Computers in Biology and Medicine 154 (2023) 106549

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Pediatric sleep apnea: Characterization of apneic events and sleep stages using heart rate variability

Adrián Martín-Montero [a,b,*], Pablo Armañac-Julián [b,c], Eduardo Gil [b,c], Leila Kheirandish-Gozal [d], Daniel Álvarez [a,b], Jesús Lázaro [b,c], Raquel Bailón [b,c], David Gozal [d], Pablo Laguna [b,c], Roberto Hornero [a,b], Gonzalo C. Gutiérrez-Tobal [a,b]

[a] Biomedical Engineering Group, University of Valladolid, Valladolid, Spain
[b] CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Valladolid, Spain
[c] Biomedical Signal Interpretation and Computational Simulation (BSICoS) Group, Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain
[d] Department of Child Health, The University of Missouri School of Medicine, Columbia, MO, USA

ARTICLE INFO

Keywords:
Obstructive sleep apnea
Children
Heart rate variability
Apneic events
Sleep stages
Childhood adenotonsillectomy trial

ABSTRACT

Heart rate variability (HRV) is modulated by sleep stages and apneic events. Previous studies in children compared classical HRV parameters during sleep stages between obstructive sleep apnea (OSA) and controls. However, HRV-based characterization incorporating both sleep stages and apneic events has not been conducted. Furthermore, recently proposed novel HRV OSA-specific parameters have not been evaluated. Therefore, the aim of this study was to characterize and compare classic and pediatric OSA-specific HRV parameters while including both sleep stages and apneic events. A total of 1610 electrocardiograms from the Childhood Adenotonsillectomy Trial (CHAT) database were split into 10-min segments to extract HRV parameters. Segments were characterized and grouped by sleep stage (wake, W; non-rapid eye movement, NREMS; and REMS) and presence of apneic events (under 1 apneic event per segment, e/s; 1–5 e/s; 5–10 e/s; and over 10 e/s). NREMS showed significant changes in HRV parameters as apneic event frequency increased, which were less marked in REMS. In both NREMS and REMS, power in BW2, a pediatric OSA-specific frequency domain, allowed for the optimal differentiation among segments. Moreover, in the absence of apneic events, another defined band, BWRes, resulted in best differentiation between sleep stages. The clinical usefulness of segment-based HRV characterization was then confirmed by two ensemble-learning models aimed at estimating apnea-hypopnea index and classifying sleep stages, respectively. We surmise that basal sympathetic activity during REMS may mask apneic events-induced sympathetic excitation, thus highlighting the importance of incorporating sleep stages as well as apneic events when evaluating HRV in pediatric OSA.

1. Introduction

The spectral characteristics of cardiac function differ between wakefulness and sleep and are closely modulated by sleep stage. Typically, there is a sympathetic inhibition along with parasympathetic predominance during non-rapid eye movement sleep (NREMS), leading to a decrease in heart rate (HR) and blood pressure (BP) [1,2]. Conversely, the trends during wakefulness (W) and rapid eye movement sleep (REMS) consist of sympathetic predominance with parasympathetic inhibition, leading to increased heart rate and BP [2]. This modulation of both branches of the autonomic nervous system (ANS) can be non-invasively monitored by heart rate variability (HRV) analyses [3]. Accordingly, previous studies have identified these ANS alterations during sleep as diminished high frequency (HF) HRV activity and elevated low frequency (LF) and LF/HF ratios during W and REMS, and reciprocal changes during NREMS in both healthy adults and children [1,2,4–6].

Together with the changes in the cardiac dynamics that take place through sleep stages, children cardiac behaviors during sleep can also be altered by pediatric obstructive sleep apnea (OSA) [7,8]. In its original description in adults, OSA-related alterations in heart rhythm were defined as a pattern of continuing bradycardia during apneic episodes,

* Corresponding author. Biomedical Engineering Group, Facultad de Medicina, Universidad de Valladolid, Av. Ramón y Cajal 7, 47003, Valladolid, Spain.
E-mail address: adrian.martin@gib.tel.uva.es (A. Martín-Montero).

https://doi.org/10.1016/j.compbiomed.2023.106549
Received 19 September 2022; Received in revised form 19 December 2022; Accepted 11 January 2023
Available online 15 January 2023

followed by an abrupt tachycardia when the apneic events end [9]. In children, these patterns have also been detected, but with a high degree of variability depending on the presence and duration of the apneic events [7,8,10–12]. Thus, these alterations in the healthy ANS performance threaten the cardiac health in children, with pediatric OSA having been linked to increased cardiovascular risks [13,14], thereby emphasizing the usefulness of HRV analysis in the pediatric OSA context.

Previous studies have analyzed pediatric OSA-related alterations by comparing HRV patterns across sleep stages between children suffering from OSA and healthy children [15–20]. However, these studies addressed the differences emerging across the sleep stages but did not assess the specific effects of apneic events [15–17]. Furthermore, some of these reports excluded those segments in the recordings containing apneic episodes from their analyses [18,19,21]. Therefore, HRV assessments that include both sleep stages and apneic events or their absence has not yet been conducted. Furthermore, we have recently defined and characterized three pediatric OSA-specific HRV spectral bands [22]. Those bands have been proposed as an alternative to the classic very low frequency (VLF), LF, and HF ranges in pediatric OSA studies, establishing the changes in the activity of those bands and concomitant changes in OSA severity and resolution of OSA after treatment [23]. Accordingly, the analysis of the evolution of the HRV activity in these specific-spectral bands across sleep stages, and using segments with different number of apneic events, is a natural step forward required to gain insights into the behavior of these novel frequency bands.

Therefore, we hypothesized that HRV analysis at segmented time intervals that segregate and include the information on sleep stages and apneic events and employ analyses of both classic and OSA-specific HRV parameters could reveal previously unknown information of ANS alterations during night. A secondary goal was to evaluate the potential clinical application of the extracted features from these segments for the automatic diagnosis of pediatric OSA and for sleep stage classification.

2. Subjects and signals

This study involved 1610 polysomnographic (PSG) studies from children ages 5–9.9 years from the Childhood Adenotonsillectomy Trial (CHAT) database (Number of Clinical Trial NCT00560859). All the information regarding rationale, design, and primary outcomes of the original CHAT study can be consulted in the published literature [24,25]. The CHAT data is publicly available under request at https://sleepdata.org/datasets/chat. The children included in the CHAT study were referred to a sleep laboratory for a nocturnal PSG due clinical symptoms suggestive of the presence of OSA. The CHAT population involved in the current study was: i) 451 children with OSA included the baseline group, who underwent an initial nocturnal PSG, met the inclusion criteria of the original study, and were randomized to different OSA treatments (see Ref. [25] for more details); ii) 755 children from the nonrandomized group, who did not meet the inclusion criteria of the original study but whose first nocturnal PSG was available; and iii) 404 children from the follow-up group, who completed a second overnight PSG, 7-months after the initial PSG with approximately half having undergo adenotonsillectomy as standard treatment of OSA and the rest being allocated to the watchful waiting group. For the nonrandomized group, 75% of the subjects were assigned to the training set (567 recordings), 12.5% to the validation set (94 recordings), and the remaining 12.5% were assigned to the test set (94 recordings). For the baseline and follow-up groups, 50% of the 404 subjects with a follow-up study were assigned to the training set (202 recordings from each baseline and follow-up), 25% were assigned to the validation set (101 recordings from each baseline and follow-up), and 25% to the test set (101 recordings from each baseline and follow-up). Of note, for each child with both follow-up and baseline recordings, we ensured that both recordings were included in the same group to avoid biases. The remaining 47 recordings from baseline without a follow-up study were assigned to the training set. Demographic and clinical data of the children included in the study are shown in Table 1.

The sleep studies were scored and annotated following the scoring rules established by the American Academy of Sleep Medicine (AASM) [26]. Based on these annotations, the diagnosis of OSA was based on the apnea-hypopnea index (AHI), defined as the number of apneic and hypopneic events per hour of sleep (e/h) [26]. Thus, in this study we assigned each child to one out of four commonly used OSA severity groups, as follows: no OSA (AHI <1 e/h), mild OSA ($1 \leq$ AHI <5 e/h), moderate OSA ($5 \leq$ AHI <10 e/h), and severe OSA (AHI ≥ 10 e/h).

3. Methods

The methods employed herein can be partitioned into three stages. First, we performed a signal processing stage to partition all of the recordings into timed segments with artifact removal and extract the ECG channel for subsequent processing. Secondly, we conducted a feature extraction stage to achieve ANS characterization of the segments included in the study. Finally, we assessed the clinical relevance of the features by evaluating its ability to conduct pediatric OSA diagnosis and to automatically classify sleep stages. Fig. 1 shows a global overview of the protocol followed across the study, from the ECG acquisition to the evaluation of the clinical applicability of the HRV segments characterization based on the two approaches considered. Details of the whole protocol are included in the next subsections.

3.1. Signal processing and segmentation

For HRV analysis, the electrocardiogram (ECG) channel from each PSG was extracted and pre-processed. The ECG signals were originally recorded at sampling frequencies of 200, 256 or 512 Hz. Signals were split into 10-min segments. The duration of these segments was selected since it constitutes a trade-off between a complete description of the VLF fluctuations in the HRV spectral domain and limiting the number of segments with two or more sleep stages. Then, for each segment, a R peak detection algorithm was applied, as proposed by Benitez et al. [27].

Table 1
Demographic and clinical data of the children included in the CHAT database. Data are presented as median [interquartile range] or n (percentage); BMI: body mass index; AHI: apnea–hypopnea index; e/s: apneic events per segment; W: Wake; NREMS: non rapid eye-movement; REMS: rapid eye movement.

	Training set	Validation set	Test set
Subjects (n)	1018	296	296
Age (years)	7.0 [2.1]	6.9 [2.0]	7.0 [2.0]
Males (n)	500 (49.12%)	129 (43.58%)	145 (48.99%)
BMI (kg/m^2)	17.28 [5.81]	17.68 [6.22]	17.09 [6.61]
AHI (e/h)	2.23 [4.15]	3.8 [7.76]	1.46 [2.07]
AHI ≥ 1 (e/h)	771 (75.74%)	246 (83.11%)	249 (84.12%)
AHI ≥ 5 (e/h)	264 (25.93%)	96 (32.43%)	108 (36.49%)
AHI ≥ 10 (e/h)	124 (12.18%)	46 (15.54%)	45 (15.20%)
#Segments < 1 e/s	40105 (79.71%)	11391 (77.69%)	11326 (77.16%)
#Segments 1 to 5 e/s	8419 (16.73%)	2677 (18.26%)	2748 (18.72%)
#Segments 5 to 10 e/s	1023 (2.03%)	371 (2.53%)	394 (2.68%)
#Segments ≥ 10 e/s	770 (1.53%)	223 (1.52%)	211 (1.44%)
#Segments W	8733 (17.36%)	2590 (17.66%)	2604 (17.74%)
#Segments NREMS	34961 (69.48%)	10151 (69.23%)	10177 (69.33%)
< 1 e/s	27766 (79.42%)	7833 (77.17%)	7792 (76.57%)
1 to 5 e/s	6164 (17.63%)	1988 (19.58%)	2036 (20.00%)
5 to 10 e/s	602 (1.72%)	210 (2.07%)	243 (2.39%)
≥ 10 e/s	429 (1.23%)	120 (1.18%)	106 (1.04%)
#Segments REMS	6623 (13.16%)	1921 (13.10%)	1898 (12.93%)
< 1 e/s	3606 (54.45%)	968 (50.39%)	930 (49.00%)
1 to 5 e/s	2255 (34.05%)	689 (35.87%)	712 (37.51%)
5 to 10 e/s	421 (6.35%)	161 (8.38%)	151 (7.96%)
≥ 10 e/s	341 (5.15%)	103 (5.36%)	105 (5.53%)

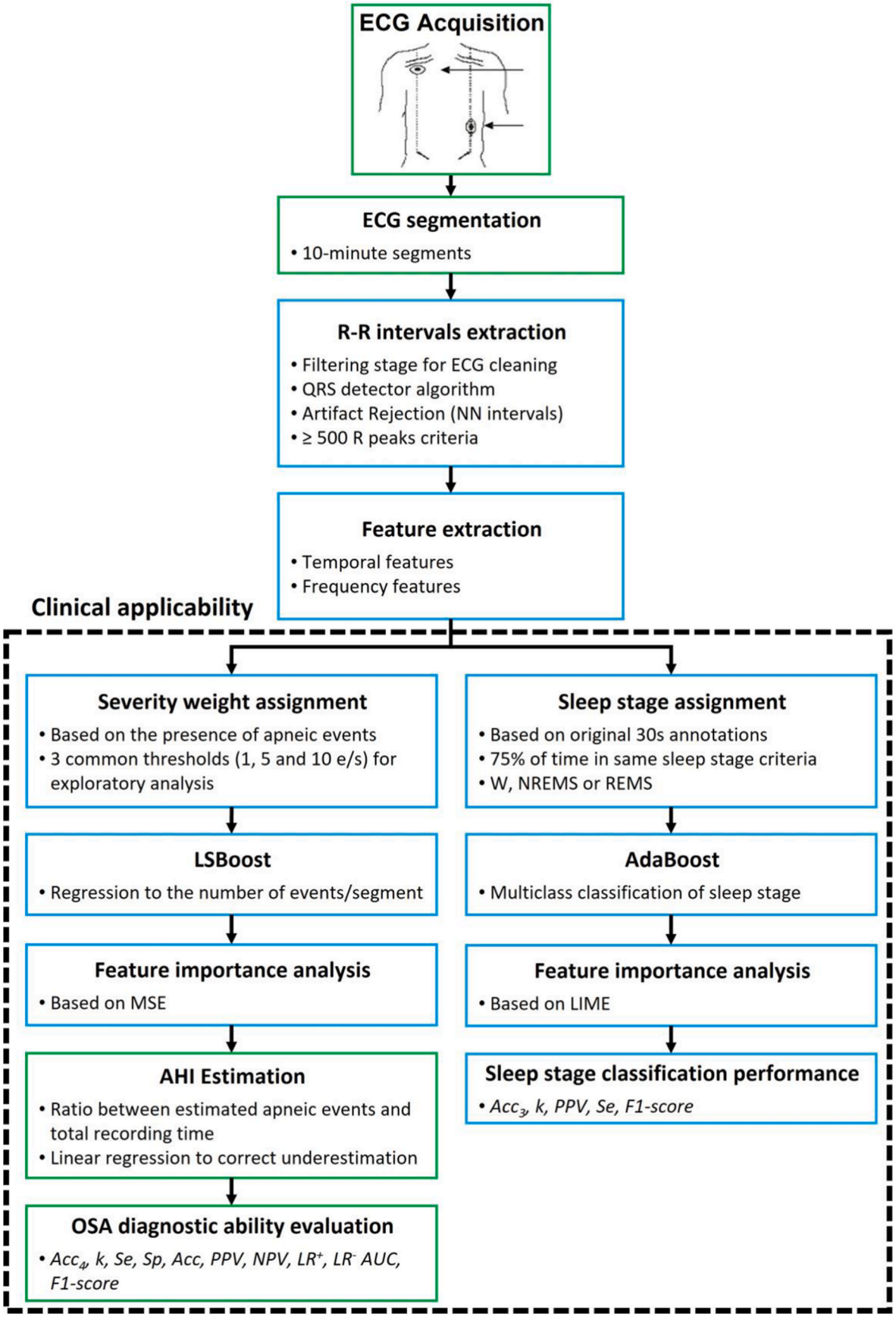

Fig. 1. Flowchart of the protocol followed across the study from the ECG acquisition to the evaluation of the clinical applicability of the characterization of HRV segments following two approaches: OSA diagnostic ability (left branch) and sleep stage classification (right branch). Green boxes means that the corresponding stage of the protocol was applied per subject. Blue boxes means that it was a segment-level stage.

This algorithm has been previously implemented to extract the signals required for HRV characterization in pediatric OSA [22,23,28]. After the R peak detection, we computed the R-R intervals, and conducted an artifact rejection procedure to include only physiologically coherent intervals (N–N intervals). To this effect, R-R intervals that did not fit the following rules were discarded [29]: (i) 0.33 s < R-R interval <1.5 s and (ii) maximum difference between a R-R interval and the previous one of 0.66 s. Following artifact rejection, those segments containing less than 500 normal R peaks were removed. This threshold was selected by establishing 50 beats per minute (bpm) as the minimum heart rate for children between 5 and 10 years old [30].

The analysis of signals in the frequency domain requires a uniform sampling rate. Therefore, after the erroneous intervals removal, each segment was resampled to a constant rate of 3.41 Hz, based on HRV signal interpolation. Also, through this interpolation the erroneous beats that were removed are replaced, fulfilling with the recommendation of the guidelines for the analysis of HRV signal [3]. The sampling rate of 3.41 Hz was selected in order to use a power of two window-length to estimate the power spectral density (PSD) using the Fast Fourier Transform (FFT) [22,23,29]. Then, the normalized PSD (PSDn) to the whole spectrum power of each segment was estimated using a Hamming window of 2^{11} samples (2048 samples), making PSD estimation computationally efficient.

Later, a continuous score was assigned to each segment. Hence, if a 10-min segment contained one apneic event, the assigned score was one. If it contains two and a half apneic events, the score assigned was 2.5, and so on. Then, to conduct the exploration of the pediatric OSA effects in each segment, we assigned them into four severity groups. For this purpose, we adopted the classification commonly established to categorize OSA severity in pediatric subjects [22,31–33]: no apneic segments (<1 apneic event per segment, e/s), 1 e/s ≤ apneic events <5 e/s, 5 e/s ≤ apneic events <10 e/s, and ≥10 e/s.

Finally, we also assigned one of three sleep stages to each 10-min segment (W, NREMS or REMS). To this effect, the original CHAT sleep stage scoring was available. This initial sleep stage annotation was performed by medical experts based on the AASM rules for scoring [26], and following an epoch-by-epoch approach. In the original study, PSG recordings were divided into consecutive 30s epochs, assigning a specific sleep stage to each one of them. Therefore, based on this 30s annotations, we considered that a 10-min segment belongs to a particular sleep stage if 75% of the time was labeled with that stage (7.5 min, i.e., at least 15 of the original 30s epochs). This was a trade-off between reducing the overall number of segments containing sleep stage transitions and also not to increase the complexity introduced by 3 different NREMS sub-stages. Based on these rules a total of 11.4% of the segments could not be assigned to any sleep stage and were discarded from the study. Of note, this percentage is relatively low compared to the large number of segments available in this study.

3.2. ANS characterization of the different type of segments

3.2.1. Time domain features

We used three common time-domain metrics to evaluate differences between HRV activity in each type of segment under study:

- Mean heart rate (*mHR*), measured in bpm.
- Standard deviation of normal-to-normal interval time series (*SDNN*). This parameter globally reflects the variability in the NN time series acting as an overall ANS activity measure [3].
- Root Mean Square of successive differences of NN intervals (*RMSSD*), which is mainly influenced by parasympathetic activity [3].

3.2.2. Frequency domain features

The remaining parameters included in the study were computed in the frequency domain. We have included classical frequency domain HRV measures, as well as relative power (RP) in the three recently defined pediatric OSA-specific spectral bands, all of them extracted from the PSDns:

- RP inside VLF band (RP_{VLF}, 0–0.04 Hz). The physiological meaning of the activity along this band is controversial, hypothesized to be related to slow regulatory mechanism, such as thermoregulation or renin-angiotensin system [34], but the frequency and amplitude of its oscillations could also be influenced by the sympathetic nervous system (SNS) [35]. This parameter has been shown to be highly correlated with *SDNN* [35].
- RP inside LF band (RP_{LF}, 0.04–0.15 Hz). This band may reflect both SNS and parasympathetic nervous system (PNS) activity, and also baroreceptors activity regulating BP [34,35]. LF activity has also been correlated with *SDNN* [35].
- RP inside HF band (RP_{HF}, 0.15–0.40 Hz). HF activity reflects HR changes due to respiratory cycles and has been related to PNS activity, being correlated with *RMSSD* [34,35].
- Normalized power in LF (LF_n), computed as the ratio between RP in LF and the sum of RP inside LF and HF. This parameter is commonly used as an index of sympathovagal balance, as the normalization emphasizes the balanced activation of the two branches of the ANS [3].
- RP inside BW1 band (RP_{BW1}, 0.001–0.005 Hz). This band was the first out of three pediatric OSA-specific frequency bands that we reported in a previous work, and has been linked to macro-sleep disruptions [22].
- RP inside BW2 band (RP_{BW2}, 0.028–0.074 Hz). This pediatric OSA-related frequency band has been linked to the duration and number of apneic events [22], and has been proposed as a potential biomarker of pediatric OSA resolution [23].
- RP inside BWRes band (RP_{BWRes}, adaptive band of ±0.04 Hz around the frequency of the maximum amplitude inside the HF band). This individual adaptive band, designed to consider changes of the respiratory peak inside HF due to age, has been correlated with oxygen desaturations [22].

3.3. Evaluation of potential clinical applicability

In light of the HRV alterations attributable to the presence of OSA in children and the effects of sleep stages, we selected two ensemble-learning boosting algorithms to evaluate the clinical applicability of the HRV segment characterization performed in the previous steps described above: least-squares boosting (LSBoost) for pediatric OSA diagnosis, and adaptive boosting (AdaBoost) for sleep stage classification. Both of these methodologies have previously proven useful in the context of OSA [32,36–39].

3.3.1. LSBoost for pediatric OSA diagnosis

We selected LSBoost as a regression algorithm to estimate the number of apneic events in each time segment. Ensemble-learning methods combine multiple weak base-learners decisions, leading to a robust algorithm with high generalization ability [40]. When using boosting algorithms, the estimations of the base-learners are sequentially computed. Accordingly, the next learner is trained based on the estimations of the previous ones [41]. In the present study, the target variable y_i was the number of apneic events of each HRV segment, the estimated output $f^m(x)$ was the estimated number of apneic events for each segment, and x_i the feature vector for each HRV segment composed of the ten extracted features detailed in the previous section. We decided to use decision stumps (trees with three nodes, one parent and two children) as base-learners as this approach conduct a feature selection stage *de facto* when training the models [36]. The formal definition of LSBoost can be specified this way [41,42]:

1. Being $f^m(x)$ the estimated output, set m to 0 and initialize the corresponding estimated output $f^0(x)$.
2. Increase m (the number of learners) by 1 and obtain the residuals as: $U_i = y_i - f^{m-1}(x_i)$ for i = 1,2 …,N, where N refers to the number of segments in the training set.
3. Fit the residual vector with least squares loss function, the weak learner h, and the predictors for each segment x_i: $(\lambda_m, a_m) = argmin_{\lambda,a} \sum_{i=1}^{N} [U_i^m - \lambda h(x_i; a)]^2$, where λ is a regularization parameter ranging from 0 to 1, and a the set of parameters of h.
4. Update $f^m(x) = f^{m-1}(x) + \lambda_m h(x; a_m)$.
5. Repeat iteratively steps 2 to 4 until $m = N_{LSB}$, being N_{LSB} the number of learners included.

Once the estimation for each HRV segment was obtained, the AHI for each subject can be computed as the rate between the estimated apneic events and the total recording time. However, using the total recording time leads to an underestimation of the actual AHI. Accordingly, we added a subsequent linear regression stage between the real and estimated AHI, which was modeled using the training set. Once that the model was fitted, it was applied to the AHI estimated in the validation and test sets, to correct for the underestimation trend imposed by the total recording time [31,43].

Two hyperparameters that needed to be optimized included: N_{LSB} and λ. We varied λ from 0.1 to 1 in steps of 0.1, and N_{LSB} was varied from 1 to 10,000 increasing the step each multiple of 10 (from 1 to 9 in steps of 1, from 10 to 100 in steps of 10, and so on). All the trained models for each (N_{LSB}, λ) pair with the training set were subsequently evaluated using the validation set, and the optimum (N_{LSB}, λ) pair was selected as the one that maximized multiclass Cohen's kappa (k).

Furthermore, as an effort to explain the role of each feature in the automated AHI estimation, we computed the relative importance of the included features. When using decision stumps, every $h(x; a_m)$ is function of a unique feature, so a procedure of feature selection is performed by default at each iteration [41,42]. Based on the mean squared error (MSE) of the empirical improvement across the trees, the relative importance can be estimated as [36,44]:

$$\widehat{I}_j^2 = \frac{1}{N_{LSB}} \sum_{m=1}^{N_{LSB}} MSE_m(x_j) w_m - \left(MSE_m^p(x_j) w_m^p + MSE_m^r(x_j) w_m^r\right), \quad (1)$$

where MSE_m is the mean squared error for the m regression tree linked to x_j, w_m the weight of the parent node probability, and $p-r$ the parameters associated with the children nodes. Once that $\widehat{I}_j^2$ of each feature is obtained, the predictors importance can be scaled as a percentage of contribution, with higher values meaning higher influence in the LSBoost model [45].

3.3.2. AdaBoost for sleep stage classification

The other approach to assess the clinical applicability of the HRV extracted features was to conduct per-segment sleep stage classification. To do this, we selected the boosting classification algorithm AdaBoost. Similar to LSBoost, AdaBoost combines several weak base classifier decisions sequentially, obtaining a more robust final classification decision, that is, a more generalized decision [40]. In this case, we selected linear discriminant analysis (LDA) as weak classifiers that have proven their applicability along with AdaBoost in previous studies addressing OSA [32,37].

At each m iteration of the AdaBoost algorithm, it assigns a weight w_i^m to each vector x_i in the training group. Then, the classifier of that iteration is trained with the corresponding weighted features, and the performance is evaluated computing an error, ε_m. This ε_m is used to determine the weighted vote of the classifier trained in that iteration, α_m [40], so that the smaller the ε_m, the higher the contribution of the classifier to the final decision. When the iteration ends, the weights of the misclassified x_i are updated (w_i^{m+1}) [40]. At this point, the weights of all features are normalized, maintaining their original distribution [46]. Through this reweighting, the LDA classifiers in the next iterations give more importance to those x_i misclassified in previous ones, thereby increasing the probability of being rightly classified [40,46].

As our aim was to classify each segment into one out of three sleep stages (W, NREMS and REMS), we used AdaBoost.M2, which is the AdaBoost algorithm version designed for multiclass classification [46]. When using AdaBoost.M2, ε_m is computed as follows [46]:

$$\varepsilon_m = \frac{1}{2} \sum_{k=1}^{N} \sum_{l \neq l_{true}} w_{i,l}^m (1 - c_m(x_i, l_{true}) + c_m(x_i, l)), \quad (2)$$

where l is a categorical variable representing the multiple classes, l_{true} is the real class labeled for x_i, and c_m is the confidence of LDA prediction for a vector x_i and a given class. To conduct the final classification task, the class with the highest sum of votes from all the LDA classifiers is returned, considering the weight of their predictions α_m as [46]:

$$\alpha_m = \ln(\beta_m) \quad (3)$$

with β_m defined as $\frac{(1-\varepsilon_m)}{\varepsilon_m}$ [46]. In order address the potential issue of overfitting, we also included a learning rate υ to redefine β_m in each iteration $(\beta_m)^{\upsilon}$. As in the case of LSBoost, we needed to fit two hyperparameters, the number of LDA classifiers (N_{AB}) and υ. The combinations of (N_{AB}, υ) values trained were the same than (N_{LSB}, λ), and we selected the combination of N_{AB} and υ that maximized the multiclass k in the validation set.

Finally, we also conducted a feature importance analysis in the sleep stage classification task. In this case, as we use a linear weak learner, we decided to use local interpretable model-agnostic explanation (LIME) technique with the AdaBoost model [47]. LIME is an explanation technique for machine-learning models that provides explanations of the predictions performed in an interpretable and faithful way [47]. The rationale behind LIME is to learn an interpretable model of the classifier locally around each prediction. Thus, for each instance of the test set, we computed LIME fitting a linear model locally, obtaining a weighted coefficient for each instance and feature, W_{ij}. Then, the global importance of each feature can be obtained as [47]:

$$I_j = \sqrt{\sum_{i=1}^{n} |W_{ij}|}. \quad (4)$$

After the computation of I_j, the global importance can be scaled as a percentage of the contribution, obtaining the relative importance for the AdaBoost Model as well.

3.4. Statistical analysis

The features computed in each segment in the training set did not fit either normality or homoscedasticity assumptions. Consequently, the non-parametric Mann-Whitney U test was used to assess two-by-two statistical differences between segment severity groups in each sleep stage, as well as between sleep stages in each segment severity group. Statistically significant differences between segment features were defined as those p-values <0.05 after Bonferroni correction (eighteen comparisons). However, when large sample sizes are available, the p-value by itself would not be sufficient for a comprehensive interpretation of the results [48]. Thus, we decided to complement the evaluation of differences between segment groups reporting the effect size of each comparison through the non-parametric Cohen's d measure [49]. Based on the Cohen's d, results can be interpreted as small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$) or large ($d \geq 0.8$) effect size [48,50]. We also visually represented the differences between segment groups with boxplots of the distribution for the different features considered.

Regarding the AHI estimations obtained from LSBoost, we evaluated its OSA diagnostic ability by splitting the subjects into four severity

levels (No OSA: AHI <1 e/h; Mild OSA: 1 ≤ AHI <5 e/h; Moderate OSA: 5 ≤ AHI <10 e/h; Severe OSA: AHI ≥10 e/h), and obtaining the confusion matrix, the four-class accuracy (Acc_4) and k. Besides, the diagnostic performance using the three common AHI cutoffs (1 e/h, 5 e/h and 10 e/h) was also evaluated using sensitivity (*Se*), specificity (*Sp*), accuracy (*Acc*), positive predictive value (*PPV*), negative predictive value (*NPV*), positive and negative likelihood ratios (LR^+, LR^-), the area under the receiver operating characteristics curve (*AUC*), and F1-score, which are widely used metrics to assess OSA diagnosis [12,31–33,36–39,43,51–53].

Regarding the AdaBoost model, the overall performance in the classification task into three sleep stages (W, NREMS and REMS) was reported by means of a confusion matrix, the three-class accuracy (Acc_3) and k. Additionally, for specific sleep stages, individual performance was evaluated in terms of *PPV* (also known as precision), *Se* (also known as recall), and F1-score (harmonic mean of precision and recall), as those metrics are commonly assessed to perform this task [54–56].

4. Results

4.1. Exploratory analysis in the training set

4.1.1. Differences in the PSDns between segments

Fig. 2 shows the averaged PSDns of the segments in the training set by considering three sleep stages and differentiating between no apneic/hypopneic segments (<1 e/s) and apneic/hypopneic segments (>1 e/s). The three classical spectral bands and two of the pediatric OSA-specific frequency bands have been shaded in the background (as mentioned above, the third OSA-specific band, BWRes, adapts to each subject). It can be observed that PSDns within the REMS stage behave more similar between control and apneic segments than in the case of NREMS segments. Segments among W, NREMS, and REMS were better differentiated in frequencies below the LF range (see Fig. 2B). Inside the BW2 band, NREMS apneic event segments presented higher power spectral values than the rest of groups. Finally, inside the HF band, NREMS segments presented higher power spectral values, as expected due to parasympathetic predominance, with NREMS no apneic segments showing higher activity than NREMS segments containing apneic segments.

4.1.2. Descriptive analysis of the features

Figs. 3–5 show the boxplots for the HRV segments differentiated by sleep stage and segment severity groups in the temporal measures, classic frequency bands, and pediatric OSA-related frequency bands, respectively. As commented in the methods section, due to the high number of the HRV segments involved in the study, rather than only using p-values, a better way to quantify the differences found among the segments is considering the effect size. Thus, Table 2 shows the Cohen's d measure to assess the effect size for the different comparisons conducted, with those effect sizes corresponding to statistically significant p-values highlighted in bold. The specific p-value for each comparison can be seen in the Supplementary Table 1.

Although most of the comparisons resulted in statistically significant differences (p-value <0.05), many of these comparisons showed negligible effect size (d < 0.2, see Table 2), which emphasizes the applicability of the evaluation of the differences through d values. Regarding intra-sleep stage comparisons, it can be observed in Table 2 that while NREMS showed a high number of comparisons with considerable effect size between segments with different number of apneic events (d = 0.5 or higher), these effects were attenuated in REMS. Particularly, NREMS RP_{BW2} displayed the highest effect size in five out of six comparisons between apneic severity segments. These marked differences in NREMS can also be observed as an increment of RP_{BW2} as the presence of apneic events increased (see Fig. 5B). Furthermore, the increased tendency in NREMS was also apparent in *mHR*, *SDNN*, RP_{LF}, and *LFn*, and to a lesser degree, i.e., lower d values. In contrast, RP_{HF} and RP_{BWRes} showed decreasing tendencies inside these frequency ranges as the frequency of apneic events increased, with RP_{BWRes} showing medium or large effects for all the comparisons considered. *RMSSD* was the only measure that showed negligible effect size within the two sleep stages. Inside REMS, *SDNN* and RP_{VLF} also showed negligible values in all the comparisons performed, and RP_{HF} and RP_{BW1} obtained negligible or small effect sizes

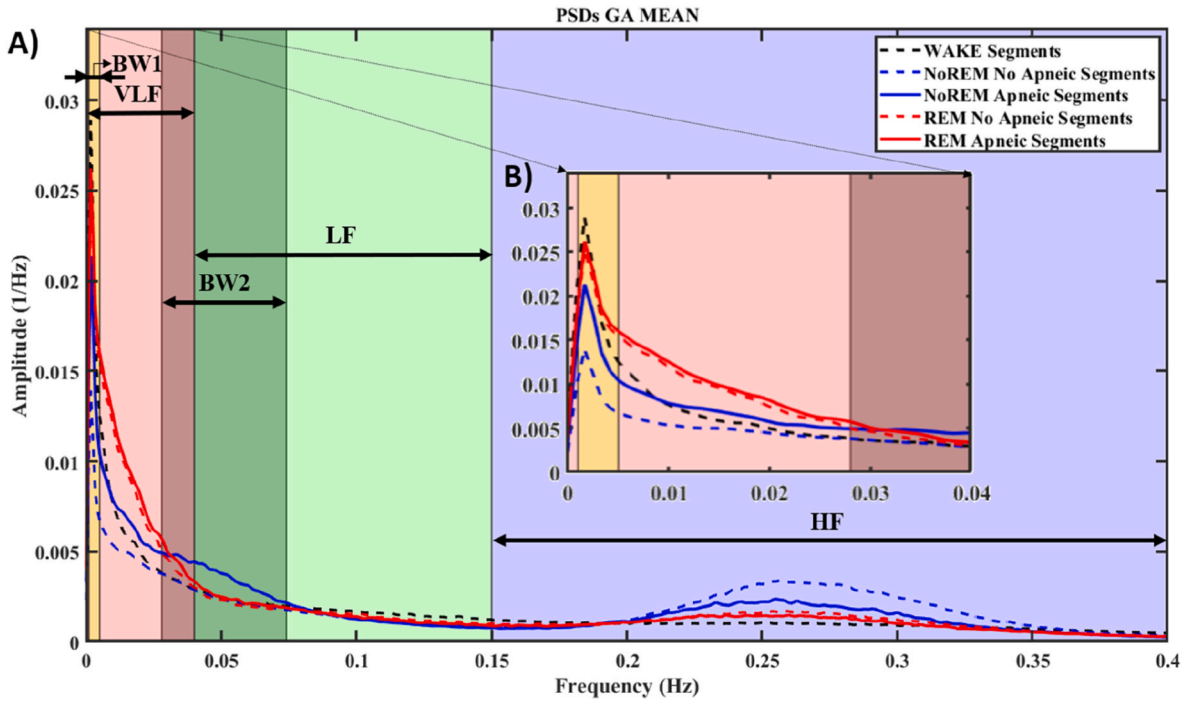

Fig. 2. A) Averaged PSDns in the 0–0.4 Hz range for five types of segments in the training set. Shaded areas represent those frequencies corresponding with the three classic spectral HRV bands (VLF: 0–0.04 Hz; LF: 0.04–0.15 Hz; HF: 0.15–0.4 Hz) and two of the pediatric OSA-specific frequency bands (BW1: 0.001–0.005 Hz; BW2: 0.028–0.074 Hz). B) Zoomed area in the VLF range.

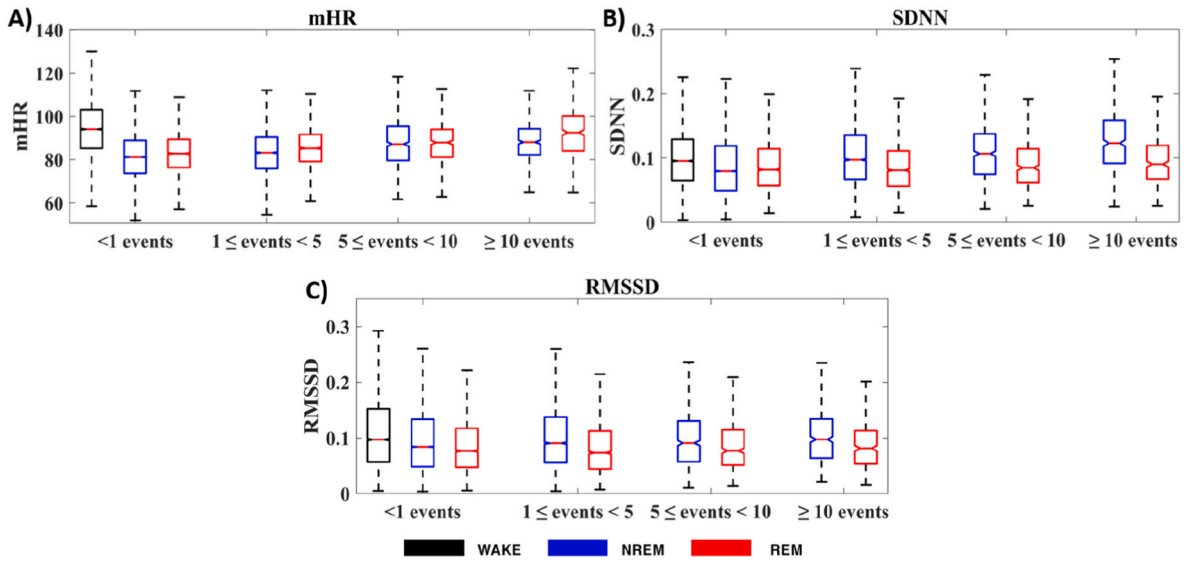

Fig. 3. Boxplot distribution of the temporal features computed for each type of segment included in the study in the training set. A) *mHR* boxplots; B) *SDNN* boxplots; C) *RMSSD* boxplots.

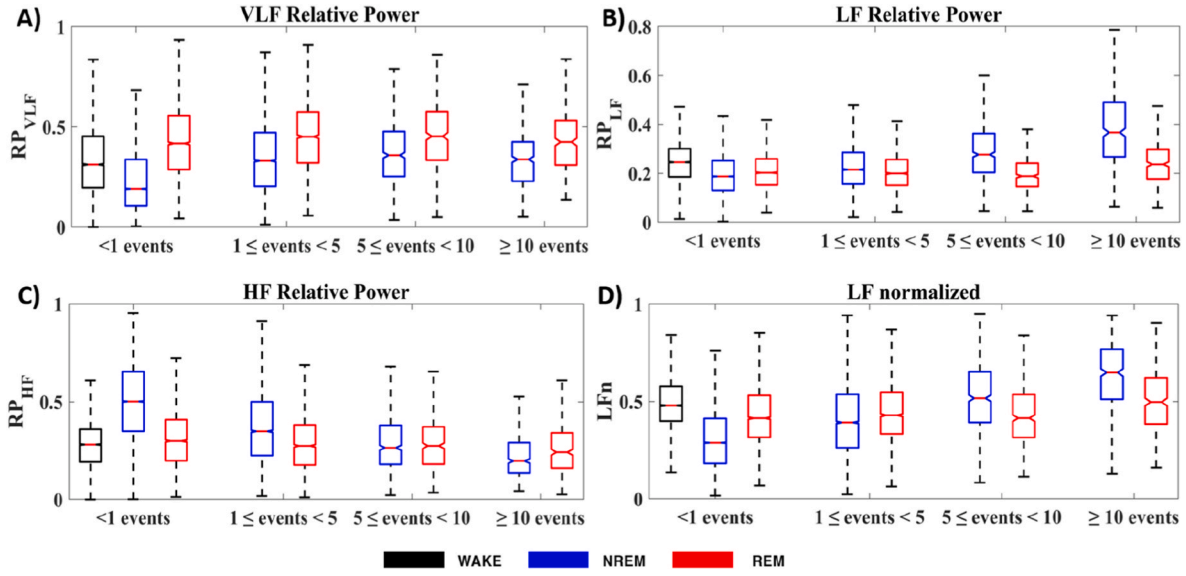

Fig. 4. Boxplot distribution of the frequency features in the classic HRV frequency bands computed for each type of segment included in the study in the training set. A) RP_{VLF} boxplots; B) RP_{LF} boxplots; C) RP_{HF} boxplots; D) *LFn*.

too. The reduction of the effect size inside REMS can also be observed in the boxplots (Figs. 3–5) where the parameters followed similar tendencies than in NREMS but are less pronounced. RP_{BW2} and *mHR* displayed the largest differences between apneic severity segments, with *mHR* consisting of the only parameter that allowed to differentiate between 1 ≤ apneic events <5 segments and 5 ≤ apneic events <10 segments (p-value <0.05), albeit with an associated small effect size (d = 0.263).

In the inter sleep stages comparisons, when assessing sleep stages in the absence of apneic events, RP_{BW2} was the only measure showing negligible effect size between sleep stages, thus corroborating the high dependence of this information on apneic events. RP_{BWRes} showed the highest effect size ($d > 0.8$) for W *vs.* NREMS, and NREMS *vs.* REMS segments. For the W *vs.* REMS comparison, only *mHR* and RP_{BWRes} showed at least a medium effect size, being higher in *mHR*. When including the presence of apneic events, RP_{BWRes} and RP_{VLF} allowed to differentiate ($d > 0.5$) between NREMS and REMS for 1 ≤ apneic events <5 segments, with higher effect size for RP_{BWRes}. For the 5 ≤ apneic events <10 segments and ≥10 apneic events segments, RP_{BW2} was again the parameter that allowed for a better differentiation between NREMS and REMS, showing large effect size in both cases.

4.1.3. LSBoost and AdaBoost models training and validation

Following extraction of the ten features for each segment, we optimized both the LSBoost and AdaBoost models to evaluate the clinical applicability of the previous HRV characterization. As previously mentioned, this optimization was based on the pair of hyperparameters that maximized Cohen's *k* in the validation set. The combinations of (N_{AB}, υ) values trained were the same than (N_{LSB}, λ), varying N_{AB} and N_{LSB} from 1 to 10,000, increasing the step each multiple of 10, and υ and λ from 0.1 to 1 in steps of 0.1. Supplemental Fig. 1A shows the evolution of the Cohen's *k* in the validation set for each (N_{LSB}, λ) pair. The optimum values selected for the LSBoost model were $\lambda = 0.3$ and $N_{LSB} = 300$. The optimization process was similar for AdaBoost. The evolution of the

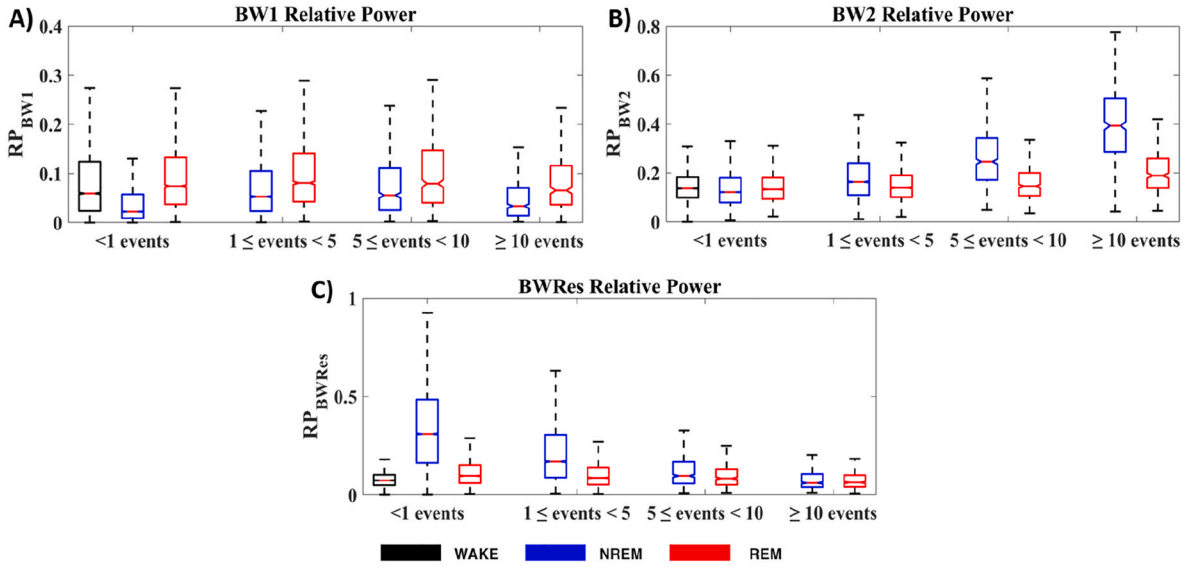

Fig. 5. Boxplot distribution of the frequency features in the pediatric OSA-specific HRV frequency bands computed for each type of segment included in the study in the training set. A) RP_{BW1} boxplots; B) RP_{BW2} boxplots; C) RP_{BWRes} boxplots.

Table 2
Cohen's *d* measures obtained to assess the effect size associated to the comparisons performed between all the types of segments considered in the study in the training set.

Intra-Stages	NREMS Segments					
Feature	<1 vs 1≤events<5	<1 vs 5≤events<10	<1 vs ≥ 10 events	1≤events<5 vs 5≤events<10	1≤events<5 vs ≥ 10 events	5≤events<10 vs ≥ 10 events
mHR	**0.147**	**0.539****	**0.634****	**0.412***	**0.516****	0.100
SDNN	**0.318***	**0.456***	**0.745****	**0.140**	**0.433***	**0.312***
RMSSD	**0.076**	0.065	**0.123**	0.011	0.050	0.068
RP_{VLF}	**0.610***	**0.754***	**0.579***	**0.137**	0.035	**0.201**
RP_{LF}	**0.325**	**1.007*****	**2.038*****	**0.628****	**1.553*****	**0.733****
RP_{HF}	**0.633****	**1.042*****	**1.335*****	**0.452***	**0.773****	**0.413***
LFn	**0.578****	**1.233*****	**1.912*****	**0.584****	**1.196*****	**0.625****
RP_{BW1}	**0.440***	**0.483***	**0.072**	0.028	**0.337**	**0.391**
RP_{BW2}	**0.546****	**1.531*****	**3.106*****	**0.779****	**2.017*****	**0.963*****
RP_{BWRes}	**0.608****	**1.023*****	**1.272*****	**0.541****	**0.859*****	**0.565****
Intra-Stages	REMS Segments					
Feature	<1 vs 1≤events<5	<1 vs 5≤events<10	<1 vs ≥ 10 events	1≤events<5 vs 5≤events<10	1≤events<5 vs ≥ 10 events	5≤events<10 vs ≥ 10 events
mHR	**0.242***	**0.500****	**0.904*****	**0.263***	**0.671****	**0.398***
SDNN	0.036	0.043	**0.148**	0.082	**0.190**	0.115
RMSSD	0.070	0.000	0.028	0.071	0.101	0.031
RP_{VLF}	**0.142**	**0.194**	0.064	0.054	0.080	0.142
RP_{LF}	0.046	**0.177**	**0.423***	0.134	**0.473***	**0.587****
RP_{HF}	**0.156**	**0.202***	**0.388***	0.049	**0.243***	0.213*
LFn	**0.097**	0.043	**0.503****	0.054	**0.392***	**0.459***
RP_{BW1}	**0.067**	0.084	0.134	0.017	**0.203***	0.224*
RP_{BW2}	**0.105**	**0.220***	**0.875*****	0.113	**0.751****	**0.573****
RP_{BWRes}	**0.152**	**0.243**	**0.511****	0.103	**0.401**	**0.379**

Inter-Stages	<1 event Segments			1≤events<5 Segments	5≤events<10 Segments	≥10 events Segments
Feature	W vs NREMS	W vs REMS	NREMS vs REMS	NREMS vs REMS	NREMS vs REMS	NREMS vs REMS
mHR	**0.958*****	**0.755****	**0.115**	**0.190**	0.007	**0.298**
SDNN	**0.223***	**0.204***	**0.023**	**0.343***	**0.452***	**0.709****
RMSSD	**0.126**	**0.279***	**0.144**	**0.299***	**0.243***	**0.300**
RP_{VLF}	**0.547****	**0.441***	**1.074*****	**0.597****	**0.569****	**0.647****
RP_{LF}	**0.417***	**0.261***	**0.158**	**0.210***	**0.927*****	**1.096*****
RP_{HF}	**1.233*****	**0.286**	**0.954*****	**0.480***	0.060	**0.176**
LFn	**1.168*****	**0.453***	**0.728****	**0.209***	**0.465***	**0.700****
RP_{BW1}	**0.589****	**0.008**	**0.739****	**0.334***	**0.313***	**0.539****
RP_{BW2}	**0.062**	0.021	**0.080**	**0.351***	**1.005*****	**1.420*****
RP_{BWRes}	**1.460*****	**0.514****	**1.133*****	**0.764****	**0.361***	0.072

NREMS: Non rapid eye movement sleep; REMS: Rapid Eye Movement Sleep; mHR: Mean Heart Rate; SDNN: Standard deviation of normal-to-normal interval time series; RMSSD: Root Mean Square of successive differences of NN intervals; RP: Relative Power; VLF: Very Low Frequency; LF: Low Frequency; HF: High Frequency; W: Wake.
Statistically significant comparisons (p-value <0.05) have been highlighted in bold.
*Small effect ($0.2 \leq d < 0.5$); ** medium effect ($0.5 \leq d < 0.8$); *** large effect ($d \geq 0.8$).

Cohen's k in the validation set for those values can be observed in Supplemental Fig. 1B. The optimum (N_{AB}, υ) pair were $\upsilon = 0.1$ and $N_{AB} = 3000$.

4.2. Evaluation of the clinical applicability in the test set

4.2.1. Clinical applicability for pediatric OSA diagnosis

Fig. 6 shows the confusion matrix along with the corresponding Acc_4 and k obtained after using the LSBoost-estimated AHI to determine the OSA severity degree of the subjects in the test set. The color code of the confusion matrix allows to observe how the model performs. The darker the colors of the main diagonal, the better the diagnosis performance. It becomes apparent an overestimation of the AHI for the No OSA subjects, which achieved the lowest proportion of subjects rightly classified among the four classes. Similarly, the highest proportion was obtained for the moderate OSA subjects.

The diagnosis obtained to predict OSA presence using the three reference AHI cutoffs is shown in Table 3. The increase of the *AUC* with the increasing cutoffs reflects an increment in the overall diagnosis performance with OSA severity cutoffs. Fig. 7A shows the relative importance of the features included in the LSBoost model. It can be observed that, by far, RP_{BW2} was the feature that accounted the most for the relative importance (72.01%), followed by *SDNN* (7.09%), and RP_{LF} (6.08%). The four first features (the previous ones plus RP_{BWRes}) contributed over 90% to the final AHI estimation.

4.2.2. Clinical applicability for sleep stage classification

For the multiclass sleep stage classification task, the confusion matrix, Acc_3 and k achieved by AdaBoost in the test set are shown in Fig. 8. In this setting, the greatest proportion of segments correctly classified was obtained for the REMS segments. However, when checking the individual sleep stage classification results, which are shown in Table 4, despite the high recall obtained in the REMS stage (77.13%), this stage also achieved the lowest precision (39.17%). Thus, the highest overall classification performance was achieved in NREMS with the greatest precision (94.55%) and F1-score (0.818) among the three sleep stages included herein.

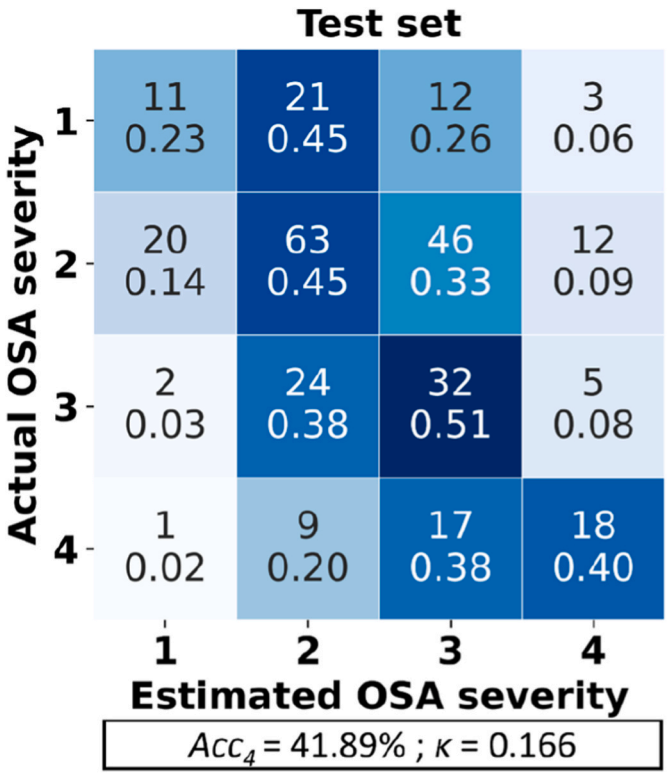

Fig. 6. Confusion matrix for pediatric OSA diagnosis using the LSBoost model in the test set. The main diagonal represents the number and proportion of subjects rightly classified. The darkness of the cells represents the proportion of the actual class assigned to each group. 1: No OSA (AHI <1 e/h); 2: Mild OSA (1 ≤ AHI <5 e/h); 3: Moderate OSA (5 ≤ AHI <10 e/h); 4: Severe OSA (AHI ≥10 e/h).

Finally, Fig. 7B shows the relative predictor importance computed using LIME for the features included in the AdaBoost model in the classification task. In the case of sleep stages, RP_{BWRes} was the feature with the highest importance among the ten included (20.04%), followed by RP_{HF} (17.52%) and RP_{VLF} (12.54%). In general, the relative importance of the ten features is higher for sleep stage classification than for the apneic event estimation, which is clearly dominated by BW2.

5. Discussion

In this study, a segment-based characterization of HRV that accounts for both sleep stages and the presence of apneic events was performed for the first time in the context of pediatric OSA. The marked differences in the HRV parameters across NREMS according to severity of respiratory disturbance were attenuated in REMS. In both cases, BW2, a pediatric OSA-specific frequency band, achieved the highest differentiation ability. However, in the absence of apneic events, another novel pediatric OSA-related frequency band, BWRes, achieved a better differentiation ability across sleep stages. The usefulness of BW2 and BWRes in HRV analysis was also confirmed when analyzing its clinical applicability to diagnose pediatric OSA and to classify sleep stages, respectively.

5.1. Physiological interpretation of the characterization of the segments

Our approach showed greater differences for NREMS compared to REMS between segments with variable numbers of apneic events. This was appreciable in the averaged PSDns of the segments (see Fig. 2), the feature boxplots (see Figs. 3–5) and in a general reduction of the effect size in the REMS features. As mentioned above, NREMS has been associated with downregulation of SNS activity along with enhanced PNS activity, whereas this trend is reversed during REMS and wakefulness [2]. Likewise, previous pediatric OSA studies have found overall increased sympathetic activity in OSA children [2,15], which has been reduced with treatment [57]. This increased sympathetic activity has been confirmed in our study and is significantly associated with an increased number of apneic events. In parallel, we identified that the parameters that measure SNS activity to some extent (RP_{BW2}, *mHR*, *SDNN*, and RP_{LF}) also show increased correlation within NREMS (see Figs. 3–5). However, these trends were reduced in REMS, with RP_{BW2} and *mHR* emerging as the only parameters reaching considerable effect size (see Table 2) in three out of the six comparisons evaluated. Accordingly, despite the fact that apneic events occur more often during REMS in the context of pediatric OSA [2], it seems that REMS basal SNS activation may be masking the effect imposed by the respiratory perturbations. In this regard, RP_{BW2} reached considerable effect sizes in all the comparisons conducted for NREMS and achieved the highest feature importance for the diagnosis of pediatric OSA. Therefore, we highlight this frequency band as a measure of specific SNS activation during apneic events, which is further accentuated when these events occur during NREMS.

According to the effect size analyses, the utility of RP_{BW2} to characterize HRV alterations in the context of pediatric OSA seems to disappear in the absence of respiratory disturbances. However, another recently defined OSA-specific band (RP_{BWRes}) was the only measure that performed well according to effect size in all the comparisons among sleep stages for no apneic segments (see Table 2). One possible explanation for property is that W, NREMS, and REMS can be readily differentiated by their effects on respiratory patterning [58–61]. During wakefulness, respiratory rate is regulated by both voluntary and

Table 3
Diagnostic performance by the LSBoost Model in the test set for binary classification in the three apnea-hypopnea index cutoffs 1, 5 and 10 events/hour (e/h).

LSBoost Model: $\lambda = 0.3$; $N_{LSB} = 300$										
	Cutoff	Se	Sp	Acc	PPV	NPV	LR^+	LR^-	AUC	F1-score
Test Set	**1 e/h**	90.76	23.40	80.07	86.26	32.35	1.18	0.39	0.651	0.885
	5 e/h	66.67	61.17	63.18	49.66	76.16	1.72	0.54	0.677	0.569
	10 e/h	40.00	92.03	84.12	47.37	89.53	5.02	0.65	0.742	0.434

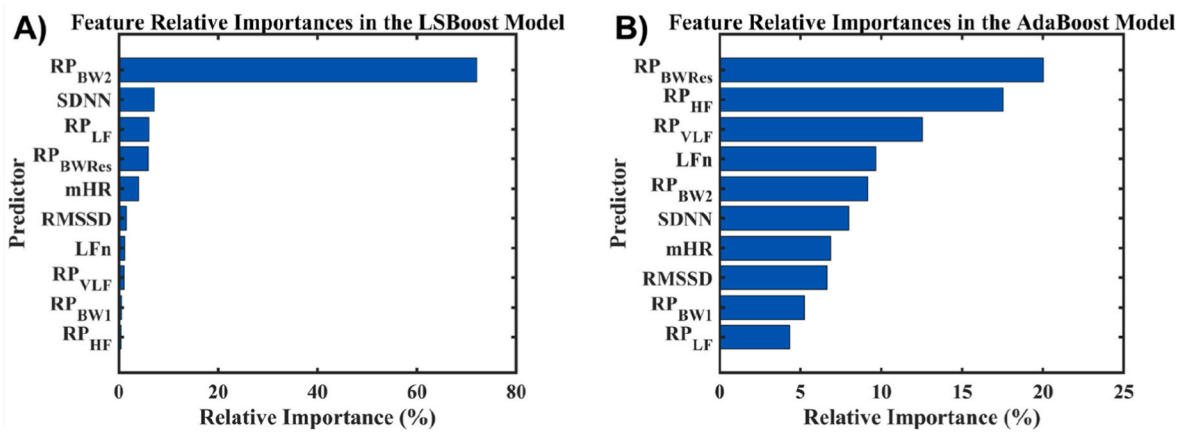

Fig. 7. Feature importance for both the LSBoost and AdaBoost models. A) Relative importance of the ten features through the tree base classifiers in the LSBoost Model. B) Relative importance of the ten features included in the AdaBoost Model, computed using LIME.

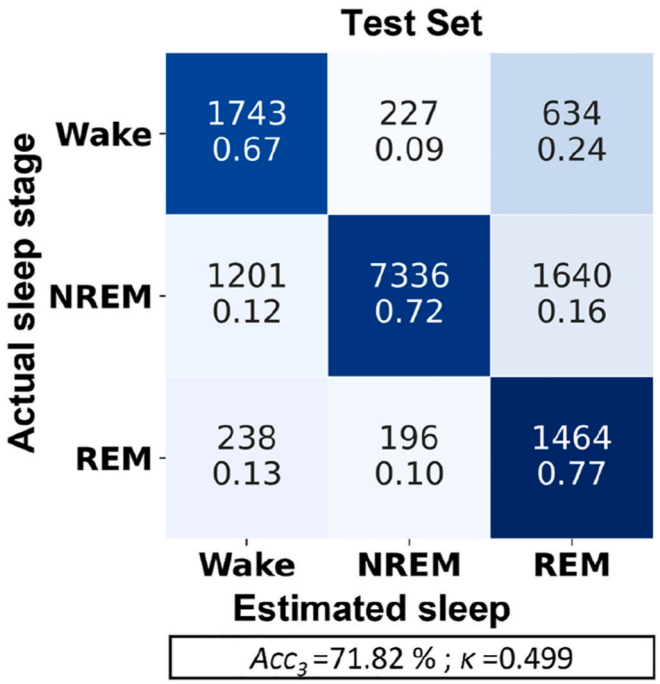

Fig. 8. Confusion matrix of the classification in sleep stages by the AdaBoost model in the test set. The main diagonal represents the number and proportion of subjects rightly classified. The darkness of the cells represents the proportion of the actual class assigned to each group.

automatic controllers, but the voluntary regulation disappears while asleep [62]. During NREMS, breathing is characterized by regular patterns with decreased variability. However, during REMS, irregular breathing and short respiratory pauses appear in healthy children, with increments in the respiratory rate [62]. The HF band reflects HR and blood pressure oscillations induced by such changes respiratory activity,

Table 4
AdaBoost model classification performance in the test set for individual sleep stage classification in the three sleep stages considered: W, NREMS and REMS.

AdaBoost Model: $\upsilon = 0.1$; $N_{AB} = 3000$				
	Sleep Stage	Precision (%)	Recall (%)	F1-score
Test Set	**W**	54.78	66.94	0.603
	NREMS	94.55	72.08	0.818
	REMS	39.17	77.13	0.516

and is a marker of PNS activity [62]. Nonetheless, the inter-individual variability of the respiratory frequencies due to other aspects such as age makes the adaptive RP_{BWRes} better suited to assess respiratory influence on HRV [22], as highlighted in the relative importance analysis conducted in the AdaBoost model. Taken together, the importance of respiration to discriminate sleep stages is robustly represented in our study as RP_{BWRes} and RP_{HF} achieved the highest relative importance in the classification of sleep stages, with *LFn* (also influenced by HF) achieving a high relative importance as well.

Non-apneic NREMS segments reflected an enhanced PNS activity compared to W and REMS (see RP_{HF} and RP_{BWRes} in Figs. 4–5). However, when such segments included apneic events, the PNS activity during NREMS measured by the HF band and BWRes decreased, reaching REMS-like levels in the 5 $\leq$ apneic events <10 e/s and $\geq$10 e/s segments. This reduction resulted in effect sizes under the 0.5 threshold in NREMS *vs.* REMS comparisons for both RP_{HF} and RP_{BWRes}. Remarkably, as the presence of apneic events increased, RP_{BW2} also increased, achieving the highest differentiation ($d > 0.8$) between NREMS and REMS for the 5 $\leq$ apneic events <10 e/s and $\geq$10 e/s segments. These results highlight the importance of BW2 beyond detecting segments with respiratory disturbances and, specifically, to differentiate NREMS from REMS segments even when apneic events are present. Accordingly, we have shown that both BWRes and BW2 play important roles to differentiate sleep stages in the context of pediatric OSA.

5.2. Clinical applicability of the characterization of the nocturnal HRV segments

In the present study, the clinical applicability of the information extracted through the characterization of HRV has been highlighted. In the case of the diagnosis of pediatric OSA, we observed an increment of performance with disease severity, achieving the highest diagnostic yield in the 10 e/h cutoff with an $Acc = 84.12\%$ and $AUC = 0.742$, but with an unbalanced *Se/Sp* pair (40.00/92.03%). This increase in the performance with the severity threshold may be due to the well-tuned characterization that RP_{BW2} showed in the presence of apneic events. However, considering the modest overall diagnostic performance, the most useful approach for an automatic strategy using our LSBoost model would be detecting OSA presence (90.76% *Se* and 86.26% *PPV* for an AHI cutoff of 1 e/h) and discarding severe OSA (92.03% *Sp* and 89.53% *NPV* for an AHI cutoff of 10 e/h). Relative to the sleep stage classification task, the performance metrics were remarkably higher than in the diagnostic approach, as higher values of both multiclass k (0.499 *vs.* 0.166) and accuracy ($Acc_3 = 71.82\%$ *vs.* $Acc_4 = 41.89\%$) were reached. NREMS was the sleep stage with the best overall classification performance (72.08% recall and 94.55% precision), which is coherent with the largest differences found after the characterization step.

According to these findings, a real clinical implementation using a single-signal approach based on HRV in the context of pediatric OSA diagnosis could benefit from an initial step such as to detect NREMS HRV followed by a second step to estimate AHI from the recorded segments, since these are the segments with the highest discriminant ability. In summary, we have shown that the characterization of HRV segments, is a promising approach to classify sleep stages, as well as to detect the presence of OSA in children, and to discard severe OSA, rather than to automatically perform multiclass classification of pediatric OSA. Further studies in large prospective cohorts should further allow for confirmation of these assumptions.

5.3. Comparison with previous work

To the best of our knowledge, this is the first study in which a HRV segment-based evaluation considering both sleep stages and apneic events presence was conducted in a pediatric population. However, some similarities with previous studies evaluating changes among classical HRV parameters across sleep stages should be mentioned. Kontos et al. [6] reported segment-based HRV during sleep of healthy children and adolescents, and their results can be compared with our findings when excluding the presence of apneic events. These investigators observed the same tendencies in HRV spectral power within LF and HF bands. Also, they measured mean NN, which is the inverse of our *mHR*, obtaining opposing tendencies, i.e., the same outcomes. Despite these similarities, dissimilar results were obtained from *SDNN* and *RMSSD* measures. These discrepancies could be due to the fact that they only considered the first W pre-sleep segment, as well as the different segment size they used in each sleep stage (3 min for W and REMS, and 6 min for NREMS). Also, they evaluated 7 segments for each child, with a total of 75 children included in the study. This means that the conclusions by Kontos and collaborators were based on 525 segments, against the 40,105 segments without apneic events analyzed by us in the training set.

When HRV segments were compared between healthy and OSA children, the studies excluded segments with apneic events. Two of these studies agreed that for all the severity groups (healthy, primary snorers, mild OSA, and moderate-severe OSA) there was a decrease in the HF band power from NREMS to REMS [18,21]. This decrease was also observed in our study in RP_{HF} for the no apneic segments. There is also another study that excluded apneic episodes from its analysis and compared HRV parameters in different sleep stages across severity groups [19]. Following this approach, they reported an absence of differences in LF power between groups, and only reported differences between moderate-severe OSA and controls in the power of the HF band and in the LF/HF ratio across all sleep stages. Notwithstanding the differences found, the authors concluded that pediatric OSA did not manifest autonomic dysfunction [19]. However, in the present study we have shown that the presence of apneic events modulates HRV across several HRV measures. Thus, in the aforementioned study [19], exclusion of apneic events could have hidden the evidence pointing to autonomic dysfunction due to OSA, highlighting the importance of considering the presence of apneic events to detect HRV differences across the night.

Other published studies included apneic events, but their analyses included segments with and without apneic events clumped together. Nonetheless, some affinity between these studies and the present study are worthy of mention. Baharav et al. [15] compared the overnight HRV in OSA and healthy children, and observed higher LF activity and LF/HF ratio in pediatric OSA in all the sleep stages considered. In the present study, we observed increasing tendencies in RP_{LF} and *LFn* as the presence of apneic events increased, particularly during NREMS. Similarly, Horne et al. [16] also compared HRV throughout the night but differentiating additional children groups: control non snorers, normal weight primary snorers, normal weight OSA, overweight primary snorers, and overweight OSA children. They showed that overweight OSA presented elevated overnight HR compared to healthy children, and lower HF activity than normal weight primary snorers in NREMS. Also, LF activity was higher for normal weight OSA than overweight primary snorers. Again, as the presence of apneic events is minimal in healthy children and primary snorers, the trends observed in the present work as the density of apneic events increased in those parameters are in accordance with the results reported by Horne and colleagues [16]. Recently, Wu et al. [17] also reported an evolution of HRV parameters across sleep stages comparing healthy children to mild and moderate-to-severe OSA. In their results, the mean RR (that is, the inverse of *mHR*) decreased with OSA severity in all sleep stages, while LF/HF ratio increased with OSA severity, thus agreeing with the results of the above-mentioned studies [15,16], and with the present study.

Regarding clinical usefulness assessment, Shouldice et al. [12] reported per-subject classification results, obtaining $Acc = 84\%$, $Se = 85.7\%$, $PPV = 85.7\%$, $Sp = 81.8\%$, and $NPV = 81.8\%$ when using a 12.5 e/h AHI cutoff. In addition, three studies from another research group [51–53] also conducted per-subject classification in pediatric OSA using HRV features derived from decreases in the fluctuation of the photoplethysmography signal. They reported *Acc* in the range 73.3–80%, *Se* in the range 62.5–87.5%, *PPV* in the range 75.0–85.71%, *Sp* in the range 71.45–85.7%, and *NPV* in the range 66.7–83.3% when classifying OSA (>18 e/h) *vs.* healthy subjects (<5 e/h). Although lower *Se* was identified in the present study, Table 3 shows that our approach is more useful to discard children without severe OSA, as reflected by higher *Sp* and *NPV* with a stricter severity threshold (10 e/h). It would allow to reduce the subjects sent to a normal evaluation, reducing waiting lists. Of course, population differences among the studies, as well as the criteria used to establish OSA hinder any further comparisons. Therefore, our previous studies examining HRV in the context of pediatric OSA [22,28] will need to serve as best comparators. Although these studies were not segment-based, the diagnostic performance in the same AHI cutoffs used here were reported. Overall diagnostic performance in the 5 and 10 e/h cutoffs was lower here than in these previous studies. However, the highest diagnostic performance was achieved in the present study for a cut-off of 1 e/h in terms of *Acc*, *AUC*, *Se* and *PPV*. These findings highlight the potentially advantageous clinical applicability of the characterization of HRV segments to detect the presence of OSA even in its mildest forms.

Sleep stage classification using only cardiac measures has been scarcely investigated in children. To the best of our knowledge, this is the first study in which HRV metrics were used to evaluate classification of sleep stages in the context of pediatric OSA. However, sleep stage classification in healthy infants has been previously investigated.

Haddad et al. [63] conducted the assessment of sleep stages in 9 infants (1–4 months of age) using cardiorespiratory measures to distinguish REMS and quiet sleep. After a preliminary evaluation, they observed that the variation of the respiratory cycle time had the highest chances to differentiate sleep stages, reporting *Se* of 93% in quiet sleep, and ~99% in REMS. However, the classification performance of cardiac measures alone was not reported [63]. Harper et al. [64] also conducted sleep stage classification through cardiorespiratory measures in 25 infants (up to 6 months of life). They differentiated 1-min segments of W, quiet sleep, and REMS, reporting and Acc_3 = 84.8% when using 7 cardiorespiratory measures. This overall accuracy was reduced to 82% using cardiac measures alone [64]. Finally, Lewicke et al. [65] conducted sleep *vs*. W classification in 30-s segments of 190 infants using several machine learning models with HRV measures. They reported accuracies of ~78%, which increased to 85–87% when rejecting 30% of segments difficult to classify [65]. Unfortunately, the nature and size of the sample of the previous studies, as well as the kind of sleep stages considered, make it virtually impossible to compare to our present findings.

5.4. Limitations and future work

Several limitations deserve mention. First, although the main aim of this study was to characterize HRV using sleep-specific segments, the results obtained from the LSBoost and AdaBoost models are not sufficiently robust for widespread implementation. Thus, future machine-learning focused efforts will be needed to determine their usefulness in the tasks of achieving reliable determination of a diagnosis of pediatric patients and in sleep stage classification. Furthermore, there is imbalance between the number of different type of segments considered. Although the number of segments included for each class is large (see Table 1), there were more NREMS and <1 apneic event segments. To test if balancing the dataset would change the performance, we performed a reanalysis of both LSBoost and AdaBoost models with a balanced subset of the segments, but these results did not lead to an improvement in any of the performances for both tasks. In the future, inclusion of more segments with apneic events, as well as more segments corresponding to W and REMS would be desirable to increase the generalizability of our results. Besides, the original annotations of sleep stages were performed on 30 s epochs, but some of the features included cannot be computed for this length of register, so we performed sleep stage annotations based on 10-min segments. Thus, the selection of metrics that cannot be computed for 30 s of HRV signal, despite being a common practice in the literature, constitutes another limitation of our study. In addition, we have only considered W, NREMS, or REMS sleep stages. Although this is an approach followed in many studies aimed at sleep stage classification and characterization [6,18,63,64], the AASM establishes that NREMS sleep stage can be divided into N1, N2 and N3 [66]. Therefore, further HRV segment-based analyses may benefit from inclusion of NREMS sub-stages. Of note, evaluating the clinical applicability of current findings in prediction of OSA-associated cardiovascular risks may provide further value if applicable.

6. Conclusions

This is the first study in which a segment-based evaluation of HRV incorporating sleep stages and apneic events has been conducted in children. In addition, the evolution of HRV measures in pediatric OSA-specific frequency bands across sleep stages is introduced. Besides, the reliability of models trained with HRV OSA-specific frequency measures through ensemble-learning algorithms in the context of pediatric OSA has also been assessed for the first time. This approach allowed us to observe that two of these novel spectral bands, BW2 and BWRes, displayed increased relevance when compared to conventional spectral bands when establishing pediatric OSA severity and to classify sleep stages, respectively. Although an increased effect of sympathetic activation would be expected during REMS in the presence of respiratory disturbances, the characteristically increased basal sympathetic activity of REMS appears to mask the sympathetic excitation induced by apneic or hypopneic events. Such phenomenon is therefore easier to distinguish during NREMS. Accordingly, when evaluating HRV in pediatric OSA, both the sleep stage and the presence of apneic events need to be considered. Furthermore, the analysis of the pediatric-OSA specific spectral bands may prove particularly useful in both the automated diagnosis of OSA and in machine-based sleep stage classification in children.

Ethical approval

This work has been carried out according to the Declaration of Helsinki. The clinical trial identifier of the original CHAT database is NCT00560859. In all patients, a written consent for parental permission for the research, along with assent for those children over 7 years of age, was obtained as part of the research protocol, which can be found in the supplementary material of Marcus et al. [25].

Authorship contribution statement

Study design: A. Martín-Montero, G.C. Gutiérrez-Tobal, P. Laguna, and R. Hornero; Implementation: A. Martín-Montero and P. Armañac-Julián; Data Analysis: A. Martín-Montero, P. Armañac-Julián, E. Gil, D. Álvarez, J. Lázaro, R. Bailón, G.C. Gutiérrez-Tobal; Manuscript Writing: A. Martín-Montero, P. Armañac-Julián, E. Gil, L. Kheirandish-Gozal, D. Álvarez, J. Lázaro, R. Bailón, D. Gozal, P. Laguna, R. Hornero, and G.C. Gutiérrez-Tobal; Manuscript review: A. Martín-Montero, P. Armañac-Julián, E. Gil, L. Kheirandish-Gozal, D. Álvarez, J. Lázaro, R. Bailón, D. Gozal, P. Laguna, R. Hornero, and G.C. Gutiérrez-Tobal; Funding acquisition: L. Kheirandish-Gozal, D. Gozal, P. Laguna, R. Hornero, and G.C. Gutiérrez-Tobal. All authors gave their final approval of this version of the manuscript.

Conflict of interest

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

Acknowledgements

This work was supported by 'Ministerio de Ciencia, Innovación y Universidades' and 'European Regional Development Fund (FEDER)' under projects PID2019-104881RB-I00, PID2020-115468RB-I00, PDC2021-120775-I00 and PID2021-126734OB-C21, by 'Sociedad Española de Sueño (SES)' under project "Beca de Investigación SES 2019", by 'CIBER -Consorcio Centro de Investigación Biomédica en Red-(CB19/01/00012)' through 'Instituto de Salud Carlos III' co-funded with FEDER funds, as well as under the project SleepyHeart from 2020 valorization call, and by Gobierno de Aragón (Reference Group BSICoS T39-20R). The Childhood Adenotonsillectomy Trial (CHAT) was supported by the National Institutes of Health (HL083075, HL083129, UL1-RR-024134, UL1 RR024989). The National Sleep Research Resource was supported by the National Heart, Lung, and Blood Institute (R24 HL114473, 75N92019R002). A. Martín-Montero 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. Álvarez is supported by a "Ramón y Cajal" grant (RYC2019-028566-I) from the 'Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación' co-funded by the European Social Fund. Leila Kheirandish-Gozal and David Gozal are supported by the Leda J Sears Foundation for Pediatric Research and DG is also supported by National Institutes of Health (NIH) grant AG061824, and a Tier 2 grant from the University of Missouri.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106549.

References

[1] M.H. Bonnet, D.L. Arand, Heart rate variability: sleep stage, time of night, and arousal influences, Electroencephalogr. Clin. Neurophysiol. 102 (1997) 390–396, https://doi.org/10.1016/S0921-884X(96)96070-1.

[2] H. Qin, N. Steenbergen, M. Glos, N. Wessel, J.F. Kraemer, F. Vaquerizo-Villar, T. Penzel, The different facets of heart rate variability in obstructive sleep apnea, Front. Psychiatr. 12 (2021) 1–20, https://doi.org/10.3389/fpsyt.2021.642333.

[3] Task Force of the ESC the NAS of Pacing Electrophysiology, Heart rate variability. Standards of measurement, physiological interpretation, and clinical use, Circulation 93 (1996) 1043–1065, https://doi.org/10.1093/oxfordjournals.eurheartj.a014868.

[4] J. Milagro, M. Deviaene, E. Gil, J. Lázaro, B. Buyse, D. Testelmans, P. Borzée, R. Willems, S. Van Huffel, R. Bailón, C. Varon, Autonomic dysfunction increases cardiovascular risk in the presence of sleep apnea, Front. Physiol. 10 (2019) 1–11, https://doi.org/10.3389/fphys.2019.00620.

[5] A. Baharav, S. Kotagal, V. Gibbons, B.K. Rubin, G. Pratt, J. Karin, S. Akselrod, Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability, Neurology 45 (1995) 1183–1187, https://doi.org/10.1212/WNL.45.6.1183.

[6] A. Kontos, M. Baumert, K. Lushington, D. Kennedy, M. Kohler, D. Cicua-Navarro, Y. Pamula, J. Martin, The inconsistent nature of heart rate variability during sleep in normal children and adolescents, Front. Cardiovasc. Med. 7 (2020) 1–11, https://doi.org/10.3389/fcvm.2020.00019.

[7] G. Aljadeff, D. Gozal, V.L. Schechtman, B. Burrell, R.M. Harper, S.L. Davidson Ward, Heart rate variability in children with obstructive sleep apnea, Sleep 20 (1997) 151–157, https://doi.org/10.1093/sleep/20.2.151.

[8] D. Gozal, F. Hakim, L. Kheirandish-Gozal, Chemoreceptors, baroreceptors and autonomic deregulation in children with obstructive sleep apnea, Respir. Physiol. Neurobiol. 185 (2013) 177–185, https://doi.org/10.4213/mzm87.

[9] C. Guilleminault, R. Winkle, S. Connolly, K. Melvin, A. Tilkian, Cyclical variation of the heart rate in sleep apnoea syndrome, Lancet 323 (1984) 126–131, https://doi.org/10.1016/s0140-6736(84)90062-x.

[10] D.M. O'Driscoll, A.M. Foster, M.L. Ng, J.S.C. Yang, F. Bashir, S. Wong, G.M. Nixon, M.J. Davey, V. Anderson, A.M. Walker, J. Trinder, R.S.C. Horne, Central apnoeas have significant effects on blood pressure and heart rate in children, J. Sleep Res. 18 (2009) 415–421, https://doi.org/10.1111/j.1365-2869.2009.00766.x.

[11] O. Vitelli, M. Del Pozzo, G. Baccari, J. Rabasco, N. Pietropaoli, M. Barreto, M. P. Villa, Autonomic imbalance during apneic episodes in pediatric obstructive sleep apnea, Clin. Neurophysiol. 127 (2016) 551–555, https://doi.org/10.1016/j.clinph.2015.05.025.

[12] R.B. Shouldice, L.M. O'Brien, C. O'Brien, P. De Chazal, D. Gozal, C. Heneghan, Detection of obstructive sleep apnea in pediatric subjects using surface lead electrocardiogram features, Sleep 27 (2004) 784–792, https://doi.org/10.1093/sleep/27.4.784.

[13] R. Tauman, D. Gozal, Obstructive sleep apnea syndrome in children, Expet Rev. Respir. Med. 5 (2011) 425–440, https://doi.org/10.1586/ers.11.7.

[14] C.L. Marcus, L.J. Brooks, S.D. Ward, K.A. Draper, D. Gozal, A.C. Halbower, J. Jones, C. Lehmann, M.S. Schechter, S. Sheldon, R.N. Shiffman, K. Spruyt, Diagnosis and management of childhood obstructive sleep apnea syndrome, Pediatrics 130 (2012) e714–e755, https://doi.org/10.1542/peds.2012-1672.

[15] A. Baharav, S. Kotagal, B.K. Rubin, J. Pratt, S. Akselrod, Autonomic cardiovascular control in children with obstructive sleep apnea, Clin. Auton. Res. 9 (1999) 345–351, https://doi.org/10.1007/BF02318382.

[16] R.S.C. Horne, G. Shandler, K. Tamanyan, A. Weichard, A. Odoi, S.N. Biggs, M. J. Davey, G.M. Nixon, L.M. Walter, The impact of sleep disordered breathing on cardiovascular health in overweight children, Sleep Med. 41 (2018) 58–68, https://doi.org/10.1016/j.sleep.2017.09.012.

[17] Y. Wu, L. Tian, D. Ma, P. Wu, Y. Tang, X. Cui, Z. Xu, Autonomic nervous function and low-grade inflammation in children with sleep-disordered breathing, Pediatr. Res. (2021) 1–7, https://doi.org/10.1038/s41390-021-01691-4.

[18] D. Liao, X. Li, A.N. Vgontzas, J. Liu, S. Rodriguez-Colon, S. Calhoun, E.O. Bixler, Sleep-disordered breathing in children is associated with impairment of sleep stage-specific shift of cardiac autonomic modulation, J. Sleep Res. 19 (2010) 358–365, https://doi.org/10.1111/j.1365-2869.2009.00807.x.

[19] L.C. Nisbet, S.R. Yiallourou, G.M. Nixon, S.N. Biggs, M.J. Davey, J. Trinder, L. M. Walter, R.S.C. Horne, Nocturnal autonomic function in preschool children with sleep-disordered breathing, Sleep Med. 14 (2013) 1310–1316, https://doi.org/10.1016/j.sleep.2013.07.010.

[20] L.M. Walter, G.M. Nixon, M.J. Davey, V. Anderson, A.M. Walker, R.S.C. Horne, Autonomic dysfunction in children with sleep disordered breathing, Sleep Breath. 17 (2013) 605–613, https://doi.org/10.1007/s11325-012-0727-x.

[21] A. Vlahandonis, S.R. Yiallourou, S.A. Sands, G.M. Nixon, M.J. Davey, L.M. Walter, R.S.C. Horne, Long-term changes in heart rate variability in elementary school-aged children with sleep-disordered breathing, Sleep Med. 15 (2014) 76–82, https://doi.org/10.1016/j.sleep.2013.06.023.

[22] A. Martín-Montero, G.C. Gutiérrez-Tobal, L. Kheirandish-Gozal, J. Jiménez-García, D. Álvarez, F. del Campo, D. Gozal, R. Hornero, Heart rate variability spectrum characteristics in children with sleep apnea, Pediatr. Res. 89 (2021) 1771–1779, https://doi.org/10.1038/s41390-020-01138-2.

[23] A. Martín-Montero, G.C. Gutiérrez-Tobal, L. Kheirandish-Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo, D. Gozal, R. Hornero, Heart rate variability as a potential biomarker of pediatric obstructive sleep apnea resolution, Sleep 45 (2022) 1–9, https://doi.org/10.1093/sleep/zsab214.

[24] S. Redline, R. Amin, D. Beebe, R.D. Chervin, S.L. Garetz, B. Giordani, C.L. Marcus, R.H. Moore, C.L. Rosen, R. Arens, D. Gozal, E.S. Katz, R.B. Mitchell, H. Muzumdar, H.G. Taylor, N. Thomas, S. Ellenberg, The childhood adenotonsillectomy trial (CHAT): rationale, design, and challenges of a randomized controlled trial evaluating a standard surgical procedure in a pediatric population, Sleep 34 (2011) 1509–1517, https://doi.org/10.5665/sleep.1388.

[25] C.L. Marcus, R.H. Moore, C.L. Rosen, B. Giordani, S.L. Garetz, H.G. Taylor, R. B. Mitchell, R. Amin, E.S. Katz, R. Arens, S. Paruthi, H. Muzumdar, D. Gozal, N. H. Thomas, D.B. Janice Ware, K. Snyder, L. Elden, R.C. Sprecher, P. Willging, D. Jones, J.P. Bent, T. Hoban, R.D. Chervin, S.S. Ellenberg, S. Redline, A randomized trial of adenotonsillectomy for childhood sleep apnea, N. Engl. J. Med. 368 (2013) 2366–2376, https://doi.org/10.1056/NEJMoa1215881.

[26] C. Iber, S. Ancoli-Israel, A.L. Chesson, S.F. Quan, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, AASM Man, 2007.

[27] D. Benitez, P.A. Gaydecki, A. Zaidi, A.P. Fitzpatrick, The use of the Hilbert transform in ECG signal analysis, Comput. Biol. Med. 31 (2001) 399–406. http://www.ncbi.nlm.nih.gov/pubmed/11535204.

[28] A. Martín-Montero, G.C. Gutiérrez-Tobal, D. Gozal, V. Barroso-García, D. Álvarez, F. del Campo, L. Kheirandish-Gozal, R. Hornero, Bispectral analysis of heart rate variability to characterize and help diagnose pediatric sleep apnea, Entropy 23 (2021) 1016, https://doi.org/10.3390/e23081016.

[29] T. Penzel, J.W. Kantelhardt, L. Grote, J.H. Peter, A. Bunde, Comparison of detrended fluctuation analysis and spectral analysis of heart rate variability in sleep and sleep apnea, IEEE Trans. Biomed. Eng. 50 (2003) 1143–1151, https://doi.org/10.1109/cic.2003.1291152.

[30] S. Fleming, M. Thompson, R. Stevens, C. Heneghan, A. Plüddemann, I. MacOnochie, L. Tarassenko, D. Mant, Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies, Lancet 377 (2011) 1011–1018, https://doi.org/10.1016/S0140-6736(10)62226-X.

[31] F. Vaquerizo-Villar, D. Alvarez, L. Kheirandish-Gozal, G.C. Gutierrez-Tobal, V. Barroso-Garcia, E. Santamaria-Vazquez, F. Del Campo, D. Gozal, R. Hornero, A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea, IEEE J. Biomed. Heal. Inf. 25 (2021) 2906–2916, https://doi.org/10.1109/JBHI.2020.3048901.

[32] J. Jiménez-García, G.C. Gutiérrez-Tobal, M. García, L. Kheirandish-Gozal, A. Martín-Montero, D. Álvarez, F. del Campo, D. Gozal, R. Hornero, Assessment of airflow and oximetry signals to detect pediatric sleep apnea-hypopnea syndrome using AdaBoost, Entropy 22 (2020) 670, https://doi.org/10.3390/e22060670.

[33] R. Hornero, L. Kheirandish-Gozal, G.C. Gutiérrez-Tobal, M.F. Philby, M.L. Alonso-Álvarez, D. Alvarez, E.A. Dayyat, Z. Xu, Y.S. Huang, M.T. Kakazu, A.M. Li, A. Van Eyck, P.E. Brockmann, Z. Ehsan, N. Simakajornboon, A.G. Kaditis, F. Vaquerizo-Villar, A.C. Sedano, O.S. Capdevila, M. Von Lukowicz, J. Terán-Santos, F. Del Campo, C.F. Poets, R. Ferreira, K. Bertran, Y. Zhang, J. Schuen, S. Verhulst, D. Gozal, Nocturnal oximetry-based evaluation of habitually snoring children, Am. J. Respir. Crit. Care Med. 196 (2017) 1591–1598, https://doi.org/10.1164/rccm.201705-0930OC.

[34] U.R. Acharya, K.P. Joseph, N. Kannathal, C.M. Lim, J.S. Suri, Heart rate variability: a review, Med. Biol. Eng. Comput. 44 (2006) 1031–1051, https://doi.org/10.1007/s11517-006-0119-0.

[35] F. Shaffer, J.P. Ginsberg, An overview of heart rate variability metrics and norms, Front. Public Health 5 (2017) 1–17, https://doi.org/10.3389/fpubh.2017.00258.

[36] G.C. Gutiérrez-Tobal, D. Álvarez, F. Vaquerizo-Villar, A. Crespo, L. Kheirandish-Gozal, D. Gozal, F. del Campo, R. Hornero, Ensemble-learning regression to estimate sleep apnea severity using at-home oximetry in adults, Appl. Soft Comput. 111 (2021), 107827, https://doi.org/10.1016/j.asoc.2021.107827.

[37] G.C. Gutierrez-Tobal, D. Alvarez, A. Crespo, F. del Campo, R. Hornero, Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings, IEEE J. Biomed. Heal. Inf. 23 (2019) 882–892, https://doi.org/10.1109/JBHI.2018.2823384.

[38] G.C. Gutierrez-Tobal, D. Alvarez, F. del Campo, R. Hornero, Utility of AdaBoost to detect sleep apnea-hypopnea syndrome from single-channel airflow, IEEE Trans. Biomed. Eng. 63 (2016) 636–646, https://doi.org/10.1109/TBME.2015.2467188.

[39] V. Barroso-García, G.C. Gutiérrez-Tobal, D. Gozal, F. Vaquerizo-Villar, D. Álvarez, F. del Campo, L. Kheirandish-Gozal, R. Hornero, Wavelet analysis of overnight airflow to detect obstructive sleep apnea in children, Sensors 21 (2021) 1491, https://doi.org/10.3390/s21041491.

[40] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, third ed., Elsevier, 2011 https://doi.org/10.1016/C2009-0-19715-5.

[41] P. Bühlmann, T. Hothorn, Boosting algorithms: regularization, prediction and model fitting, Stat. Sci. 22 (2007) 477–505, https://doi.org/10.1214/07-STS242.

[42] P. Bühlmann, B. Yu, Boosting with the L 2 loss, J. Am. Stat. Assoc. 98 (2003) 324–339, https://doi.org/10.1198/016214503000125.

[43] M. Deviaene, D. Testelmans, B. Buyse, P. Borzee, S. Van Huffel, C. Varon, Automatic screening of sleep apnea patients based on the SpO 2 signal, IEEE J. Biomed. Heal. Inf. 23 (2019) 607–617, https://doi.org/10.1109/JBHI.2018.2817368.

[44] J.H. Friedman, J.J. Meulman, Multiple additive regression trees with application in epidemiology, Stat. Med. 22 (2003) 1365–1381, https://doi.org/10.1002/sim.1501.
[45] J. Elith, J.R. Leathwick, T. Hastie, A working guide to boosted regression trees, J. Anim. Ecol. 77 (2008) 802–813, https://doi.org/10.1111/j.1365-2656.2008.01390.x.
[46] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.
[47] M. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?": explaining the predictions of any classifier, in: Proc. 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Demonstr., Association for Computational Linguistics, Stroudsburg, PA, USA, 2016, pp. 97–101, https://doi.org/10.18653/v1/N16-3020.
[48] G.M. Sullivan, R. Feinn, Using effect size—or why the P value is not enough, J. Grad. Med. Educ. 4 (2012) 279–282, https://doi.org/10.4300/JGME-D-12-00156.1.
[49] M. Tomczak, E. Tomczak, The need to report effect size estimates revisited. An overview of some recommended measures of effect size, Trends Sport Sci. 1 (2014) 19–25. http://www.wbc.poznan.pl/Content/325867/5_Trends_Vol21_2014_no1_20.pdf.
[50] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, Routledge, 1998.
[51] E. Gil, R. Bailón, J.M. Vergara, P. Laguna, PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children, IEEE Trans. Biomed. Eng. 57 (2010) 1079–1088, https://doi.org/10.1109/TBME.2009.2037734.
[52] J. Lázaro, E. Gil, J.M. Vergara, P. Laguna, 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. Inf. 18 (2014) 240–246, https://doi.org/10.1109/JBHI.2013.2267096.
[53] E. Gil, M. Mendez, J.M. Vergara, S. Cerutti, A.M. Bianchi, P. Laguna, Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of ppg signal in children by HRV analysis, IEEE Trans. Biomed. Eng. 56 (2009) 1005–1014, https://doi.org/10.1109/TBME.2008.2009340.
[54] H. Korkalainen, J. Aakko, B. Duce, S. Kainulainen, A. Leino, S. Nikkonen, I. O. Afara, S. Myllymaa, J. Töyräs, T. Leppänen, Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea, Sleep 43 (2020) 1–10, https://doi.org/10.1093/sleep/zsaa098.
[55] H. Korkalainen, T. Leppanen, J. Aakko, S. Nikkonen, S. Kainulainen, A. Leino, B. Duce, I.O. Afara, S. Myllymaa, J. Toyras, Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea, IEEE J. Biomed. Heal. Inf. 24 (2019), https://doi.org/10.1109/JBHI.2019.2951346, 1–1.
[56] F. Vaquerizo-Villar, D. Álvarez, G.C. Gutiérrez-Tobal, F. Del Campo, L. Kheirandish-Gozal, D. Gozal, T. Penzel, R. Hornero, A convolutional neural network to classify sleep stages in pediatric sleep apnea from pulse oximetry signals, in: 21St IEEE Mediterr. Electrotech. Conf., IEEE Melecon 2022, Palermo (Italia), 2022, pp. 108–113.
[57] H.V. Muzumdar, S. Sin, M. Nikova, G. Gates, D. Kim, R. Arens, Changes in heart rate variability after adenotonsillectomy in children with obstructive sleep apnea, Chest 139 (2011) 1050–1059, https://doi.org/10.1378/chest.10-1555.
[58] V.L. Schechtman, R.M. Harper, The maturation of correlations between cardiac and respiratory measures across sleep states in normal infants, Sleep 15 (1992) 41–47, https://doi.org/10.1093/sleep/15.1.41.
[59] N.A. Sazonova, E.E. Sazonov, B. Tan, S.A.C. Schuckers, CHIME study group, sleep state scoring in infants from respiratory and activity measurements, in: 2006 Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE, 2006, pp. 2462–2465, https://doi.org/10.1109/IEMBS.2006.259719.
[60] P.I. Terrill, S.J. Wilson, S. Suresh, D.M. Cooper, C. Dakin, Application of recurrence quantification analysis to automatically estimate infant sleep states using a single channel of respiratory data, Med. Biol. Eng. Comput. 50 (2012) 851–865, https://doi.org/10.1007/s11517-012-0918-4.
[61] J.R. Isler, T. Thai, M.M. Myers, W.P. Fifer, An automated method for coding sleep states in human infants based on respiratory rate variability, Dev. Psychobiol. 58 (2016) 1108–1115, https://doi.org/10.1002/dev.21482.
[62] C.L. Marcus, J.L. Carroll, D. Donnelly, G.M. Loughlin, Sleep in Children and Sleep and Breathing in Children: Developmental Changes in Sleep Patterns, Second, CRC Press, New York, 2008.
[63] G.G. Haddad, H.J. Jeng, T.L. Lai, R.B. Mellins, Determination of sleep state in infants using respiratory variability, Pediatr. Res. 21 (1987) 556–562, https://doi.org/10.1203/00006450-198706000-00010.
[64] R.M. Harper, V.L. Schechtman, K.A. Kluge, Machine classification of infant sleep state using cardiorespiratory measures, Electroencephalogr. Clin. Neurophysiol. 67 (1987) 379–387, https://doi.org/10.1016/0013-4694(87)90126-X.
[65] A. Lewicke, E. Sazonov, M.J. Corwin, M. Neuman, S. Schuckers, Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models, IEEE Trans. Biomed. Eng. 55 (2008) 108–118, https://doi.org/10.1109/TBME.2007.900558.
[66] R.B. Berry, R. Budhiraja, D.J. Gottlieb, D. Gozal, C. Iber, V.K. Kapur, C.L. Marcus, R. Mehra, S. Parthasarathy, S.F. Quan, S. Redline, K.P. Strohl, S.L.D. Ward, M. M. Tangredi, 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) 597–619, https://doi.org/10.5664/jcsm.2172, 08.