









# Obstructive Sleep Apnea Screening by Joint Saturation Signal Analysis and PPG-Derived Pulse Rate Oscillations

Diego Cajal , Eduardo Gil , Pablo Laguna , *Fellow, IEEE*, Carolina Varon , Dries Testelmans , Bertien Buyse , Chris Jensen, Rohan Hoare, Raquel Bailón , *Member, IEEE*, and Jesús Lázaro 

**Abstract**—Obstructive sleep apnea (OSA) is a high-prevalence disease in the general population, often underdiagnosed. The gold standard in clinical practice for its diagnosis and severity assessment is the polysomnography, although in-home approaches have been proposed in recent years to overcome its limitations. Today's ubiquitously presence of wearables may become a powerful screening tool in the general population and pulse-oximetry-based techniques could be used for early OSA diagnosis. In this work, the peripheral oxygen saturation together with the pulse-to-pulse interval (PPI) series derived from photoplethysmography (PPG) are used as inputs for OSA diagnosis. Different models are trained to classify between normal and abnormal breathing segments (binary decision), and between normal, apneic and hypopneic segments (multi-class decision). The models obtained 86.27% and 73.07% accuracy for the binary and multiclass segment classification, respectively. A novel index, the cyclic variation of the heart rate index (CVHRI), derived from PPI's spectrum, is computed on the segments containing disturbed breathing,

representing the frequency of the events. CVHRI showed strong Pearson's correlation ( $r$ ) with the apnea-hypopnea index (AHI) both after binary ( $r=0.94$ ,  $p<0.001$ ) and multiclass ( $r=0.91$ ,  $p<0.001$ ) segment classification. In addition, CVHRI has been used to stratify subjects with AHI higher/lower than a threshold of 5 and 15, resulting in 77.27% and 79.55% accuracy, respectively. In conclusion, patient stratification based on the combination of oxygen saturation and PPI analysis, with the addition of CVHRI, is a suitable, wearable friendly and low-cost tool for OSA screening at home.

**Index Terms**—Obstructive sleep apnea (OSA), Hjorth parameters, pulse photoplethysmography (PPG), pulse-to-pulse interval (PPI), oxygen saturation, cyclic variation of the heart rate index (CVHRI).

## I. INTRODUCTION

**O**BSTRUCTIVE Sleep Apnea (OSA) is a syndrome caused by repetitive episodes of total or partial interruption of the respiratory flow during sleep due to blockades produced by intermittent relaxation of throat muscles. Obstructive respiratory events are the cause of sleep fragmentation, hypoxemia, hypercapnia and increased sympathetic activity [1]. The list of symptoms can include daytime sleepiness, cognitive impairment, memory loss [2], together with comorbidities such as hypertension, cerebrovascular artery disease, coronary artery disease, congestive heart failure and atrial fibrillation [3]. OSA prevalence ranges from 9% to 38% in the general adult population, being much higher in the elderly groups [4]. Furthermore, prevalence is expected to increase in the general population due to obesity and overweight epidemic [5]. OSA underdiagnosis was estimated as 93% for women and 82% for men by Young et al. [6], however, the increase in obesity prevalence together with the generalization of screening are factors that may have altered these statistics from then until now. For decades, the gold standard for diagnosis included polysomnography (PSG) performed in a clinical environment. The patient is requested to sleep in a medical center while he or she is continuously monitored, making this test uncomfortable and with some impact in the natural sleep. Recently, the use of out-of-center sleep testing with limited channels was included in the diagnostic criteria for adult OSA, although it commonly underestimates the number of obstructive respiratory events per hour as compared

Manuscript received 5 June 2023; revised 12 September 2023 and 30 October 2023; accepted 7 November 2023. Date of publication 10 November 2023; date of current version 5 January 2024. This work was supported in part by MCIN/ AEI /10.13039/501100011033/ under Grants PID2021-126734OB-C21 and PID-2022-138585OA-C32, in part by ERDF A way of making Europe, in part by MCIN/ AEI /10.13039/501100011033/ Grant PDC2021-120775-I00 and TED2021-131106B-I00, in part by the European Union NextGenerationEU/PRTR, in part by Gobierno de Aragón (Reference Group BSIcOs T39-20R), and in part by the University of Zaragoza under Project UZ2022-IAR-06. (Corresponding author: Diego Cajal.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by UZ Leuven under Application No. S60319.

Diego Cajal, Eduardo Gil, Pablo Laguna, Raquel Bailón, and Jesús Lázaro are with the Biomedical Signal Interpretation and Computational Simulation (BSICoS) Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, 50018 Zaragoza, Spain, and also with the Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), 28029 Madrid, Spain (e-mail: dcajal@unizar.es; edugilh@unizar.es; laguna@unizar.es; rbailon@unizar.es; jlazarop@unizar.es).

Carolina Varon is with the STADIUS, Department of Electrical Engineering, KU Leuven, 3000 Leuven, Belgium (e-mail: carolina.varon@esat.kuleuven.be).

Dries Testelmans and Bertien Buyse are with the Department of Pneumology, Leuven University Centre for Sleep and Wake Disorders, UZ Leuven, 3000 Leuven, Belgium (e-mail: dries.testelmans@uzleuven.be; bertien.buyse@uzleuven.be).

Chris Jensen and Rohan Hoare are with Brainmatterz, McKinney, TX 75071 USA (e-mail: chris.jensen@velentium.com; rhoare@brainmatterz.health).

Digital Object Identifier 10.1109/JBHI.2023.3331947

to PSG [3]. Obstructive respiratory events are usually measured by the apnea-hypopnea index (AHI). This index, being the total count of apneas and hypopneas normalized by the sleep time in hours, has been a matter of controversy since its introduction in OSA diagnosis and severity rating [7], [8]. Despite of this, AHI is still the main measurement in OSA diagnosis, as OSA is defined as a combination of symptoms or comorbidities together with an  $AHI \geq 5$ ; or an  $AHI \geq 15$ , even in absence of symptoms [3].

Early diagnosis of OSA is important as it can cause several major health issues [9]. OSA underdiagnosed would be reduced by the development of novel techniques for massive screenings in the general population. Among these techniques, the assessment of heart rate variability (HRV) is appealing since it can be applied to signals recorded at home using wearables. Variability of the heartbeat period is known to be related to sleep breathing disorders. Zwillich et al. [10] discovered that most apneas –excluding those without oxygen desaturations– are associated with bradycardia episodes, and that bradycardias became more marked when apnea length and oxyhemoglobin desaturation increases. In 1984, Guilleminault et al. [11] described the Cyclic Variation of the Heart Rate (CVHR), a pattern of bradycardia during apnea, followed by abrupt tachycardia on airflow restoration. This pattern has been an object of study, including frequency-domain analysis [12], morphology variations [13] and automatic detection [14]. Shiomi et al. [12] discovered an augmented very low frequency (VLF) component of heart rate (0.008–0.04 Hz) in OSA patients synchronized with episodes of absence of air exchange or hypoxemia, that occurred at a cycle length of 25–120 seconds. They also described a VLF peak during episodes of OSA, likely related to the CVHR oscillation frequency, itself related to the frequency of the apneas. Stein et al. [13] set a 20% of the sleeptime with CVHR as a threshold to predict  $AHI \geq 15$ . They suggested that, despite Guilleminault described CVHR as an effective monitoring of the OSA, it had not been used by the end of the 20th century due to technical difficulties. They also appointed, back in 2003, that this technique should be included as a part of routine Holter reports.

Pulse rate variability (PRV) is a well-known alternative that offers a high correlation with HRV even in non-stationary situations [15]. The main advantage is that acquisition is made by an optic sensor placed on the skin, rather than attached electrodes. This technique, called pulse photoplethysmography (PPG), is the most popular in wrist-worn devices worldwide. Khandoker et al. [2] demonstrated that PRV could be used to distinguish OSA events from normal breathing during sleep, although several variability measures were significantly different from the HRV reference during OSA events. Analogously, Lázaro et al. [16] demonstrated that PRV can be used as HRV surrogate in apnea detectors based on decreases of amplitude fluctuations of the PPG (DAP). Later, Lazazzera et al. [17] combined DAP, PRV and peripheral capillary oxygen saturation ( $SpO_2$ ) for OSA screening purposes in adults. In [18], Hayano et al. presented an automatic detection of the CVHR pattern from a PPG signal for its use in a commercial wearable watch device. This algorithm is based on the detection of every cycle on the pulse-to-pulse interval (PPI) signal. MagnUSDottir et al. [19] used

CVHR combined with cardiopulmonary coupling to identify sleep apnea.

In [20], a novel method of OSA screening based on CVHR was proposed and preliminarily evaluated with recordings from 15 subjects. CVHR was detected from PPI signal using its Hjorth parameters as inputs of a bagged trees model. Moreover, a frequency-based metric, the CVHR Index (CVHRI), was proposed for severity stratification, obtaining a Pearson's correlation ( $r$ ) of  $r = 0.68$  ( $p < 0.05$ ) with AHI. The  $SpO_2$  signal is added to the model in this work, hypothesizing that it may considerably improve segment classification outcomes given that it provides a different source of information of the apnea-generated hypoxia. In addition, the further inclusion of frequency-domain PRV metrics as predictors is studied in Appendix A. Different combinations of PPI and  $SpO_2$  inputs are used in order to understand their individual contribution. CVHRI is evaluated with recordings from 96 subjects as a potential metric for stratification of subjects with AHI higher or lower than a threshold, considered either 5 or 15, corresponding with OSA diagnosis thresholds with and without symptoms or comorbidities, respectively. The novelties of this work are summarized as follows:

- The use of Hjorth parameters as the only features of  $SpO_2$  and PPI signals for the classification of segments.
- The inclusion of the  $SpO_2$  signal to the model preliminary presented in [20]. A larger dataset of 96 subjects is also used, in comparison with the previous 15-subject dataset.
- The inclusion of PRV metrics as model inputs is also studied.
- The use of a new PPI-derived index, the CVHRI, for stratifying subjects between OSA/non-OSA.

The manuscript structure is as follows. The methodology is explained in Section II, where the database is detailed in Section II-A, the classification of segments is addressed in subsections II-B, and II-C, and II-D, as well as the stratification of subjects in Section II-E, and the statistical metrics to evaluate the performance of segment classification and subject stratification in Section II-F. In Section III, the results of segment classification are studied in Section III-A, correlation of CVHRI with AHI in Section III-B, and subject stratification in Section III-C. Section IV is dedicated to the discussion of the results and is organized discussing segment classification in Section IV-A, correlation of CVHR with AHI in Section IV-B, and subject stratification in Section IV-C. This section includes an additional Section IV-D focused on the study of limitations. The manuscript is finished with conclusions presented in Section V.

## II. METHODS

### A. Dataset

The dataset is composed of 96 subjects (age  $44.5 \pm 11.4$  years, 62 males) suspected to suffer from OSA, who underwent a PSG (Medatec, Brainnet II, Brussels, Belgium) at the sleep laboratory of the University Hospitals Leuven (UZ Leuven). All included patients did not suffer from any of the following comorbidities: atrial fibrillation, hypertension, stroke, myocardial infarction, hyperlipidemia or diabetes. All signals were sampled at 500 Hz,

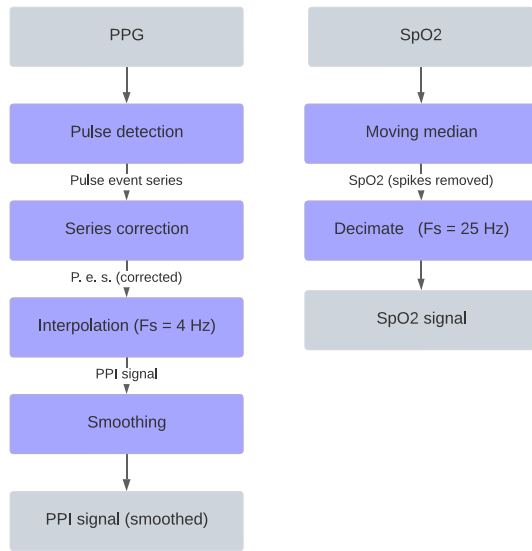


Fig. 1. Signal processing flowchart of signals used in for segment classification.

including nasal pressure and oronasal flow (thermistor). Hypnograms are also available. PPG and SpO<sub>2</sub> signals were recorded using a Nonin 8000 J sensor at 500 Hz. One subject was removed from dataset due to he/she was wake most of the test time, whereas another one was removed due to an unreliable nasal pressure signal. Thus, in total, 94 subjects were used. 72% of the subjects had  $AHI \geq 5$ , while 50% had  $AHI \geq 15$ , based on AASM annotation rules [21]. The inclusion of these data sets was approved by the ethical committee of UZ Leuven (S60319) and all patients signed an informed consent.

## B. Signals for Segment Classification

1) *Pulse-to-Pulse Intervals*: The PPG signal is processed to obtain pulse event series using an adaptive threshold pulse detector [16]. Then, pulse series are checked using the algorithm described in [22], correcting both false positives and false negatives. Finally, the PPI signal is obtained by evenly sampling the pulse series at 4 Hz using linear interpolation. PPI signal is also smoothed using a second-order polynomial fitting with a moving window of 20 seconds (see Fig. 1).

2) *SpO<sub>2</sub>*: SpO<sub>2</sub> values are quantified in integer units and the lack of hysteresis provokes large quantization noise. A 3-second median filter is used to reduce noise in the SpO<sub>2</sub> signal. SpO<sub>2</sub> is also decimated to 25 Hz, following the AASM 2012 recommendations [21].

## C. References for Performance Evaluation

1) *Airflow*: The AASM [21] recommends different sensors to annotate apneas and hypopneas: apneas are proposed to be annotated from oronasal thermistor, while hypopneas are from nasal pressure. However, oronasal thermistor signals were saturated in most of the cases, making them not reliable. Thus, following the AASM guidelines for this cases, both apneas and hypopneas are annotated from nasal pressure signals.

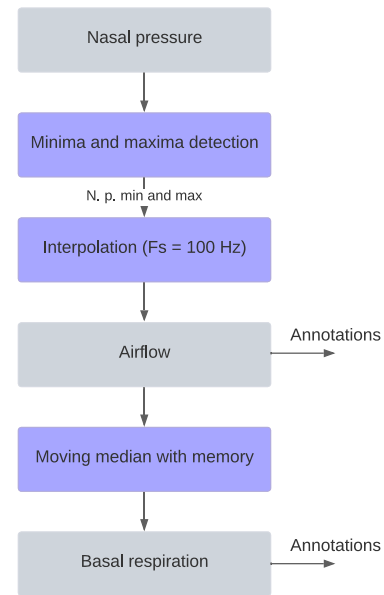
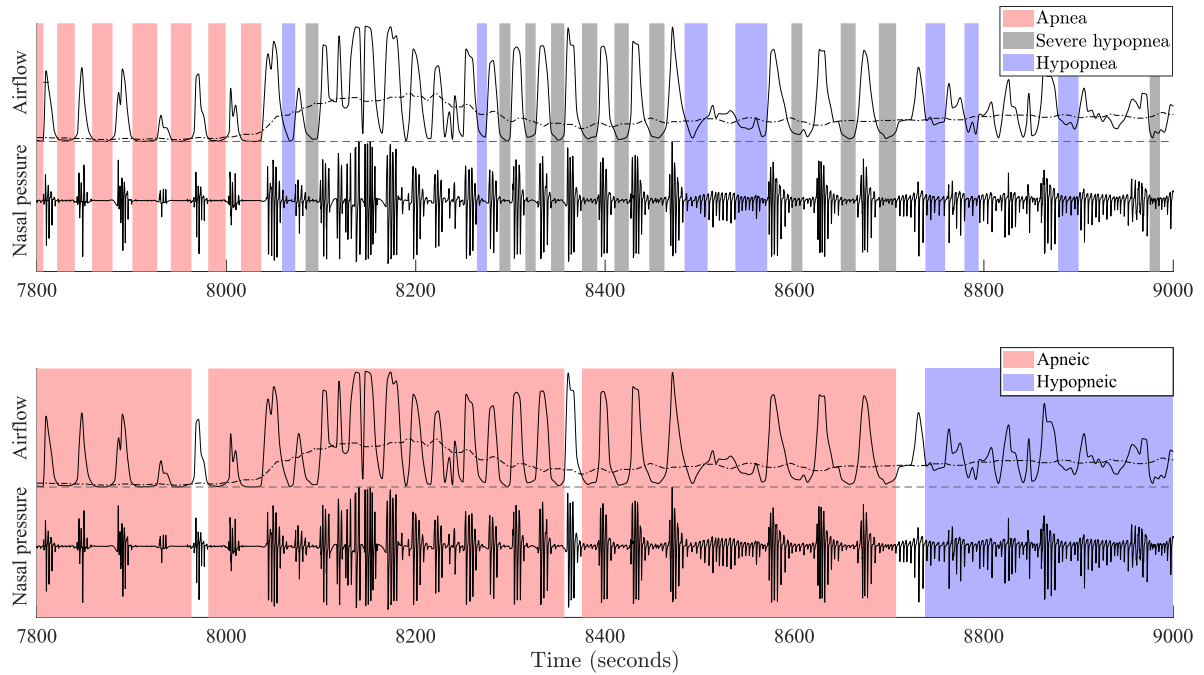


Fig. 2. Signal processing flowchart of reference signals.

First, nasal pressure signals are low-pass filtered at 15 Hz for noise removal, and detrended by a high-pass filter at 0.1 Hz, using 3-order Butterworth filters (see Fig. 2). Airflow is computed from the nasal pressure signal by the algorithm described in [23] by detection of maxima (for positive segments) or minima (for negative segments) between consecutive zero crossings. These minima and maxima are interpolated using piecewise cubic Hermite interpolating polynomials, obtaining a positive and a negative envelope. Finally, the airflow is defined as the difference between the envelopes. Airflow signal is decimated to 100 Hz, following the AASM 2012 recommendations [21].

2) *Basal Respiration*: Running basal respiration was used as reference for annotating reductions of the airflow signal. This signal is obtained by an algorithm that computes the median of the airflow signal in 1-minute segments. The result of the sum of each segment median, multiplied by a weight of 0.4, plus the previous segment median, multiplied by 0.6, is stored. These weights were chosen empirically, looking for the smoothest line that at the same time allowed to follow the variations in the baseline in a subset of 10 random subjects. The use of averaging with memory helps to obtain a more accurate basal respiration in signals with the presence of apneas. Without these disruptions, averaging is a simple task and no weighting is required. However, subjects with apneas present regions with large variations, sometimes composed of a succession of events. Therefore, a compromise must be reached that allows these disruptions not to raise or lower the basal respiration to arbitrary values, while allowing the average to follow the changes over time. The more challenging decision is in those cases with burst events, where respiration does not return to a basal respiration between events. In these cases, basal values before and after the burst must be taken into account to make a correct approximation. Thus, once the algorithm obtains a value for each airflow segment, it is run again backwards. Finally, basal respiration is the mean of the



**Fig. 3.** *Top panel:* Event-based annotations. Solid lines represent airflow (top row) and nasal pressure (bottom row), the dashed-dotted line represents basal respiration and the dashed line represents zero-reference for airflow and basal respiration. Annotation onsets and endsets correspond to airflow reductions and restorations, respectively. *Bottom panel:* Grouped annotations. Events are grouped in apneic/hypopneic bursts. Note that new bursts are initiated every 8 events, allowing precise apneic/hypopneic characterization. These annotations are later transformed into segment-based annotations.

forward and backward results. This allows for more accurate transitions between normal and disrupted breathing segments.

**3) Annotations:** Events are labeled as *apnea* if airflow decreases  $\geq 90\%$  from basal respiration, during  $\geq 10$  seconds, while they are labeled as *hypopnea* if the decrease is  $\geq 30\%$  during  $\geq 10$  seconds and there is an associated  $\geq 3\%$  desaturation. A third label, *severe hypopnea*, is applied to airflow decreases  $\geq 70\%$  during  $\geq 10$  seconds, regardless of saturation. These are borderline cases, in which airflow does not completely disappear, although the reduction is considerably greater than in most hypopneas (See Section IV-A). Hypopneas related to arousals were not annotated. No distinction was made between central and obstructive apnea/hypopnea annotations, although respiratory effort was assessed (89.7% of apneas/hypopneas were obstructive [17]).

Event-based annotations were transformed into segment-based annotations. This step is performed taking into account that the objective is to annotate each segment into *abnormal/normal breathing* (binary decision) or *apneic/hypopneic/normal breathing* (multiclass decision) as a reference for segment classification. Events are grouped in bursts. Bursts are composed of at least two events, separated by a maximum of 180 seconds, and are labeled as apneic or hypopneic depending on the events forming each burst. A burst is labeled as *apneic* if it contains at least one *apnea* or at least half of the events are *severe hypopneas*. Bursts are labeled as *hypopneic* otherwise. The maximum number of grouped events is empirically set to eight, allowing precise *apneic/hypopneic* characterization. An example of event-based annotation grouping in bursts is shown in Fig. 3.

Finally, time is divided in 180-second segments with 150-second overlap (step: 30 seconds). A segment is labeled as *apneic* if it contains at least one *apneic* burst; as *hypopneic* if it contains at least one *hypopneic* burst and no *apneic* bursts; or as *normal breathing* if it does not contain any burst. *Apneic* and *hypopneic* classes are grouped into *abnormal breathing* in the binary case. Classes were balanced by randomly removing the majority class before segment classification, obtaining a total of 53.242 segments in the binary case (26.621 *normal* and 26.621 *abnormal breathing* segments), and 25.278 segments in the multiclass case (8.426 *apneic*, 8.426 *hypopneic* and 8.426 *normal breathing* segments).

#### D. Segment Classification Models

For each segment, Hjorth parameters  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  and  $\mathcal{H}_2$  of the PPI and SpO<sub>2</sub> signals, surrogates of power, dominant frequency and bandwidth, respectively [24], [25], were computed in a sliding window following the expressions:

$$\begin{aligned} \text{Activity} : \mathcal{H}_0(m) &= \bar{w}_0(m) \\ \text{Mobility} : \mathcal{H}_1(m) &= \sqrt{\frac{\bar{w}_2(m)}{\bar{w}_0(m)}} \\ \text{Complexity} : \mathcal{H}_2(m) &= \sqrt{\frac{\bar{w}_4(m)}{\bar{w}_2(m)} - \frac{\bar{w}_2(m)}{\bar{w}_0(m)}}, \end{aligned} \quad (1)$$

where  $\bar{w}_i$  is the  $i$ -th-order spectral moment.  $\bar{w}_i$  can be estimated using the temporal expression of the moments in the  $m$ -th

window of  $P$  samples:

$$\hat{w}_i(m) \approx \frac{2\pi}{P} \sum_{n=(m-1)P+1}^{mP} (x^{i/2}(n))^2, \quad (2)$$

being  $x(n)$  either the PPI or the SpO<sub>2</sub> signal and  $P$  the number of samples corresponding to 180 s. The Hjorth parameters are used as inputs for segment classification. The motivation for using these parameters is twofold. On the one hand, they are simple and low cost to compute since they can be estimated from the time domain signal. On the other hand, they are easily interpretable, being related to the signal energy, dominant frequency and bandwidth. The original hypothesis was that segments with CVHR pattern would have different Hjorth parameters from segments without CVHR [20]. In particular, it was hypothesized that the PPI in segments with CVHR would have a lower Complexity,  $\mathcal{H}_2$ , as the bradycardia-tachycardia pattern would mask the normal variability of the heart, causing it to more closely resemble a sinusoid, as shown in [20]. Analogously, desaturations in SpO<sub>2</sub> should affect all parameters, especially the Activity,  $\mathcal{H}_0$ .

The best model is selected maximizing the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve. For training, 5-fold cross-validation was performed in order to avoid bias in results due to overfitting. Models from the decision trees, discriminant analysis, logistic regression, naive bayes, support vector machine, nearest neighbor, kernel approximation, ensembles, neural networks families were tested. As result, Bagged trees outperformed the others and was the selected model, and a leave-one-subject-out testing strategy was followed. Different models were created for binary and multiclass decision. Also, for each classification strategy, three models were created, depending if they use PPI and SpO<sub>2</sub> (PPI+SpO<sub>2</sub> model), only PPI (PPI model) and only SpO<sub>2</sub> (SpO<sub>2</sub> model) as inputs (see Fig. 4).

### E. OSA Stratification by CVHRI

CVHRI is a metric proposed in [20] for apnea severity quantification. The spectrum of the PPI is computed using Fast Fourier Transformation (FFT) for each  $i$ -th 180-second segment classified as *abnormal breathing* in the binary case, or as *apneic/hypopneic* in the multiclass case. The frequency of the FFT modulus maxima,  $F_1^{\max}$ , between 0 and 0.1 Hz is obtained. Then, CVHRI is defined as the sum of the frequencies of the spectrum peaks of each *abnormal breathing/apneic/hypopneic* segment divided by the total number of segments, obtaining a single parameter which characterizes each patient, similarly to AHI.

$$\text{CVHRI} = \frac{\sum_{i=1}^{I_{ab}} F_1^{\max}}{I_{tot}}, \quad (3)$$

where  $I_{ab}$  is the total number of *abnormal breathing* segments in the binary case or the sum of *apneic* and *hypopneic* segments in the multiclass case and  $I_{tot}$  is the total number of segments. Pearson correlation coefficient (Pearson's  $r$ ) between CVHRI and AHI is computed for each model. This index was



Fig. 4. Segment classification models.

proposed in [20] as an alternative for detecting each bradycardia-tachycardia pattern individually, i.e., the CVHR pattern, in order to be less costly and more robust. The former is achieved because only one peak per segment must be detected, which is easier in the majority of the cases, while the latter benefits from the highly optimized FFT algorithms.

Finally, CVHRI is used for subject stratification. Groups of interest are those clustered by  $\text{AHI} < 5$  v.  $\text{AHI} \geq 5$  and  $\text{AHI} < 15$  v.  $\text{AHI} \geq 15$ , as are the ones used for OSA diagnosis. A CVHRI threshold for each subgroup is searched by Linear Discriminant Analysis (LDA). Class weights  $w(j)$  are computed for dealing with imbalanced data following:

$$w(j) = \frac{N}{2N(j)}, \quad (4)$$

where  $N$  is the total number of patients and  $N(j)$  the number of patients corresponding to class  $j$ . Train and test groups are selected randomly, splitting the dataset in two halves. 5-fold cross-validation is used to prevent overfitting during training.

### F. Performance Analysis

Segment classification performance is evaluated in terms of accuracy ( $Acc$ ) precision ( $P$ ) and recall ( $R$ ). These metrics have been evaluated for all subjects and also for  $\text{AHI} < 15$  and  $\text{AHI} \geq 15$  subsets. The leave-one-subject-out strategy is implemented by summing up the number of false and true events for each left subject and then computing metrics presented in Tables I and II. After segment classification, CVHRI is computed for every subject and compared to AHI. Pearson's  $r$  correlation was computed between CVHRI and AHI with a significance level of 0.05. Correlation results are divided as well in binary and multiclass depending on the segment classification prior to the CVHRI computing. Also, results are computed for all subjects and for

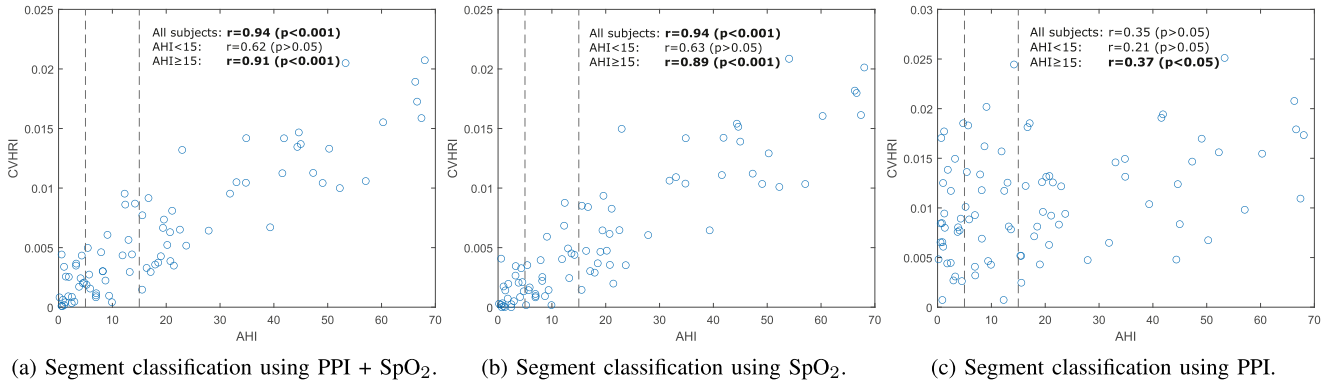


Fig. 5. AHI v. CVHRI after binary segment classification. Vertical lines separate AHI <5, 5 ≤ AHI <15 and AHI ≥15 groups.

TABLE I  
BINARY SEGMENT CLASSIFICATION METRICS

Class	Measure	Definition
All	Accuracy	$Acc = \frac{T_n + T_{ab}}{T_n + F_n + T_{ab} + F_{ab}}$
Normal breathing	Recall	$R_n = T_n / (T_n + F_{ab})$
	Precision	$P_n = T_n / (T_n + F_n)$
Abnormal breathing	Recall	$R_{ab} = T_{ab} / (T_{ab} + F_n)$
	Precision	$P_{ab} = T_{ab} / (T_{ab} + F_{ab})$

TABLE II  
MULTICLASS SEGMENT CLASSIFICATION METRICS

Class	Measure	Definition
All	Accuracy	$Acc = \frac{T_n + T_{ap} + T_h}{T_n + F_n + T_{ap} + F_{ap} + T_h + F_h}$
Normal breathing	Recall	$R_n = T_n / (T_n + F_{ap,n} + F_{h,n})$
	Precision	$P_n = T_n / (T_n + F_n)$
Apneic	Recall	$R_{ap} = T_{ap} / (T_{ap} + F_{n,ap} + F_{h,ap})$
	Precision	$P_{ap} = T_{ap} / (T_{ap} + F_{ap})$
Hypopneic	Recall	$R_h = T_h / (T_h + F_{n,h} + F_{ap,h})$
	Precision	$P_h = T_h / (T_h + F_h)$

AHI <15 and AHI ≥15 subsets separately. Finally, stratification results were computed taking the AHI clustering (AHI <5 v. AHI ≥5 and AHI <15 v. AHI ≥15) as reference. Accuracy, positive predictive value (PPV), sensitivity (Se), negative predictive value (NPV), specificity (Sp), Area Under the Curve (AUC) and Cohen's Kappa ( $\kappa$ ) are reported alongside the best CVHRI threshold for each model that better cluster the subjects.

1) **Binary Segment Classification:** The target here is *normal* and *abnormal breathing* segment classification. A number of true *normal breathing* segments,  $T_n$ , false *normal breathing* segments,  $F_n$ , true *abnormal breathing* segments,  $T_{ab}$ , and false *abnormal breathing* segments,  $F_{ab}$ , are obtained, which are quantified by the metrics in Table I.

2) **Multiclass Segment Classification:** The target here is *normal breathing*, *apneic* or *hypopneic* segment classification. We obtain a number of true *normal breathing* segments,  $T_n$ , false *normal breathing* segments,  $F_n$ , divided between those coming from *apneic* and *hypopneic* segments ( $F_n = F_{n,ap} + F_{n,h}$ ), true *apneic* segments,  $T_{ap}$ , false *apneic* segments,  $F_{ap}$ , divided between those coming from *normal breathing* and *hypopneic* segments ( $F_{ap} = F_{ap,n} + F_{ap,h}$ ), true *hypopneic*

TABLE III  
BINARY SEGMENT CLASSIFICATION RESULTS (%)

Model	Subgroup	Acc	$P_n$	$R_n$	$P_{ab}$	$R_{ab}$
PPI+ SpO <sub>2</sub>	All subjects	85.01	90.89	87.01	73.52	80.53
	AHI <15	87.33	95.21	90.57	38.06	56.01
	AHI ≥15	83.01	84.33	81.53	81.76	84.53
SpO <sub>2</sub>	All subjects	86.27	91.30	88.56	76.05	81.16
	AHI <15	89.38	95.00	93.19	44.41	52.60
	AHI ≥15	83.59	85.43	81.41	81.89	85.81
PPI	All subjects	60.30	81.13	57.34	39.36	67.50
	AHI <15	57.93	94.66	57.17	12.73	65.98
	AHI ≥15	62.47	66.07	57.61	59.46	67.75

TABLE IV  
MULTICLASS SEGMENT CLASSIFICATION RESULTS (%)

Model	Subgroup	Acc	$P_n$	$R_n$	$P_{ap}$	$R_{ap}$	$P_h$	$R_h$
PPI+ SpO <sub>2</sub>	All subjects	73.07	92.86	80.19	65.91	60.70	23.71	49.87
	AHI <15	80.39	96.00	84.53	16.25	25.93	19.62	48.22
	AHI ≥15	66.77	87.75	73.48	73.08	63.43	25.70	50.51
SpO <sub>2</sub>	All subjects	71.71	93.86	77.18	78.21	57.54	22.91	63.52
	AHI <15	79.39	96.38	82.60	32.10	16.75	19.31	65.38
	AHI ≥15	65.10	89.52	68.82	80.72	60.74	24.79	62.80
PPI	All subjects	44.39	78.39	48.49	24.82	37.06	10.19	28.87
	AHI <15	45.77	93.64	47.26	4.33	39.97	5.34	24.61
	AHI ≥15	43.12	62.94	50.46	42.10	36.83	14.36	30.55

segments,  $T_h$ , and false *hypopneic* segments,  $F_h$ , divided between those coming from *normal breathing* and *apneic* segments ( $F_h = F_{h,n} + F_{h,ap}$ ), which are quantified by the metrics in Table II.

### III. RESULTS

#### A. Segment Classification Results

The results for binary and multiclass classification are shown in Tables III and IV, respectively. Results are given for all subjects together as well as separately in the AHI <15 and AHI ≥15 subgroups.

#### B. CVHRI Correlation With AHI

1) **Binary:** Fig. 5 shows CVHRI v. AHI scatter plots for each binary segment classification model, including Person's  $r$ . A

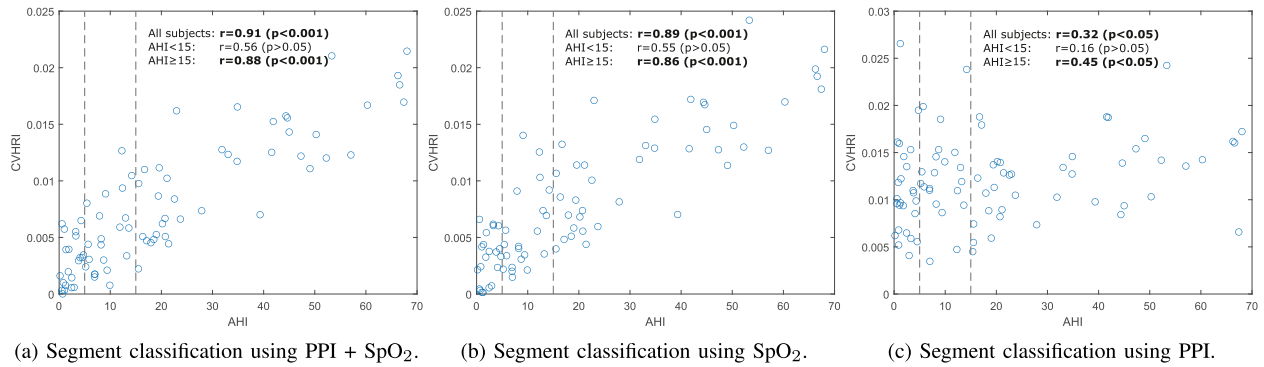


Fig. 6. AHI v. CVHRI after multiclass segment classification. Vertical lines separate AHI <5,  $5 \leq$  AHI <15 and AHI  $\geq$ 15 groups.

TABLE V  
STRATIFICATION AFTER BINARY SEGMENT CLASSIFICATION (% , EXCEPT AUC AND  $\kappa$ )

AHI $\geq$	Model	CVHRI Thresh.	Acc	Se	Sp	PPV	NPV	$\kappa$	AUC (Train)	AUC (Test)
5	PPI+SpO <sub>2</sub>	$4.0 \times 10^{-3}$	77.27	87.50	75.00	43.75	96.42	0.45	0.88	0.89
	SpO <sub>2</sub>	$4.0 \times 10^{-3}$	72.73	87.50	69.44	38.89	96.15	0.38	0.87	0.89
	PPI	$1.1 \times 10^{-3}$	51.06	66.67	47.37	23.08	85.71	0.08	0.72	0.56
15	PPI+SpO <sub>2</sub>	$5.4 \times 10^{-3}$	79.55	82.35	77.78	70.00	87.50	0.58	0.93	0.90
	SpO <sub>2</sub>	$6.3 \times 10^{-3}$	79.55	88.24	74.07	68.18	90.91	0.59	0.96	0.91
	PPI	$1.2 \times 10^{-3}$	59.57	57.14	61.54	54.55	64.00	0.19	0.67	0.57

TABLE VI  
STRATIFICATION AFTER MULTICLASS SEGMENT CLASSIFICATION (% , EXCEPT AUC AND  $\kappa$ )

AHI $\geq$	Model	CVHRI Thresh.	Acc	Se	Sp	PPV	NPV	$\kappa$	AUC (Train)	AUC (Test)
5	PPI+SpO <sub>2</sub>	$5.6 \times 10^{-3}$	75.00	75.00	75.00	40.00	93.10	0.37	0.88	0.86
	SpO <sub>2</sub>	$4.0 \times 10^{-3}$	72.73	75.00	72.22	37.50	92.86	0.34	0.86	0.86
	PPI	$1.2 \times 10^{-3}$	53.19	55.56	52.63	21.74	83.33	0.05	0.71	0.52
15	PPI+SpO <sub>2</sub>	$7.8 \times 10^{-3}$	79.55	88.24	74.07	68.18	90.91	0.59	0.92	0.88
	SpO <sub>2</sub>	$7.8 \times 10^{-3}$	77.27	82.35	74.07	66.67	86.96	0.54	0.93	0.90
	PPI	$1.4 \times 10^{-3}$	46.81	57.14	38.46	42.86	52.63	0.04	0.64	0.51

very strong correlation ( $r = 0.94$ ) was found when CVHRI segments are detected with both the PPI+SpO<sub>2</sub> model and the SpO<sub>2</sub> model. No correlation was found when CVHRI segments were detected by the PPI model, except for the group with AHI  $\geq$ 15, where a low correlation ( $r = 0.37$ ) was obtained. Correlation was slightly lower when including only the AHI  $\geq$ 15 group in comparison with all subject correlation, being 0.91 using the PPI+SpO<sub>2</sub> model and 0.89 using the SpO<sub>2</sub> model. No correlation was found in any case for AHI <15.

2) *Multiclass*: Fig. 6 shows CVHRI v. AHI scatter plots for each multiclass segment classification model. Results when using a multiclass classifier prior to the CVHRI computation are analogous to the binary case, obtaining slightly lower values. A very strong correlation was found using the PPI+SpO<sub>2</sub> model ( $r = 0.91$ ) and the SpO<sub>2</sub> model ( $r = 0.89$ ), while a low correlation was found when using the PPI model ( $r = 0.32$ ). Correlation was again slightly lower when including only the AHI  $\geq$ 15 group in comparison with all subject correlation, being  $r = 0.88$  using the PPI+SpO<sub>2</sub> model and  $r = 0.86$  using the SpO<sub>2</sub> model, with the exception of the PPI model, that increased its correlation to  $r = 0.45$ . No correlation was found in any case for AHI <15 either.

### C. OSA Stratification by CVHRI

Stratification results after binary and multiclass segment classification are shown in Tables V and VI, respectively.

## IV. DISCUSSION

The present work has found sufficient evidence for supporting the use of spectral features, extracted by Hjorth parameters, as models for OSA screening based on oximetry systems. These models stand out for their low computational cost, linearly proportional to the length  $P$  of the segment, in contrast to FFT, with a computational cost proportional to  $P \log_2(P)$ , or wavelet transform, also proportional to  $P$  but with a greater number of operations [26]–, being suitable for built-in wearable applications. Moreover, CVHRI computed on not *normal breathing* classified segments has demonstrated to be strongly correlated to AHI, implying that could be a proper AHI surrogate when the airflow is not available, specially in moderate-to-severe cases (AHI  $\geq$  15). However, the limitations that apply to AHI as a diagnosis and severity stand-alone index should also be attributed to CVHRI, making it just valuable as surrogate for

AHI. As future research, it could be interesting to include patients with comorbidities. As CVHR amplitude is mortality risk predictor [27], a new index could be included as the amplitude of the peak between 0 and 0.1 Hz of the PPI spectra used in CVHRI computation. In addition, CVHR is thought to reflect cardiac autonomic responses to cardio-respiratory perturbation caused by apneic/hypoxic episodes [27]. Consequently, it is possible that it could be useful in other disturbance assessments.

### A. Segment Classification

Both PPI+SpO<sub>2</sub> and SpO<sub>2</sub> models perform similar, obtaining 85.01% and 86.27% accuracy when including all subjects, respectively. Differences are high when comparing with the PPI model, which obtained a 60.30% accuracy. Both PPI+SpO<sub>2</sub> and SpO<sub>2</sub> models performed slightly better in the AHI<15 group, obtaining 87.33% and 89.38% accuracy, in comparison with the AHI≥15 group (83.01% and 83.51% accuracy). Despite the better accuracy, precision and recall are uneven between classes in the AHI<15 group, e.g., 95.21% (*normal* class) against 38.06% (*abnormal breathing* class) precision and 90.57% (*normal* class) against 56.01% (*abnormal breathing* class) recall in the PPI+SpO<sub>2</sub> model; being even in the AHI≥15 group, with ≥81.41% in all cases.

Multiclass results follow analogous trends as for the binary case. PPI+SpO<sub>2</sub> and SpO<sub>2</sub> models showed the best performance, with 73.07% and 71.71% accuracy, respectively, whereas PPI model obtained 44.39% accuracy. Both PPI+SpO<sub>2</sub> and SpO<sub>2</sub> models performed better in the AHI<15 group as well, obtaining 80.39% and 79.39% against 66.77.39% and 65.10% accuracy in the AHI≥15 group. Precision and recall are also uneven in the AHI<15 group in comparison with the AHI≥15 group. Most errors are confusions between *apneic* and *hypopneic breathing*: e.g., in the PPI+SpO<sub>2</sub> model, *normal breathing* precision and recall was 92.86% and 80.19%, while for *apneic/hypopneic breathing* was 65.91/23.71% precision and 60.70/49.87% recall. This differences are accentuated in the AHI < 15 group: e.g., in the PPI+SpO<sub>2</sub> model, *normal breathing* precision and recall was 96.00% and 84.53%, while for *apneic/hypopneic breathing* was 16.25/19.62% precision and 25.93/48.22% recall.

According to the results, most of the predictive capacity of the models rely on the SpO<sub>2</sub> signal, taking into account that PPI+SpO<sub>2</sub> models and SpO<sub>2</sub> models perform identically, in contrast with the poor performance of the PPI models. However, correlation between CVHRI and AHI was slightly higher using PPI+SpO<sub>2</sub> models. The inclusion of SpO<sub>2</sub>, evaluated in a higher number of subjects, has largely improve the results of the preliminary work [20]. It is possible that PPI could be useful to detect arousal-related hypopneas as [16], [28] suggest (see IV-D Section).

The overall segment classification performance worsens in the multiclass case relative to the binary case, although accuracies remain high (73.07%). Multiclass classifiers may be useful in future research, specially when including comorbidities, although nowadays there is no distinction in OSA treatment whether there is an apneic or hypopneic predominance. *Severe hypopneas* annotations (airflow reduction higher than 70% but lower than 90% related to desaturation) were introduced to improve *apneic*

TABLE VII  
MULTICLASS SEGMENT CLASSIFICATION RESULTS (%)

<i>Sev. hypo.</i>	<i>Acc</i>	<i>P<sub>n</sub></i>	<i>R<sub>n</sub></i>	<i>P<sub>ap</sub></i>	<i>R<sub>ap</sub></i>	<i>P<sub>h</sub></i>	<i>R<sub>h</sub></i>
Omitting	73.01	92.90	80.20	65.90	60.70	23.70	49.90
Using	73.07	92.86	80.19	65.91	60.70	23.71	49.87

Comparison between using/omitting *Severe hypopnea* label using the PPI+SpO<sub>2</sub> model

segment detections. 90% reduction of the airflow is an arbitrary threshold that attempts to operationalize the requirement of “absent or nearly absent airflow” [29]. This way, it was observed that borderline events labeled as *hypopneas* have comparable PPI and SpO<sub>2</sub> response to that of apneas rather than < 70% airflow reduction hypopneas. Nevertheless, analyzing the segment classification results, virtually no differences were found when the label *severe hypopnea* was omitted, i.e., strictly following the AASM rules, as it is shown in Table VII. This comparison has been computed using the best classification model (PPI+SpO<sub>2</sub>). The utility of this label for stratification will be discussed in Section IV-C.

Previous studies have classified apneic events based on SpO<sub>2</sub> and PPI. The approach of this work is different. In this case, there is no detection of apneic/hypopneic events, but rather a classification of segments, which subsequently allows CVHRI to be calculated. To the authors’ knowledge, there are no studies in which segments are classified, which is one of the main novelties of this study. The different approaches make the results not directly comparable with other works. As a reference, it is worth mentioning the event classification results of Lazazzera et al. [17], which obtained a 75.1% accuracy on the same database and using the same input signals. This result was obtained for multiclass classification, so it should be compared with the 73.03% (Table IV) of this work. Deviaene et al. [30] obtained an accuracy of 83.4% using SpO<sub>2</sub> and PPG features, in a database with 102 subjects, also recorded at UZ Leuven. In this study they reached the same conclusion that SpO<sub>2</sub> models outperform PPG models, obtaining an accuracy of 82.2% with the SpO<sub>2</sub> model. The authors concluded that it is better to use both inputs if available [30]. In a recent study by Huttunen et al. [31], the authors compare different combinations of signals used as inputs of a deep learning model that is able to simultaneously detect respiratory events and classify sleep stages. The authors compare three models: the first using PPG and SpO<sub>2</sub>, the second adding the nasal pressure, and the third using SpO<sub>2</sub>, nasal pressure and the electroencephalogram. Interestingly, the three obtain virtually the same results in estimating AHI, supporting the use of pulse oximeters in OSA screening without additional sensors.

### B. CVHRI Correlation With AHI

Best correlation between CVHRI and AHI was found in the AHI≥15 subgroup. This may appear counter-intuitive observing segment classification accuracies, that are higher in the AHI<15 subgroup. However, the *abnormal* class (or *apneic* + *hypopneic* classes in the multiclass model) is better detected in the AHI≥15 subgroup, according to precision and recall results, probably due to the fact that those are the cases with the most clear bursts of respiratory events. As CVHRI is measured only in abnormal (or



TABLE VIII  
COMPARISON WITH OTHER STUDIES (% , EXCEPT AUC AND  $\kappa$ )

AHI $\geq$	Model	Acc	Se	Sp	PPV	NPV	$\kappa$	AUC (Train)	AUC (Test)
5	PPI+SpO <sub>2</sub> binary	77	88	75	44	96	0.45	0.88	0.89
	Romem et al. [32]	-	80	86	93	68	0.67	0.91	-
	Fassbender et al. [33]	-	100	44	62	100	0.43	0.93	-
15	PPI+SpO <sub>2</sub> binary	80	82	78	70	88	0.59	0.93	0.90
	Romem et al. [32]	-	70	91	80	85	0.71	0.90	-
	Fassbender et al. [33]	-	92	77	60	96	0.59	0.95	-

apneic + hypopneic) segments, it is not as reliable in the AHI < 15 subgroup as in the AHI  $\geq$  15 subgroup.

### C. OSA Stratification by CVHRI

Analogously to the previous result sections, stratification accuracy decreases substantially when using PPI segment classification model. Similar trends are followed after multiclass decision. Established thresholds are a reliable tool for OSA diagnosis after PPI+SpO<sub>2</sub> and SpO<sub>2</sub> model, being the first slightly more accurate. Moreover, its use for screening purposes is supported by high negative predictive values. Bad results after PPI models were expected taking into account segment classification results. Binary models obtained slightly better results in comparison with multiclass models. This, in addition to the higher complexity of multiclass models suggest that binary models should be used for stratification purposes. Results are comparable with other researches that use PPG-derived metrics for OSA diagnosis (Table VIII), such as [32], [33], although none of these works used train-test splits nor cross-validation. Also, higher scores in [33] could be explained since airflow information was added to the model by using a nasal cannula. A comparison with [34] cannot be directly done as different groups were used ( $5 \leq \text{AHI} < 15$ ,  $15 \leq \text{AHI} < 30$  and  $\text{AHI} \geq 30$ ), obtaining  $\kappa$  values ranged from 0.49 to 0.79. It is reasonable to assume that these results are in the same order than in [32], as they used the same proprietary algorithm (Morpheus Ox. WideMed Ltd, Herzliya, Israel). Same AHI groups were used in [35]. In this work, AHI was estimated from SpO<sub>2</sub> using an artificial neural network. Estimated AHI classified patients in mentioned groups with 90.9% accuracy.

The inclusion of the label *severe hypopnea* also deserves discussion at this point. This label applies to events with an airflow reduction  $\geq 70\%$  for  $\geq 10$  seconds, regardless of desaturation. Therefore, by omitting this label, events with an airflow reduction between 70% and 90% –at which point they are classified as apneas regardless of desaturations– need to be linked to desaturation to be scored. Thus, there are events previously annotated as *severe hypopnea* that may change to *hypopnea* and events where annotations may be removed. Stratification results did not change substantially by omitting the *severe hypopnea* label. Only two errors arose after binary classification (one subject with AHI < 5 was stratified as with an AHI  $\geq$  5 and other with AHI < 15 was stratified as with an AHI  $\geq$  15) and one error after multiclass classification (one subject with AHI < 15 was stratified as with an AHI  $\geq$  15) with respect to the results including the *severe hypopnea* label. Considering that the dataset consisted of 94 subjects, the increase in error was 2.12% after

binary classification and 1.06% after multiclass classification. Therefore, no large differences were obtained to support the need for the label, although the type of error, i.e., false positive in all cases, may be relevant in a screening tool.

### D. Limitations

Arousal-related events are not taken into account in this work, being desaturation-related hypopneas and apneas the only events annotated. This decision was made taking into account that models rely on CVHR pattern detection and that PPI and SpO<sub>2</sub> were the only used signals. First, CVHR pattern is not present in events not associated to desaturations [10]. Second, although PPI can be used to assess arousals to some extent [25] using PPI's DAPs, it has not been demonstrated its feasibility from PPI's Hjorth parameters. Also, arousal assessment using PPI's DAPs has the limitation that some of the DAPs are not related to apneic arousals [25]. In any case, thresholds for OSA stratification should be set observing Oxygen Desaturation Index (ODI) thresholds, rather than AHI's, if arousals are not included [36]. Moreover, no distinction between central, obstructive and mixed apneas was made. The justification is rather similar to that for arousal-related hypopneas, as respiratory effort is not available with limited channel motorization.

Another limitation was introduced by the saturation of the oronasal thermistor signals, that lead to the use of the alternative oronasal pressure for airflow assessment. However, the use of an oronasal pressure sensor instead of a nasal pressure sensor may be considered a half-way solution between the recommended and the alternative, as nasal pressure is criticized because the signal may show decreased amplitude during mouth breathing [29].

The device used for the input signals is a commercial pulse oximeter. Since this work has focused on a screening tool, with the possibility of being used in at-home solutions with wearables, it is possible that the signals available may be of lower quality. Tests should be performed to calculate performance metrics for each case.

Finally, segment classification models used in this work were designed to detect bursts of respiratory events that lead to a CVHR pattern, rather than isolated events. As CVHR is mediated by the parasympathetic system [11], it cannot be detected in patients with autonomic nervous system impairments, such as autonomic neuropathy, multiple system atrophy or Guillain-Barré syndrome.

## V. CONCLUSION

A classifier for during-sleep breathing segments has been presented. This classifier exploits the differences in oscillatory

**TABLE IX**  
BINARY SEGMENT CLASSIFICATION RESULTS (%)

Model	Subgroup	Acc	$P_n$	$R_n$	$P_{ab}$	$R_{ab}$
PPI+ SpO <sub>2</sub>	All subjects	85.01	90.89	87.01	73.52	80.53
	AHI<15	87.33	95.21	90.57	38.06	56.01
	AHI≥15	83.01	84.33	81.53	81.76	84.53
PPI+ SpO <sub>2</sub> + PRV	All subjects	84.98	90.67	87.23	73.71	79.95
	AHI<15	87.27	95.13	90.59	37.77	55.22
	AHI≥15	83.01	83.95	82.06	82.09	83.98

**TABLE X**  
MULTICLASS SEGMENT CLASSIFICATION RESULTS (%)

Model	Subgroup	Acc	$P_n$	$R_n$	$P_{ap}$	$R_{ap}$	$P_h$	$R_h$
PPI+ SpO <sub>2</sub>	All subjects	73.07	92.86	80.19	65.91	60.70	23.71	49.87
	AHI<15	80.39	96.00	84.53	16.25	25.93	19.62	48.22
	AHI≥15	66.77	87.75	73.48	73.08	63.43	25.70	50.51
PPI+ SpO <sub>2</sub> + PRV	All subjects	73.13	92.69	80.52	66.12	60.85	23.11	47.81
	AHI<15	80.43	95.89	84.64	17.00	27.62	19.23	46.31
	AHI≥15	66.85	87.53	74.16	73.36	63.46	24.99	48.39

pattern characteristics of the SpO<sub>2</sub> and PPI signals by using the Hjorth parameters as features. This approach obtained 86.27% accuracy in the binary (*normal-abnormal breathing*) decision, and 73.07% accuracy in the multiclass (*normal breathing-apneic-hypopneic*) decision. A novel index, CVHRI, has been computed in not *normal breathing* segments after segment classification. This index has shown to be strongly correlated with AHI both after binary ( $r = 0.94$ ,  $p < 0.001$ ) and multiclass ( $r = 0.91$ ,  $p < 0.001$ ) segment classification. A better performance has been found in subjects with AHI≥15 rather than in the AHI<15 subgroup. In addition, CVHRI has been used to stratify AHI≥5 and AHI≥15 subgroups, resulting in 77.27% and 79.55% accuracy, respectively. These results suggest that the presented methods provide value for OSA limited-channel screening, allowing monitoring with wearables at home.

#### APPENDIX INCLUSION OF FREQUENCY-DOMAIN PULSE RATE VARIABILITY METRICS AS PREDICTORS

The inclusion of frequency-domain PRV metrics in the study for the detection of apneic segments is justified by the known relationship of OSA with sympathetic overactivity. For this reason, HRV has previously been used as a method to assess cardiac autonomic changes during sleep [37]. To find out whether the inclusion of these metrics can provide even improved results, the best resulting model (PPI+SpO<sub>2</sub> model) was taken as a starting point and power in the low frequency band ( $P_{LF}$ ), in the high frequency band ( $P_{HF}$ ) and  $P_{LF}/P_{HF}$  ratio were added to the inputs. These features were calculated in the same segments as the other models to make the results comparable. Both  $P_{LF}$  and  $P_{HF}$  were computed by trapezoidal integration of the power spectral density estimate obtained by periodogram within the classic windows, i.e., 0.04–0.15 Hz for the low frequency and 0.15–0.4 Hz for the high frequency.

To facilitate comparison with the PPI+SpO<sub>2</sub> model, both model results are shown including and excluding the PRV metrics (see Tables IX and X). The variations in the predictor output

are minimal, demonstrating that the inclusion of these metrics does not significantly improve the model. For simplicity, results related to the subsequent computation of CVHRI and its ability to predict OSA are not shown since the variation in the outcome is as imperceptible as for segment classification.

It should be noted that the use of the classic high-frequency band has been criticized [38]. It is known that respiration affects the boundaries of this autonomic component, therefore it should be studied whether the inclusion of respiratory frequency information to the PRV analysis would allow a more accurate classification.

#### ACKNOWLEDGMENT

Computations were performed by the ICTS NANBIOSIS (HPC Unit at University of Zaragoza).

#### REFERENCES

- [1] Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine, "Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults," *J. Clin. Sleep Med.*, vol. 5, no. 3, pp. 263–276, Jun. 2009.
- [2] A. H. Khandoker, C. K. Karmakar, and M. Palaniswami, "Comparison of pulse rate variability with heart rate variability during obstructive sleep apnea," *Med. Eng. Phys.*, vol. 33, no. 2, pp. 204–209, Mar. 2011.
- [3] *International Classification of Sleep Disorders*, 3rd ed. Darien, IL, USA: American Academy of Sleep Medicine, 2014, pp. 53–62.
- [4] C. V. Senaratna et al., "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, pp. 70–81, Aug. 2017.
- [5] P. E. Peppard, T. Young, J. H. Barnett, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults," *Amer. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, May 2013.
- [6] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705–706, Sep. 1997.
- [7] D. A. Pevernagie et al., "On the rise and fall of the apnea–hypopnea index: A historical review and critical appraisal," *J. Sleep Res.*, vol. 29, no. 4, Feb. 2020, Art. no. e13066.
- [8] W. Randerath et al., "Challenges and perspectives in obstructive sleep apnoea: Report by an ad hoc working group of the sleep disordered breathing group of the European Respiratory Society and the European Sleep Research Society," *Eur. Respir. J.*, vol. 52, no. 3, Sep. 2018.
- [9] S. Ahmadzadeh, J. Luo, and R. Wiffen, "Review on biomedical sensors, technologies and algorithms for diagnosis of sleep disordered breathing: Comprehensive survey," *IEEE Rev. Biomed. Eng.*, vol. 15, pp. 4–22, 2022.
- [10] C. Zwillich, T. Delvin, D. White, N. Douglas, J. Weil, and R. Martin, "Bradycardia during sleep apnea. characteristics and mechanism," *J. Clin. Investigation*, vol. 69, no. 6, pp. 1286–1292, Jun. 1982.
- [11] C. Guilleminault, S. Connolly, R. Winkle, and K. Melvin, "Cyclical variation of the heart rate in sleep apnoea syndrome: Mechanisms, and usefulness of 24 h electrocardiography as a screening technique," *Lancet*, vol. 323, no. 8369, pp. 126–131, Jan. 1984.
- [12] T. Shiomi, C. Guilleminault, R. Sasanabe, I. Hirota, M. Maekawa, and T. Kobayashi, "Augmented very low frequency component of heart rate variability during obstructive sleep apnea," *Sleep*, vol. 19, no. 5, pp. 370–377, Apr. 1996.
- [13] P. K. Stein, S. P. Duntley, P. P. Domitrovich, P. Nishith, and R. M. Carney, "A simple method to identify sleep apnea using Holter recordings," *J. Cardiovasc. Electrophysiol.*, vol. 14, no. 5, pp. 467–473, Mar. 2003.
- [14] J. Hayano et al., "Screening for obstructive sleep apnea by cyclic variation of heart rate," *Circulation Arrhythmia Electrophysiol.*, vol. 4, no. 1, pp. 64–72, Feb. 2011.
- [15] E. Gil, M. Orini, R. Bailon, J. M. Vergara, L. Mainardi, and P. Laguna, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiol. Meas.*, vol. 31, no. 9, Aug. 2010, Art. no. 1271.

- [16] J. Lázaro, E. Gil, J. M. Vergara, and 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. Health Inform.*, vol. 18, no. 1, pp. 240–246, Jun. 2014.
- [17] R. Lazazzera et al., "Detection and classification of sleep apnea and hypopnea using PPG and SpO<sub>2</sub> signals," *IEEE. Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1496–1506, May 2021.
- [18] J. Hayano et al., "Quantitative detection of sleep apnea with wearable watch device," *PLoS One*, vol. 15, no. 11, Nov. 2020, Art. no. e0237279.
- [19] S. Magnúsdóttir and H. Hilmisson, "Ambulatory screening tool for sleep apnea: Analyzing a single-lead electrocardiogram signal (ECG)," *Sleep Breathing*, vol. 22, pp. 421–429, Sep. 2017.
- [20] D. Cajal et al., "Sleep apnea severity stratification by an FFT-based PPG-derived index," in *Proc. IEEE 12th Conf. Eur. Study Group Cardiovasc. Oscillations*, 2022, pp. 1–2.
- [21] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. V. Vaughn, *The AASM Manual for the Scoring of Sleep and Associated Events*, Rules, Terminol. Tech. Specifications Amer. Acad. Sleep Med., Darien, IL, USA, vol. 176. 2012, Art. no. 2012.
- [22] D. Cajal, D. Hernando, J. Lázaro, P. Laguna, E. Gil, and R. Bailón, "Effects of missing data on heart rate variability metrics," *Sensors*, vol. 22, no. 15, Aug. 2022, Art. no. 5774.
- [23] J. Lázaro et al., "Electrocardiogram derived respiration for tracking changes in tidal volume from a wearable armband," in *Proc. IEEE 42nd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2020, pp. 596–599.
- [24] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephal. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, Sep. 1970.
- [25] E. Gil, J. M. Vergara, and P. Laguna, "Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children," *Biomed. Signal Process. Control*, vol. 3, no. 3, pp. 267–277, Jul. 2008.
- [26] J. P. R. Leite and R. L. Moreno, "Heartbeat classification with low computational cost using Hjorth parameters," *IET Signal Process.*, vol. 12, no. 4, pp. 431–438, Jun. 2018.
- [27] J. Hayano et al., "Blunted cyclic variation of heart rate predicts mortality risk in post-myocardial infarction, end-stage renal disease, and chronic heart failure patients," *Europace*, vol. 19, no. 8, pp. 1392–1400, Aug. 2017.
- [28] E. Gil, M. Mendez, J. M. Vergara, S. Cerutti, A. M. Bianchi, and P. Laguna, "Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of PPG signal in children by HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1005–1014, Apr. 2009.
- [29] R. B. Berry et al., "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012.
- [30] M. Deviaene et al., "Sleep apnea detection using pulse photoplethysmography," in *Proc. Comput. Cardiol. Conf.*, 2018, vol. 45, pp. 1–4.
- [31] R. Huttunen et al., "A comparison of signal combinations for deep learning-based simultaneous sleep staging and respiratory event detection," *IEEE. Trans. Biomed. Eng.*, vol. 70, no. 5, pp. 1704–1714, May 2023.
- [32] A. Romem, A. Romem, D. Koldobskiy, and S. M. Scharf, "Diagnosis of obstructive sleep apnea using pulse oximeter derived photoplethysmographic signals," *J. Clin. Sleep. Med.*, vol. 10, no. 3, pp. 285–290, Mar. 2014.
- [33] P. Faßbender, A. Haddad, S. Bürgener, and J. Peters, "Validation of a photoplethysmography device for detection of obstructive sleep apnea in the perioperative setting," *J. Clin. Monit. Comput.*, vol. 33, pp. 341–345, May 2018.
- [34] Y. Li, H. Gao, and Y. Ma, "Evaluation of pulse oximeter derived photoplethysmographic signals for obstructive sleep apnea diagnosis," *Medicine*, vol. 96, no. 18, May 2017, Art. no. e6755.
- [35] S. Nikkonen, I. O. Afara, T. Leppänen, and J. Töyräs, "Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Sep. 2019.
- [36] L. Varghese, G. Rebekah, N. Priya, A. Oliver, and R. Kurien, "Oxygen desaturation index as alternative parameter in screening patients with severe obstructive sleep apnea," *Sleep Sci.*, vol. 15(Spec1), pp. 224–228, Jan. 2022.
- [37] S. Ucak, H. U. Dissanayake, K. Sutherland, P. de Chazal, and P. A. Cistulli, "Heart rate variability and obstructive sleep apnea: Current perspectives and novel technologies," *J. Sleep Res.*, vol. 30, no. 4, Jan. 2021, Art. no. e13274.
- [38] A. Hernando et al., "Inclusion of respiratory frequency information in heart rate variability analysis for stress assessment," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1016–1025, Jul. 2016.