AF + ODI subset reached the maximum Acc in all AHI cutoffs: $Acc = 81.28\%$ ($Se = 92.06\%; Sp = 36.00\%$), $Acc = 82.05\%$ ($Se = 76.03\%; Sp = 85.66\%$), and $Acc = 90.26\%$ ($Se = 62.65\%; Sp = 97.72\%$) in 1, 5 and 10 e/h, respectively. These results were the same for the AF + SpO₂ + ODI subset in 5 e/h and 10 e/h, but it reached lower diagnostic performance in 1 e/h.

Therefore, the AF + ODI subset showed the highest diagnostic ability in all AHI cutoffs, outperforming the SpO₂ + ODI and AF + SpO₂ + ODI subset in 1 e/h. Nevertheless, the SpO₂ + ODI model also reached high diagnostic performance.

Table 7. Diagnostic performances of AdaBoost models and ODI 3% in the test set in the apnea–hypopnea index cutoffs 1, 5 and 10 events/hour (e/h).

Cutoff	Subset	Se	Sp	Acc	PPV	NPV	LR+	LR-
1 e/h	AF	99.37%	1.33%	80.51%	80.88%	33.33%	1.0071	0.4762
	SpO ₂	91.43%	22.67%	78.21%	83.24%	38.64%	1.1823	0.3782
	AF + SpO ₂	90.79%	25.33%	78.21%	83.63%	39.58%	1.2160	0.3634
	AF + ODI	92.06%	36.00%	81.28%	85.80%	51.92%	1.4385	0.2205
	SpO ₂ + ODI	90.16%	34.67%	79.49%	85.29%	45.61%	1.3800	0.2839
	AF + SpO ₂ + ODI	89.21%	37.33%	79.23%	85.67%	45.16%	1.4235	0.2891
	ODI 3%	57.46%	86.67%	63.08%	94.76%	32.66%	4.3095	0.4908
5 e/h	AF	62.33%	63.11%	62.82%	50.28%	73.68%	1.6898	0.5969
	SpO ₂	66.44%	84.02%	77.44%	71.32%	80.71%	4.1567	0.3995
	AF + SpO ₂	72.60%	81.15%	77.95%	69.74%	83.19%	3.8511	0.3376
	AF + ODI	76.03%	85.66%	82.05%	76.03%	85.66%	5.3002	0.2799
	SpO ₂ + ODI	76.71%	84.84%	81.79%	75.17%	85.89%	5.0589	0.2745
	AF + SpO ₂ + ODI	76.03%	85.66%	82.05%	76.03%	85.66%	5.3002	0.2799
	ODI 3%	69.86%	88.93%	81.79%	79.07%	83.14%	6.3135	0.3389
10 e/h	AF	39.76%	89.58%	78.97%	50.77%	84.62%	3.8144	0.6725
	SpO ₂	40.96%	98.05%	85.90%	85.00%	86.00%	20.9598	0.6021
	AF + SpO ₂	44.58%	97.07%	85.90%	80.43%	86.63%	15.2062	0.5710
	AF + ODI	62.65%	97.72%	90.26%	88.14%	90.63%	27.4768	0.3822
	SpO ₂ + ODI	62.65%	97.39%	90.00%	86.67%	90.61%	24.0422	0.3835
	AF + SpO ₂ + ODI	62.65%	97.72%	90.26%	88.14%	90.63%	27.4768	0.3822
	ODI 3%	81.93%	87.62%	86.41%	64.15%	94.72%	6.6189	0.2063

5. Discussion

This study aimed to assess AF and SpO₂ signals in the context of pediatric OSA and to evaluate whether these signals can provide complementary information to predict OSA severity in children. Furthermore, the diagnostic ability of multiclass AdaBoost classifiers was evaluated using six different combinations of features extracted from AF and SpO₂ signals. Feature selection revealed that the relevant features from each signal remained non-redundant when both signals were combined, thus suggesting their complementarity. Moreover, the diagnostic ability increased when both signals were combined. Two novel contributions have been introduced in this paper. First, we have compared the diagnostic ability of the automatic signal processing of AF and SpO₂ signals in the context of pediatric OSA. Second, we have designed and validated multiclass AdaBoost classifiers to predict the severity of OSA in children. To the best of our knowledge, this is the first time that AF and SpO₂ signals are jointly evaluated in the context of pediatric OSA detection.

5.1. Feature Extraction and Selection

We characterized AF and SpO₂ signals using time-domain statistics, nonlinear measures and spectral analysis. We defined a BOI in the AF signal between 0.134–0.176 Hz. Previous studies also focused on the analysis of specific BOIs in the context of pediatric OSA. Gutiérrez-Tobal et al. found two spectral BOIs (0.119–0.192 Hz and 0.784–0.890 Hz) using an AHI cutoff of 3 e/h [30]. In our study, however, 1, 5 and 10 e/h cutoffs were used. Our BOI is consistent with the first BOI defined in that work [30] and may be related to the presence of apneic events. Intermittent disruptions of at least two cycles in the normal respiratory flow define apneas and hypopneas and can increase the power in frequencies around and below one half of the normal respiratory frequency. Both BOIs are centered in 0.155 Hz, which is approximately half of the central frequency of the normal respiratory band in children [30]. In contrast, no significant differences between severity groups were found in the PSDs in

higher frequencies. This might be due to the use of three different AHI cutoffs and the analysis of a larger cohort.

Previous studies focused on the spectral analysis of AF signals in the context of pediatric OSA found relevant features from their respective BOIs [30,33]. In our study, features from the BOI were discarded in the feature selection stage due to redundancy. CTM_{AF} and $SpecEn^2_{AF}$ were the most relevant and complementary AF features while time domain statistical moments were found redundant. Previous studies addressing irregularity and variability of AF signals in the context of pediatric OSA reported the positive association of CTM and SpecEn with OSA severity [31]. In this study, CTM_{AF} and $SpecEn^2_{AF}$ were found relevant and nonredundant among AF features, thus reinforcing previous findings.

The correlations with the AHI were higher in SpO_2 -derived features in comparison with AF-derived ones, suggesting that features from the SpO_2 signal were more relevant. However, the majority of SpO_2 features were removed due to redundancy with ODI 3%. Only $M4F_{SpO_2}$ was found non-redundant with CTM_{SpO_2} and ODI 3%. These results confirm that the most useful information of SpO_2 to detect OSA is summarized in ODI 3%. This finding is also supported in previous studies. Hornero et al. [26] assessed a similar set of features from SpO_2 recordings, resulting in ODI 3% and $M3F_{SpO_2}$ being selected. Besides, Vaquerizo-Villar et al. [29] found one SpO_2 -derived feature from Detrended Fluctuation Analysis complementary with ODI 3%.

A novel contribution of this study is the joint assessment of AF and SpO_2 signals using signal processing algorithms. It is remarkable that features from both signals were selected when AF and SpO_2 features were combined. However, two different situations need to be analyzed. When feature selection was conducted on the AF + SpO_2 set, selected features matched the features selected from AF and SpO_2 sets separately. These features were thus non-redundant and may indicate complementarity between both signals. Conversely, $SpecEn^2_{AF}$ and CTM_{SpO_2} were found redundant in settings with ODI 3%. Overall results of feature selection suggest that the information of AF and SpO_2 signals could be complementary. These findings are in accordance with previous studies combining AF-derived features with ODI 3%, which reported not only their complementarity, but also an increase in the diagnostic performance when used together [30,32]. Accordingly, the complementarity of the information from AF and SpO_2 signals in the context of adult OSA [34] is also confirmed in this study using a pediatric population.

5.2. Diagnostic Ability and Comparison with Previous Studies

In this study, novel multiclass AdaBoost classifiers have been introduced to predict OSA severity in children. The highest four-class accuracies were reached using the AF + ODI and the AF + SpO_2 + ODI subsets. These results, together with the low accuracies reached using AF, SpO_2 and ODI 3% separately, suggest that AdaBoost was able to take advantage of the information of AF and SpO_2 signals. Moreover, the most useful information of the SpO_2 seems to be summarized in ODI 3%. It is necessary to note that AF + SpO_2 + ODI reached the highest κ but AF + ODI obtained the same Acc_4 . This slight difference can be related to the calculation of κ , that gives more importance to class imbalance [47]. The AF + SpO_2 + ODI setting was slightly more accurate than AF + ODI classifying actual No OSA and Moderate OSA subjects, which were the least represented groups in our database. Thus, κ was higher in the AF + SpO_2 + ODI subset. Both Acc_4 and κ were slightly lower using SpO_2 + ODI, showing that oximetry alone can also achieve high diagnostic ability by means of AdaBoost. Nevertheless, the number of Moderate and Severe OSA subjects misclassified as No OSA was lower using AF + ODI. Another difference between these settings was observed in the number of overestimated subjects (the predicted severity of OSA was higher than the actual severity), which was also higher using SpO_2 + ODI. The rates of underestimated and overestimated subjects were the most balanced in the AF + SpO_2 + ODI setting: 20.77% and 21.28%, respectively. Using ODI 3% only, 40.26% and 14.62% of the subjects were underestimated and overestimated, respectively. Previous studies reported that ODI 3% alone underestimates the severity of OSA [29,32]. In this study, this tendency was also

observed. On the other hand, the MLP neural networks used in previous approaches tended to overestimate the severity of OSA [29,32,42]. Vaquerizo-Villar et al. reported 12.75% of underestimated subjects and 27.30% of overestimated patients [29], while in Xu et al. the rates of underestimated and overestimated severity were 15.05% and 31.25%, respectively [42]. In our study, this behavior was not observed, since AdaBoost achieved a more balanced ratio of underestimated and overestimated subjects. These AdaBoost models were aimed at predicting OSA in a pediatric population. PSG data from boys and girls up to 13 years old was equally distributed in training and test sets. In general, no significant differences ($p > 0.01$) were found in age, sex or BMI z-score between patients correctly and incorrectly classified. We only found some differences in age and BMI z-score between rightly and incorrectly classified patients, which were limited to Mild OSA patients. Overall, diagnostic performances seem not to be biased towards any specific age, sex or BMI subgroup.

Regarding the results of binary classification, the top performing subset in 1 e/h was AF + ODI, reaching the highest Acc and NPV as well as the lowest LR-. These results suggest that AF + ODI is more suitable to discard the presence of OSA in 1 e/h since it was able to reduce false negatives. In comparison with AF + SpO₂ + ODI, Se was higher and Sp slightly lower in AF + ODI. Nevertheless, differences were not high. AF + ODI and AF + SpO₂ + ODI obtained the same diagnostic performance in 5 and 10 e/h. These settings obtained the highest Acc and the most balanced PPV and NPV in both cutoffs. Moreover, the value of LR+ in 10 e/h is remarkable since it indicates a very high likelihood when the AdaBoost model predicts a subject as Severe OSA. The differences between AF + ODI, SpO₂ + ODI and AF + SpO₂ + ODI were not high, which might suggest that the benefits of including AF are minor. Nevertheless, the diagnostic performance of models combining ODI 3% and AF reflects the complementarity of both signals. The contribution of AF reduced the number of false positives in 1 e/h using AF + ODI, which may compensate the added complexity and inconvenience of recording AF in children. Previous studies have successfully evaluated the usefulness of simplified devices to detect pediatric OSA using AF and SpO₂ [10,11]. To the best of our knowledge, this is the first study that jointly evaluates the diagnostic ability of AF and SpO₂ signals in children using signal processing methods. It would be convenient to enhance the diagnostic ability of these signals using signal processing methods alternative to those used in this study [29,41].

Table 8 summarizes the results achieved in previous studies focused on OSA detection in children. Simple and widespread binary classifiers (e.g., LR, LDA) were assessed in shorter cohorts, while MLP neural networks were proposed in studies comprising a larger number of subjects and using holdout validation (i.e., training and test sets). Most of the studies employed the 5 e/h AHI cutoff for binary classification, with Acc in the range 76.0–86.6%. Our proposal reached the highest Acc in 5 e/h among approaches using three AHI cutoffs. Moreover, it was close to the highest Acc among binary classifiers. In this study, Sp in 5 e/h was close to the highest ones in comparison with both binary and multiclass approaches, while Se was similar to those reached by MLP-based methods. Nevertheless, methods with higher Sp also exhibited lower Se. Fewer studies assessed their diagnostic ability in 1 e/h and 10 e/h cutoffs. Some of them developed independent binary LR models for each cutoff, while others relied on MLP neural networks. The former group reached more balanced Se-Sp pairs in both cutoffs, while the latter achieved higher Acc. The AdaBoost classifiers evaluated in this study also reached high Acc in all cutoffs. Other MLP-based approaches tended to overestimate OSA severity, resulting in low Sp in 1 e/h [29,32,42]. Our multiclass AdaBoost classifier achieved a higher Sp while maintaining high Se in 1 e/h using AF + ODI. On the other hand, Acc in 1 e/h was close to those reached using MLP networks. Therefore, a smaller proportion of symptomatic children without polysomnographically diagnosed OSA would be incorrectly diagnosed as suffering from OSA in comparison with other studies. Overall, the results of this study suggest that our ensemble learning-based approach succeeded in achieving high diagnostic ability. The performance of our AdaBoost-based approach strengthens the usefulness of ensemble learning as a valid alternative to other machine learning algorithms.

Table 8. Diagnostic performances of state-of-the-art approaches in the context of childhood Obstructive Sleep Apnea syndrome.

Study	N	Signal	Methods (Extraction/Selection/Classification)	Validation	Cutoff	Se	Sp	Acc
Chang et al. (2013) [36]	141	SpO ₂	ODI, questionnaires/-/LR	–	5	60.0	86.0	76.6
Wu et al. (2017) [37]	311	—	Clinical parameters/-/Stepwise LR	Holdout	5	94.8	25.0	78.2
Gil et al. (2010) [38]	21	PPG	DAP events, HRV, PTTV/Wrapper/LDA	–	5	75.0	85.7	80.0
Lázaro et al. (2014) [39]	21	PPG	DAP events, spectral analysis of PRV/Wrapper/LDA	–	5	100	71.4	86.6
Garde et al. (2014) [23]	146	SpO ₂ , PRV	Time, frequency, nonlinear/-/LDA	Four-fold	5	88.4	83.6	84.9
Garde et al. (2019) [24]	207	SpO ₂ , PRV	Time, frequency, ODI (SpO ₂); standard spectral bands (PRV)/-/LR (3 binary models)	Holdout	1 5 10	68.0 58.0 90.0	86.0 89.0 87.0	71.0 78.0 88.0
Álvarez et al. (2018) [28]	142	SpO ₂	Time domain, ODI, symbolic dynamics/FSLR/LR	Bootstrap	5	73.5	89.5	83.3
Barroso-García et al. (2017) [31]	501	AF	CTM and SpecEn/FSLR/LR (3 binary models)	Holdout	1 5 10	60.5 65.0 83.3	58.6 80.6 79.0	60.0 76.0 80.0
Crespo et al. (2018) [40]	176	SpO ₂	Time, frequency, nonlinear, ODI/FCBF/LDA, QDA, LR (3 binary models)	Bootstrap	1 5	93.9 70.0	37.8 91.4	84.3 82.7
Hornero et al. (2017) [26]	4191	SpO ₂	Time, frequency, nonlinear, ODI/FCBF/MLP regression	Holdout	1 5 10	84.0 68.2 68.7	53.2 87.2 94.1	75.2 81.7 90.2
Xu et al. (2019) [42]	432	SpO ₂	ODI, M3F/-/MLP regression	Direct validation	1 5 10	95.3 77.8 73.5	19.1 80.5 92.7	79.6 79.4 88.2
Vaquerizo-Villar et al. (2018) [29]	981	SpO ₂	DFA, ODI/FCBF/MLP regression	Holdout	1 5 10	97.1 78.8 77.1	23.3 83.7 94.8	82.7 81.9 91.1
Barroso-García et al. (2020) [32]	946	AF, ODI	Recurrence plots, ODI/FCBF/Bayesian MLP regression	Holdout	1 5 10	97.7 78.7 78.8	22.2 78.3 94.3	83.2 78.5 91.0
This Study	974	AF, SpO ₂	Time, Frequency, Nonlinear, ODI/FCBF/Multiclass AdaBoost	Holdout	1 5 10	92.1 76.0 62.7	36.0 85.7 97.7	81.3 82.1 90.3

Airflow signal (AF); Central Tendency Measure (CTM); Decreases in Amplitude of Plethysmography (DAP); Detrended Fluctuation Analysis (DFA); Fast Correlation-Based Filter (FCBF); Forward Stepwise Logistic Regression (FSLR); Heart Rate Variability (HRV); Linear Discriminant Analysis (LDA); Logistic Regression (LR); third order moment in frequency domain (M3F); Multilayer Perceptron (MLP); number of subjects (N); Oxygen Desaturation Index (ODI); Photoplethysmography (PPG); Pulse Rate Variability (PRV); Pulse Transit Time Variability (PTTV); Quadratic Discriminant Analysis (QDA); Spectral Entropy (SpecEn); oxygen saturation signal (SpO₂).

5.3. Limitations and Future Work

In spite of the promising performance of our proposal, some limitations and future investigations have to be pointed out. The database employed in this study comprised 974 subjects. Although this cohort is large, all subjects were recruited in the same center. It would be desirable to expand our database including new subjects from different sleep laboratories to further generalize our results. Secondly, we successfully evaluated AF and SpO₂ signals separately and jointly in the context of childhood OSA detection. Future investigations may rely on potential incorporation of other useful signals included in the PSG. In this sense, the AF signals employed in this study were recorded using a thermistor. Comparison between nasal pressure sensor and thermistor AF signals would also constitute a future goal. In addition, the AF-derived inter breath interval series can be considered for future studies to enhance the diagnostic ability of AF. Signals have been characterized using widespread signal processing methods in the context of OSA. Future work may comprise alternative approaches like bispectrum and wavelets, as well as other nonlinear analyses. Finally, although AdaBoost classifiers yielded high diagnostic performance, other ensemble learning methodologies like bagging or stacking can also be assessed to compare their diagnostic performance using SpO₂ and AF signals.

6. Conclusions

The results of this study showed the usefulness of the joint analysis of AF and SpO₂ signals in the context of pediatric OSA. A remarkable diagnostic performance was achieved using a multiclass AdaBoost classifier fed with a combination of relevant and complementary information from both signals. The most accurate AdaBoost model successfully combined CTM_{AF} with ODI 3%, which was found the most useful parameter of the SpO₂ signal. This joint model outperformed the diagnostic ability of each of these signals separately. Furthermore, we derived an accurate and unbiased AdaBoost model able to decrease the underestimation of the OSA severity previously observed in ODI 3%. Our dual-channel approach is thus a potential alternative to single-channel methodologies, one that might be useful to deploy in the context of simplified screening methods aimed at detecting OSA in children.

Author Contributions: Conceptualization, J.J.-G., G.C.G.-T. and M.G.; methodology, J.J.-G., G.C.G.-T. and D.Á.; software, J.J.-G., G.C.G.-T. and A.M.-M.; validation, J.J.-G., G.C.G.-T. and D.Á.; formal analysis, J.J.-G., G.C.G.-T., M.G. and D.Á.; investigation, J.J.-G.; resources, L.K.-G., F.d.C. and R.H.; data curation, J.J.-G. and A.M.-M.; writing—original draft preparation, J.J.-G., G.C.G.-T., M.G.; writing—review and editing, J.J.-G., G.C.G.-T., M.G., L.K.-G., A.M.-M., D.Á., F.d.C., D.G. and R.H.; supervision, R.H.; project administration, F.d.C., D.G. and R.H.; funding acquisition, D.G. and R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the ‘Ministerio de Ciencia, Innovación y Universidades’ and ‘European Regional Development Fund (FEDER)’ under projects DPI2017-84280-R and RTC-2017-6516-1, by ‘European Commission’ and ‘FEDER’ under projects ‘Análisis y correlación entre el genoma completo y la actividad cerebral para la ayuda en el diagnóstico de la enfermedad de Alzheimer’ and ‘Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres’ (‘Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020’), and by ‘CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)’ through ‘Instituto de Salud Carlos III’ co-funded with FEDER funds. J.J.-G. was in receipt of a ‘Ayudas para la contratación de personal técnico de apoyo a la investigación’ grant from the ‘Junta de Castilla y León’ funded by the European Social Fund and Youth Employment Initiative. A.M.-M. was in receipt of a “Ayudas para contratos predoctorales para la Formación de Doctores” grant from the Ministerio de Ciencia, Innovación y Universidades (PRE2018-085219). D.G. and L.K.-G. are supported by US National Institutes of Health grants HL130984 (L.K.-G.) and HL140548 (D.G.).

Acknowledgments: The authors acknowledge Verónica Barroso-García and Fernando Vaquerizo-Villar their support during the development of the methodology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tauman, R.; Gozal, D. Obstructive sleep apnea syndrome in children. *Expert Rev. Respir. Med.* **2011**, *5*, 425–440. [[CrossRef](#)] [[PubMed](#)]
2. Marcus, C.L.; Brooks, L.J.; Ward, S.D.; Draper, K.A.; Gozal, D.; Halbower, A.C.; Jones, J.; Lehmann, C.; Schechter, M.S.; Sheldon, S.; et al. Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome. *Pediatrics* **2012**, *130*, e714–e755. [[CrossRef](#)] [[PubMed](#)]
3. DelRosso, L.M. Epidemiology and Diagnosis of Pediatric Obstructive Sleep Apnea. *Curr. Probl. Pediatr. Adolesc. Health Care* **2016**, *46*, 2–6. [[CrossRef](#)] [[PubMed](#)]
4. Hunter, S.J.; Gozal, D.; Smith, D.L.; Philby, M.F.; Kaylegian, J.; Kheirandish-Gozal, L. Effect of Sleep-disordered Breathing Severity on Cognitive Performance Measures in a Large Community Cohort of Young School-aged Children. *Am. J. Respir. Crit. Care Med.* **2016**, *194*, 739–747. [[CrossRef](#)] [[PubMed](#)]
5. Blechner, M.; Williamson, A.A. Consequences of Obstructive Sleep Apnea in Children. *Curr. Probl. Pediatr. Adolesc. Health Care* **2016**, *46*, 19–26. [[CrossRef](#)]
6. Brockmann, P.E.; Schaefer, C.; Poets, A.; Poets, C.F.; Urschitz, M.S. Diagnosis of obstructive sleep apnea in children: A systematic review. *Sleep Med. Rev.* **2013**, *17*, 331–340. [[CrossRef](#)]
7. Tan, H.-L.; Kheirandish-Gozal, L.; Gozal, D. Pediatric Home Sleep Apnea Testing. *Chest* **2015**, *148*, 1382–1395. [[CrossRef](#)]
8. Tan, H.-L.; Bandla, H.P.R.; Ramirez, H.M.; Gozal, D.; Kheirandish-Gozal, L. Overnight Polysomnography versus Respiratory Polygraphy in the Diagnosis of Pediatric Obstructive Sleep Apnea. *Sleep* **2014**, *37*, 255–260. [[CrossRef](#)]

9. Kaditis, A.G.; Alonso Alvarez, M.L.; Boudewyns, A.; Alexopoulos, E.I.; Ersu, R.; Joosten, K.; Larramona, H.; Miano, S.; Narang, I.; Trang, H.; et al. Obstructive sleep disordered breathing in 2- to 18-year-old children: Diagnosis and management. *Eur. Respir. J.* **2016**, *47*, 69–94. [[CrossRef](#)]
10. Alonso-Álvarez, M.L.; Terán-Santos, J.; Ordax Carbajo, E.; Cordero-Guevara, J.A.; Navazo-Egüia, A.I.; Kheirandish-Gozal, L.; Gozal, D. Reliability of Home Respiratory Polygraphy for the Diagnosis of Sleep Apnea in Children. *Chest* **2015**, *147*, 1020–1028. [[CrossRef](#)]
11. Stehling, F.; Keull, J.; Olivier, M.; Große-Onnebrink, J.; Mellies, U.; Stuck, B.A. Validation of the screening tool ApneaLink® in comparison to polysomnography for the diagnosis of sleep-disordered breathing in children and adolescents. *Sleep Med.* **2017**, *37*, 13–18. [[CrossRef](#)] [[PubMed](#)]
12. Kaditis, A.; Kheirandish-Gozal, L.; Gozal, D. Pediatric OSAS: Oximetry can provide answers when polysomnography is not available. *Sleep Med. Rev.* **2016**, *27*, 96–105. [[CrossRef](#)]
13. Van Eyck, A.; Verhulst, S.L. Improving the diagnosis of obstructive sleep apnea in children with nocturnal oximetry-based evaluations. *Expert Rev. Respir. Med.* **2018**, *12*, 165–167. [[CrossRef](#)] [[PubMed](#)]
14. Berry, R.B.; Budhiraja, R.; Gottlieb, D.J.; Gozal, D.; Iber, C.; Kapur, V.K.; Marcus, C.L.; Mehra, R.; Parthasarathy, S.; Quan, S.F.; et al. Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J. Clin. Sleep Med.* **2012**, *8*, 597–619. [[CrossRef](#)] [[PubMed](#)]
15. Uddin, M.B.; Chow, C.M.; Su, S.W. Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: A systematic review. *Physiol. Meas.* **2018**, *39*, 03TR01. [[CrossRef](#)] [[PubMed](#)]
16. Mendonça, F.; Mostafa, S.S.; Ravelo-García, A.G.; Morgado-Dias, F.; Penzel, T. A Review of Obstructive Sleep Apnea Detection Approaches. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 825–837. [[CrossRef](#)]
17. del Campo, F.; Crespo, A.; Cerezo-Hernández, A.; Gutiérrez-Tobal, G.C.; Hornero, R.; Álvarez, D. Oximetry use in obstructive sleep apnea. *Expert Rev. Respir. Med.* **2018**, *12*, 665–681. [[CrossRef](#)]
18. Gutiérrez-Tobal, G.C.; Álvarez, D.; Marcos, J.V.; Del Campo, F.; Hornero, R. Pattern recognition in airflow recordings to assist in the sleep apnoea-hypopnoea syndrome diagnosis. *Med. Biol. Eng. Comput.* **2013**, *51*, 1367–1380. [[CrossRef](#)]
19. Gutierrez-Tobal, G.C.; Alvarez, D.; del Campo, F.; Hornero, R. Utility of AdaBoost to Detect Sleep Apnea-Hypopnea Syndrome From Single-Channel Airflow. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 636–646. [[CrossRef](#)]
20. Álvarez, D.; Hornero, R.; Víctor Marcos, J.; Delcampo, F. Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2816–2824. [[CrossRef](#)] [[PubMed](#)]
21. Marcos, J.V.; Hornero, R.; Álvarez, D.; Aboy, M.; Del Campo, F. Automated Prediction of the Apnea-Hypopnea Index from Nocturnal Oximetry Recordings. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 141–149. [[CrossRef](#)] [[PubMed](#)]
22. Gutiérrez-Tobal, G.C.; Álvarez, D.; Crespo, A.; Del Campo, F.; Hornero, R. Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity From At-Home Oximetry Recordings. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 882–892. [[CrossRef](#)] [[PubMed](#)]
23. Garde, A.; Dehkordi, P.; Karlen, W.; Wensley, D.; Ansermino, J.M.; Dumont, G.A. Development of a Screening Tool for Sleep Disordered Breathing in Children Using the Phone Oximeter™. *PLoS ONE* **2014**, *9*, e112959. [[CrossRef](#)]
24. Garde, A.; Hoppenbrouwer, X.; Dehkordi, P.; Zhou, G.; Rollinson, A.U.; Wensley, D.; Dumont, G.A.; Ansermino, J.M. Pediatric pulse oximetry-based OSA screening at different thresholds of the apnea-hypopnea index with an expression of uncertainty for inconclusive classifications. *Sleep Med.* **2019**, *60*, 45–52. [[CrossRef](#)]
25. Álvarez, D.; Alonso-Álvarez, M.L.; Gutiérrez-Tobal, G.C.; Crespo, A.; Kheirandish-Gozal, L.; Hornero, R.; Gozal, D.; Terán-Santos, J.; Del Campo, F. Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polygraphy in Unattended Settings. *J. Clin. Sleep Med.* **2017**, *13*, 693–702. [[CrossRef](#)] [[PubMed](#)]
26. Hornero, R.; Kheirandish-Gozal, L.; Gutiérrez-Tobal, G.C.; Philby, M.F.; Alonso-Álvarez, M.L.; Álvarez, D.; Dayyat, E.A.; Xu, Z.; Huang, Y.-S.; Tamae Kakazu, M.; et al. Nocturnal Oximetry-based Evaluation of Habitually Snoring Children. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, 1591–1598. [[CrossRef](#)]
27. Crespo, A.; Álvarez, D.; Gutiérrez-Tobal, G.C.; Vaquerizo-Villar, F.; Barroso-García, V.; Alonso-Álvarez, M.L.; Terán-Santos, J.; Hornero, R.; Campo, F.D. Multiscale Entropy Analysis of Unattended Oximetric Recordings to Assist in the Screening of Paediatric Sleep Apnoea at Home. *Entropy* **2017**, *19*, 284. [[CrossRef](#)]

28. Álvarez, D.; Crespo, A.; Vaquerizo-Villar, F.; Gutiérrez-Tobal, G.C.; Cerezo-Hernández, A.; Barroso-García, V.; Ansermino, J.M.; Dumont, G.A.; Hornero, R.; del Campo, F.; et al. Symbolic dynamics to enhance diagnostic ability of portable oximetry from the Phone Oximeter in the detection of paediatric sleep apnoea. *Physiol. Meas.* **2018**, *39*, 104002. [[CrossRef](#)]
29. Vaquerizo-Villar, F.; Álvarez, D.; Kheirandish-Gozaal, L.; Gutiérrez-Tobal, G.C.; Barroso-García, V.; Crespo, A.; del Campo, F.; Gozal, D.; Hornero, R. Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis. *Physiol. Meas.* **2018**, *39*, 114006. [[CrossRef](#)]
30. Gutiérrez-Tobal, G.C.; Alonso-Álvarez, M.L.; Álvarez, D.; del Campo, F.; Terán-Santos, J.; Hornero, R. Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home. *Biomed. Signal Process. Control* **2015**, *18*, 401–407. [[CrossRef](#)]
31. Barroso-García, V.; Gutiérrez-Tobal, G.; Kheirandish-Gozaal, L.; Álvarez, D.; Vaquerizo-Villar, F.; Crespo, A.; del Campo, F.; Gozal, D.; Hornero, R. Irregularity and Variability Analysis of Airflow Recordings to Facilitate the Diagnosis of Paediatric Sleep Apnoea-Hypopnoea Syndrome. *Entropy* **2017**, *19*, 447. [[CrossRef](#)]
32. Barroso-García, V.; Gutiérrez-Tobal, G.C.; Kheirandish-Gozaal, L.; Álvarez, D.; Vaquerizo-Villar, F.; Núñez, P.; del Campo, F.; Gozal, D.; Hornero, R. Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis. *Comput. Methods Programs Biomed.* **2020**, *183*, 105083. [[CrossRef](#)] [[PubMed](#)]
33. Jiménez-García, J.; Gutiérrez-Tobal, G.C.; García, M.; Álvarez, D.; Barroso-García, V.; Vaquerizo-Villar, F.; Martín-Montero, A.; Del Campo, F.; Kheirandish-Gozaal, L.; Gozal, D.; et al. Evaluación de la información espectral de las señales de flujo aéreo y saturación de oxígeno en sangre para la ayuda al diagnóstico de la apnea del sueño infantil. In Proceedings of the XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, Santander, Spain, 27–29 November 2019; pp. 25–28.
34. Álvarez, D.; Cerezo-Hernández, A.; Crespo, A.; Gutiérrez-Tobal, G.C.; Vaquerizo-Villar, F.; Barroso-García, V.; Moreno, F.; Arroyo, C.A.; Ruiz, T.; Hornero, R.; et al. A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow. *Sci. Rep.* **2020**, *10*, 1–12. [[CrossRef](#)] [[PubMed](#)]
35. Rosen, C.L.; D'andrea, L.; Haddad, G.G. Adult Criteria for Obstructive Sleep Apnea Do Not Identify Children with Serious Obstruction. *Am. Rev. Respir. Dis.* **1992**, *146*, 1231–1234. [[CrossRef](#)]
36. Chang, L.; Wu, J.; Cao, L. Combination of symptoms and oxygen desaturation index in predicting childhood obstructive sleep apnea. *Int. J. Pediatr. Otorhinolaryngol.* **2013**, *77*, 365–371. [[CrossRef](#)]
37. Wu, D.; Li, X.; Guo, X.; Qin, J.; Li, S. A simple diagnostic scale based on the analysis and screening of clinical parameters in paediatric obstructive sleep apnoea hypopnea syndrome. *J. Laryngol. Otol.* **2017**, *131*, 363–367. [[CrossRef](#)] [[PubMed](#)]
38. Gil, E.; Bailon, R.; Vergara, J.M.; Laguna, P. PTT Variability for Discrimination of Sleep Apnea Related Decreases in the Amplitude Fluctuations of PPG Signal in Children. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1079–1088. [[CrossRef](#)] [[PubMed](#)]
39. Lazaro, J.; Gil, E.; Vergara, J.M.; Laguna, P. Pulse Rate Variability Analysis for Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of Pulse Photoplethysmographic Signal in Children. *IEEE J. Biomed. Heal. Inform.* **2014**, *18*, 240–246. [[CrossRef](#)]
40. Crespo, A.; Álvarez, D.; Kheirandish-Gozaal, L.; Gutiérrez-Tobal, G.C.; Cerezo-Hernández, A.; Gozal, D.; Hornero, R.; del Campo, F. Assessment of oximetry-based statistical classifiers as simplified screening tools in the management of childhood obstructive sleep apnea. *Sleep Breath.* **2018**, *22*, 1063–1073. [[CrossRef](#)]
41. Vaquerizo-Villar, F.; Álvarez, D.; Kheirandish-Gozaal, L.; Gutiérrez-Tobal, G.C.; Barroso-García, V.; Crespo, A.; del Campo, F.; Gozal, D.; Hornero, R. Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome. *PLoS ONE* **2018**, *13*, e0208502. [[CrossRef](#)]
42. Xu, Z.; Gutiérrez-Tobal, G.C.; Wu, Y.; Kheirandish-Gozaal, L.; Ni, X.; Hornero, R.; Gozal, D. Cloud algorithm-driven oximetry-based diagnosis of obstructive sleep apnoea in symptomatic habitually snoring children. *Eur. Respir. J.* **2019**, *53*, 1801788. [[CrossRef](#)] [[PubMed](#)]
43. Vaquerizo-Villar, F.; Álvarez, D.; Kheirandish-Gozaal, L.; Gutiérrez-Tobal, G.C.; Barroso-García, V.; Crespo, A.; del Campo, F.; Gozal, D.; Hornero, R. Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings. *Comput. Methods Programs Biomed.* **2018**, *156*, 141–149. [[CrossRef](#)] [[PubMed](#)]
44. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann/Elsevier: Burlington, MA, USA, 2011; ISBN 978-0-12-374856-0.

45. Xie, B.; Minn, H. Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 469–477. [[CrossRef](#)] [[PubMed](#)]
46. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2014; ISBN 9781118914540.
47. Cohen, M.E.; Hudson, D.L.; Deedwania, P.C. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Eng. Med. Biol. Mag.* **1996**, *15*, 97–102. [[CrossRef](#)]
48. Lempel, A.; Ziv, J. On the Complexity of Finite Sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81. [[CrossRef](#)]
49. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)]
50. Alcaraz, R.; Rieta, J.J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomed. Signal Process. Control* **2010**, *5*, 1–14. [[CrossRef](#)]
51. Welch, P.D. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [[CrossRef](#)]
52. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
53. Yu, L.; Liu, H. Efficient Feature Selection via Analysis of Relevance and Redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.
54. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
55. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).